Estimating the probability that a given vector is in the convex hull of a random sample

Satoshi Hayakawa[1] · Terry Lyons[1] · Harald Oberhauser[1]

© The Author(s) 2023

Abstract

For a d-dimensional random vector X, let $p_{n,X}(\theta)$ be the probability that the convex hull of n independent copies of X contains a given point θ. We provide several sharp inequalities regarding $p_{n,X}(\theta)$ and $N_X(\theta)$ denoting the smallest n for which $p_{n,X}(\theta) \geq 1/2$. As a main result, we derive the totally general inequality $1/2 \leq \alpha_X(\theta)N_X(\theta) \leq 3d+1$, where $\alpha_X(\theta)$ (a.k.a. the Tukey depth) is the minimum probability that X is in a fixed closed halfspace containing the point θ. We also show several applications of our general results: one is a moment-based bound on $N_X(\mathbb{E}[X])$, which is an important quantity in randomized approaches to cubature construction or measure reduction problem. Another application is the determination of the canonical convex body included in a random convex polytope given by independent copies of X, where our combinatorial approach allows us to generalize existing results in random matrix community significantly.

Keywords Random convex hull · Tukey depth · Berry–Esseen theorem · Floating body · Cubature

Mathematics Subject Classification Primary 60D05 · Secondary 65C05

1 Introduction

Consider generating independent and identically distributed d-dimensional random vectors. How many vectors do we have to generate in order that a point $\theta \in \mathbb{R}^d$ is contained in the convex hull of the sample with probability at least $1/2$? More generally,

✉ Satoshi Hayakawa
   hayakawa@maths.ox.ac.uk

[1] University of Oxford, Oxford, UK

what is the probability of the event with an n-point sample for each n? These questions were first solved for a general distribution which has a certain symmetry about θ by Wendel [45]. Let us describe the problem more formally.

Let X be a d-dimensional random vector and $X_1, X_2, \ldots$ be independent copies of X. For each $\theta \in \mathbb{R}^d$ and positive integer n, define

$$p_{n,X}(\theta) := \mathbb{P}(\theta \in \text{conv}\{X_1, \ldots, X_n\}),$$

where conv $A := \{\sum_{i=1}^m \lambda_i x_i \mid m \geq 1, \ x_i \in A, \ \lambda_i \geq 0, \ \sum_{i=1}^m \lambda_i = 1\}$ denotes the convex hull of a set $A \subset \mathbb{R}^d$. We also define

$$N_X(\theta) := \inf\{n \mid p_{n,X}(\theta) \geq 1/2\}$$

as the reasonable number of observations we need. As $p_{n,X}$ and N_X are only dependent on the probability distribution of X, we also write $p_{n,\mu}$ and N_μ when X follows the distribution μ. We want to evaluate $p_{n,X}$ as well as N_X for a general X.

Wendel [45] showed that

$$p_{n,X}(0) = 1 - \frac{1}{2^{n-1}} \sum_{i=0}^{d-1} \binom{n-1}{i} \tag{1}$$

holds for an X such that X and $-X$ have the same distribution and $X_1, \ldots, X_d$ are almost surely linearly independent. In particular, $N_X(0) = 2d$ holds for such random vectors. For an X with an absolutely continuous distribution with respect to the Lebesgue measure, Wagner and Welzl [44] showed more generally that the right-hand side of (1) is indeed an upper bound of $p_{n,X}$, and they also characterized the condition for equality (see Theorem 6). Moreover, Kabluchko and Zaporozhets [20] recently gave an explicit formula for $p_{n,X}$ when X is a shifted Gaussian.

In this paper, our aim is to give generic bounds of $p_{n,X}$ and N_X, and we are particularly interested in the upper bound of N_X, which is opposite to the bound given by [44]. Estimating $p_{n,X}$ and N_X is of great interest from application, which ranges from numerical analysis to statistics, and compressed sensing. As a by-product, we also give a general result explaining the deterministic body included in the random polytope conv$\{X_1, \ldots, X_n\}$, which is a sharp generalization of a recent work in the random matrix community [15]. The remainder of this section will explain more detailed motivation from related fields and implications of our results.

Throughout the paper, let $\langle \cdot, \cdot \rangle$ be any inner product on $\mathbb{R}^d$, and $\| \cdot \|$ be the norm it induces.

1.1 Cubature and measure reduction

Let μ be a Borel probability measure on some topological space $\mathcal{X}$. Consider d integrable functions $f_1, \ldots, f_d : \mathcal{X} \to \mathbb{R}$. Then, we know the existence of "good reduction" of μ by Tchakaloff's theorem [3, 40]:

Theorem 1 (Tchakaloff) *There are $d + 1$ points $x_1, \ldots, x_{d+1} \in \operatorname{supp} \mu$ and weights $w_1, \ldots, w_{d+1} \geq 0$ such that $w_1 + \cdots + w_{d+1} = 1$ and*

$$\int_{\mathcal{X}} f_i(x) \, \mathrm{d}\mu(x) = \sum_{j=1}^{d+1} w_j f_i(x_j) \tag{2}$$

holds for each $i = 1, \ldots, d$.

The proof is essentially given by classical Carathéodory's theorem. The points and weights treated in Tchakaloff's theorem is an important object in the field of numerical integration, called cubature [39]. An equivalent problem is also treated as a beneficial way of data compression in the field of data science [8, 26]. A typical choice of test function f_i is monomials when $\mathcal{X}$ is a subset of an Euclidean space, so the integration with respect to the measure $\sum_{j=1}^{d+1} w_j \delta_{x_j}$ is a good approximation of $\int_{\mathcal{X}} f \, \mathrm{d}\mu$ for a smooth integrand f. However, constructions under general setting are also useful; for example, in the cubature on Wiener space [25], $\mathcal{X}$ is the space of continuous paths, μ is the Wiener measure, and the test functions are iterated integrals of paths.

To this generalized cubature construction (or measure reduction) problem, there are efficient deterministic approaches [22, 26, 41] when μ is discrete. Using randomness for construction is recently considered [8, 17] and it is important to know $p_{n,X}(\mathbb{E}[X])$ for the d-dimensional random variable

$$X = \boldsymbol{f}(Y) = (f_1(Y), \ldots, f_d(Y))^\top,$$

where Y is drawn from μ. Indeed, once we have $\mathbb{E}[X] \in \operatorname{conv}\{X_1, \ldots, X_n\}$ ($X_i = \boldsymbol{f}(Y_i)$ are independent copies of X), then we can choose $d + 1$ points and weights satisfying (2) by solving a simple linear programming problem. Evaluation of N_X is sought for estimating the computational complexity of this naive scheme.

1.2 Statistical depth

From the statistical context, $p_{d+1,X}(\theta)$ for a d-dimensional X is called the simplicial depth of $\theta \in \mathbb{R}^d$ with respect to the (population) distribution of X [5, 24], which can be used for mathematically characterizing the intuitive "depth" of each point θ when we are given the distribution of X. For an empirical measure, it corresponds to the number of simplices (whose vertices are in the data) containing θ.

There are also a various concepts measuring depth, all called statistical depth [5, 31]. One of the first such concepts is the halfspace depth proposed by [42]:

$$\alpha_X(\theta) := \inf_{c \in \mathbb{R}^d \setminus \{0\}} \mathbb{P}(\langle c, X - \theta \rangle \leq 0),$$

which can equivalently defined as the minimum measure of a halfspace containing θ. Donoho and Gasko [11] and Rousseeuw and Ruts [36] extensively studied general features of α_X. We call it the Tukey depth throughout the paper.

Our finding is that these two depth notions are indeed deeply related. We prove the rate of convergence $p_{n,X} \to 1$ is essentially determined by α_X (Proposition 13), and we have a beautiful relation $1/2 \le \alpha_X N_X \le 3d + 1$ in Theorem 16.

1.3 Inclusion of deterministic convex bodies

Although we have seen the background of the $p_{n,X}(\theta)$, which only describes the probability of a single vector contained in the random convex polytope, several aspects of such random polytopes have been studied [19, 27]. In particular, people also studied deterministic convex bodies associated with the distribution of a random vector. For example, one consequence of well-known Dvoretzky–Milman's Theorem (see, e.g., [43, Chapter 11]) is that the convex hull of n independent samples from the d-dimensional standard normal distribution is "approximately" a Euclidean ball of radius $\sim \sqrt{\log n}$ with high probability for a sufficiently large n.

Mainly from the context of random matrices, there have been several researches on the interior convex body of $\mathrm{conv}\{X_1, \ldots, X_n\}$ or its "absolute" version $\mathrm{conv}\{\pm X_1, \ldots, \pm X_n\}$ for various classes of X such as Gaussian, Rademacher or vector with i.i.d. subgaussian entries [10, 13, 14, 16, 23]. One result about the Rademacher vector is the following:

Theorem 2 [13] *Let d be a sufficiently large positive integer and $X_1, X_2, \ldots$ be independent samples from the uniform distribution over the set $\{-1, 1\}^d \subset \mathbb{R}^d$. Then, there exists an absolute constant $c > 0$ such that, for each integer $n \ge d(\log d)^2$, we have*

$$\mathrm{conv}\{\pm X_1, \ldots, \pm X_n\} \supset c \left(\sqrt{\log(n/d)} B_2^d \cap B_\infty^d \right)$$

with probability at least $1 - e^{-d}$. Here, B_2^d is the Euclidean unit ball in $\mathbb{R}^d$ and $B_\infty^d = [-1, 1]^d$.

Although each of those results in literature was based on its specific assumptions on the distribution of X, Guédon et al. [15] found a possible way of treating the results in a unified manner under some technical assumptions on X. They introduced the floating body associated with X

$$\tilde{K}^\alpha(X) := \{s \in \mathbb{R}^d \mid \mathbb{P}(\langle s, X \rangle \ge 1) \le \alpha\}$$

to our context (the notation here is slightly changed from the original one), and argued that, under some assumptions on X, with high probability, $\mathrm{conv}\{X_1, \ldots, X_n\}$ includes a constant multiple of the polar body of $\tilde{K}^\alpha(X)$ with $\log(1/\alpha) \sim 1 + \log(n/d)$. Note that their main object of interest is the absolute convex hull, but their results can be extended to the ordinary convex hull (see [15, Remark 1.7]).

Let us explain more formally. Firstly, for a set $A \subset \mathbb{R}^d$, the polar body of A is defined as

$$A^\circ := \{x \in \mathbb{R}^d \mid \langle a, x \rangle \le 1 \text{ for all } a \in A\}.$$

Secondly, we shall describe the assumptions used in [15]. Let $\|\|\cdot\|\|$ be a norm on $\mathbb{R}^d$ and $\gamma, \delta, r, R > 0$ be constants. Their assumptions are as follows:

- (γ, δ) small-ball condition: $\mathbb{P}(|\langle t, X \rangle| \geq \gamma \|\|t\|\|) \geq \delta$ holds for all $t \in \mathbb{R}^d$.
- L_r condition with constant R: $\mathbb{E}\big[|\langle t, X \rangle|^r\big]^{1/r} \leq R\|\|t\|\|$ holds for all $t \in \mathbb{R}^d$.

Under these conditions, they proved the following assertion by using concentration inequalities.

Theorem 3 [15] *Let X be a d-dimensional symmetric random vector that satisfies the small-ball condition and L_r condition for a norm $\|\|\cdot\|\|$ and constants $\gamma, \delta, r, R > 0$. Let $\beta \in (0, 1)$ and set $\alpha = (en/d)^{-\beta}$. Then, there exist a constant $c_0 = c_0(\beta, \delta, r, R/\gamma)$ and an absolute constant $c_1 > 0$ such that, for each integer $n \geq c_0 d$,*

$$\mathrm{conv}\{X_1, \ldots, X_n\} \supset \frac{1}{2}\big(\tilde{K}^\alpha(X)\big)^\circ$$

holds with probability at least $1 - 2\exp(-c_1 n^{1-\beta} d^\beta)$, where $X_1, X_2, \ldots$ are independent copies of X.

Though computing $\big(\tilde{K}^\alpha(X)\big)^\circ$ for individual X is not necessarily an easy task, this gives us a unified understanding of existing results in terms of the polar of the floating body $\tilde{K}^\alpha(X)$. However, its use is limited due to the technical assumptions. In this paper, we show that we can completely remove the assumptions in Theorem 3 and obtain a similar statement only with explicit constants (see Proposition 22 and Corollary 25, or the next section).

Finally, we add that this interior body of random polytopes or its radius is recently reported to be essential in the robustness of sparse recovery [15] and the convergence rate of greedy approximation algorithms [6, 29] when the data is random.

1.4 Organization of the paper

In this paper, our aim is to derive general inequalities for $p_{n,X}$ and N_X. The main part of this paper is Sects. 2, 3, 4 and 5. The following is a broad description of the contents of each section.

- Section 2: General bounds of $p_{n,X}$ without specific quantitative assumptions
- Section 3: Bounds of $p_{n,X}$ uniformly determined by α_X
- Section 4: Bounds of $N_X(\mathbb{E}[X])$ uniformly determined by the moments of X
- Section 5: Results on deterministic convex bodies included in random polytopes

Let us give more detailed explanation about each section. Section 2 provides generalization of the results of [44], and we give generic bounds of $p_{n,X}(\theta)$ under a mild assumption $p_{d,X}(\theta) = 0$, which is satisfied with absolutely continuous distributions as well as typical empirical distributions. Our main result in Sect. 2 is as follows (Theorem 8):

Theorem *Let X be an arbitrary d-dimensional random vector and $\theta \in \mathbb{R}^d$. If $p_{d,X}(\theta) = 0$ holds, then, for any $n \geq m \geq d + 1$, inequalities*

$$p_{n,X}(\theta) \leq 1 - \frac{1}{2^{n-1}} \sum_{i=0}^{d-1} \binom{n-1}{i},$$

$$\frac{1}{2^{n-m}} \frac{\binom{n}{d+1}}{\binom{m}{d+1}} p_{m,X}(\theta) \leq p_{n,X}(\theta) \leq \frac{\binom{n}{d+1}}{\binom{m}{d+1}} p_{m,X}(\theta)$$

hold.

In Sect. 3, we introduce $p_{n,X}^\varepsilon$ and α_X^ε for an $\varepsilon \geq 0$, which are "ε-relaxation" of $p_{n,X}$ and α_X in that $p_{n,X}^0 = p_{n,X}$ and $\alpha_X^0 = \alpha_X$ hold. For this generalization, we prove that the convergence of $p_{n,X}^\varepsilon \to 1$ is uniformly evaluated in terms of α_X^ε (Proposition 13), and obtain the following result (Theorem 14):

Theorem *Let X be an arbitrary d-dimensional random vector and $\theta \in \mathbb{R}^d$. Then, for each $\varepsilon \geq 0$ and positive integer $n \geq 3d/\alpha_X^\varepsilon(\theta)$, we have*

$$p_{n,X}^\varepsilon(\theta) > 1 - \frac{1}{2^d}.$$

Although we do not define ε-relaxation version here, we can see from the case $\varepsilon = 0$ that, for example, $N_X(\theta) \leq \lceil 3d/\alpha_X(\theta) \rceil$ generally holds (see also Theorem 16).

In Sect. 4, we derive upper bounds of N_X without relying on α_X, which may also be unfamiliar. By using the result in the preceding section and the Berry–Esseen theorem, we show some upper bounds of N_X in terms of the (normalized) moments of X as follows (Theorem 19):

Theorem *Let X be a centered d-dimensional random vector with nonsingular covariance matrix V. Then,*

$$N_X \leq 17d \left(1 + \frac{9}{4} \sup_{c \in \mathbb{R}^d, \|c\|_2 = 1} \mathbb{E}\left[\left| c^\top V^{-1/2} X \right|^3 \right]^2 \right)$$

holds.

Here, $\| \cdot \|_2$ denotes the usual Euclidean norm on $\mathbb{R}^d$. Note that the right-hand side can easily be replaced by the moment of $\|V^{-1/2}X\|_2$ (see also Corollary 20).

Section 5 asserts that $K^\alpha(X) := \{\theta \in \mathbb{R}^d \mid \alpha_X(\theta) \geq \alpha\}$ ($\alpha \in (0, 1)$) is a canonical deterministic body included in the random convex polytope $\text{conv}\{X_1, \ldots, X_n\}$. We see in Proposition 22 that this body is essentially equivalent to the $(\tilde{K}^\alpha(X))^\circ$ mentioned in Sect. 1.3, and prove the following (Theorem 24):

Theorem *Let X be an arbitrary symmetric d-dimensional random vector, and let $\alpha, \delta, \varepsilon \in (0, 1)$. If a positive integer n satisfies*

$$n \geq \frac{2d}{\alpha} \max \left\{ \frac{\log(1/\delta)}{d} + \log \frac{1}{\varepsilon}, \ 6 \right\},$$

then we have, with probability at least $1 - \delta$,

$$\text{conv}\{X_1, \ldots, X_n\} \supset (1 - \varepsilon)K^\alpha(X),$$

where $X_1, X_2, \ldots$ are independent copies of X.

A consequence of this theorem (Corollary 25) enables us to remove the technical assumption of Theorem 3.

Note that all these results give explicit constants with reasonable magnitude, which is because of our combinatorial approach typically seen in the proof of Proposition 10 and Proposition 15. After these main sections, we give some implications of our results on motivational examples (introduced in Sects. 1.1, 1.2) in Sect. 6, and we finally give our conclusion in Sect. 7.

2 General bounds of $p_{n,X}$

In this section, we denote $p_{n,X}(0)$ by only $p_{n,X}$. As we always have $p_{n,X}(\theta) = p_{n,X-\theta}(0)$, it suffices to treat $p_{n,X}(0)$ unless we consider properties of $p_{n,X}$ as a function.

Let us start with easier observations. Proposition 4 and Proposition 5 are almost dimension-free. Firstly, as one expects, the following simple assertion holds.

Proposition 4 *For an arbitrary d-dimensional random vector X with $\mathbb{E}[X] = 0$ and $\mathbb{P}(X \neq 0) > 0$, we have*

$$0 < p_{d+1,X} < p_{d+2,X} < \cdots < p_{n,X} < \cdots \to 1.$$

The conclusion still holds if we only assume $p_{n,X} > 0$ for some n instead of $\mathbb{E}[X] = 0$.

Proof For the proof of $p_{2d,X} > 0$, see, e.g., [17]. From this and Carathéodory's theorem, we also have $p_{d+1,X} > 0$. We clearly have $p_{n+1,X} \geq p_{n,X}$ for each $n \geq d+1$.

The strict inequality also seems trivial, but we prove this for completeness. Assume $p_{n+1,X} = p_{n,X}$ for some n. This implies that $0 \notin \text{conv}\{X_i\}_{i=1}^n \Rightarrow 0 \notin \text{conv}\{X_i\}_{i=1}^{n+1}$ holds almost surely. By symmetry, for any $J \subset \{1, \ldots, n+2\}$ with $|J| = n+1$, $0 \notin \text{conv}\{X_i\}_{i=1}^{n+1} \Rightarrow 0 \notin \text{conv}\{X_i\}_{i \in J}$ holds almost surely. Therefore, we have $0 \notin \text{conv}\{X_i\}_{i=1}^n \Rightarrow 0 \notin \text{conv}\{X_i\}_{i=1}^{n+2}$ with probability one. By repeating this argument, we obtain

$$0 \notin \text{conv}\{X_1, \ldots, X_n\} \Longrightarrow 0 \notin \text{conv}\{X_1, \ldots, X_{n+d+1}\}$$
$$\Longrightarrow 0 \notin \text{conv}\{X_{n+1}, \ldots, X_{n+d+1}\}$$

with probability one, but this is only possible when $\mathbb{P}(0 \notin \text{conv}\{X_1, \ldots, X_n\}) = 0$ as $p_{d+1,X} > 0$ and the variables $X_{n+1}, \ldots, X_{n+d+1}$ are independent from the others. This is of course impossible from the assumption $\mathbb{P}(X \neq 0) > 0$ (there exists a unit vector $c \in \mathbb{R}^d$ such that $\mathbb{P}(\langle c, X \rangle > 0) > 0$), so we finally obtain $p_{n,X} < p_{n+1,X}$.

Proving $p_{n,X} \to 1$ is also easy. From the independence, we have

$$
\begin{aligned}
p_{m(d+1),X} &= 1 - \mathbb{P}\big(0 \notin \mathrm{conv}\{X_1, \ldots, X_{m(d+1)}\}\big) \\
&\geq 1 - \mathbb{P}\left(\bigcap_{k=1}^{m}\{0 \notin \mathrm{conv}\{X_{(k-1)(d+1)+1}, \ldots, X_{k(d+1)}\}\}\right) \\
&= 1 - (1 - p_{d+1,X})^m \to 1 \qquad (m \to \infty).
\end{aligned}
$$

This leads to the conclusion combined with the monotonicity of $p_{n,X}$.

Note that we have used the condition $\mathbb{E}[X] = 0$ only to ensure $p_{d+1} > 0$. Hence the latter statement readily holds from the same argument. □

The next one includes a little quantitative relation among $p_{n,X}$ and N_X.

Proposition 5 *For an arbitrary d-dimensional random vector X and integers* $n \geq m \geq d + 1$,

$$
p_{n,X} \leq \binom{n}{m} p_{m,X}, \qquad N_X \leq \frac{n}{p_{n,X}}
$$

hold.

Proof Let M be the number of m-point subsets of $\{X_1, \ldots, X_n\}$ whose convex hull contains 0. Then, we have

$$
\mathbb{E}[M] = \sum_{\substack{J \subset \{1,\ldots,n\} \\ |J|=m}} \mathbb{P}(0 \in \mathrm{conv}\{X_i\}_{i \in J}) = \binom{n}{m} p_{m,X}.
$$

As $p_{n,X} = \mathbb{P}(M \geq 1) \leq \mathbb{E}[M]$, we obtain the first inequality.

For the second part, we carry out the following rough estimate: For the minimum integer k satisfying $(1 - p_{n,X})^k \leq 1/2$, we have $N_X \leq kn$. If $p_{n,X} \geq 1/2$ holds, then $N_X \leq n$ immediately holds. Thus it suffices to prove $k \leq \left\lceil \frac{1-p_{n,X}}{p_{n,X}} \right\rceil$ when $p_{n,X} < 1/2$. Indeed, by the motonicity of $(1 + 1/x)^x$ over $x > 0$, we have

$$
\left(\frac{1}{1 - p_{n,X}}\right)^{\frac{1-p_{n,X}}{p_{n,X}}} = \left(1 + \frac{p_{n,X}}{1 - p_{n,X}}\right)^{\frac{1-p_{n,X}}{p_{n,X}}} \geq 2, \qquad (\because p_{n,X} < 1/2)
$$

so the conclusion follows. □

Remark 1 Although the estimate $N_X \leq \frac{n}{p_{n,X}}$ looks loose in general, $N_X \leq \frac{2d}{p_{2d,X}}$ is a sharp uniform bound for each dimension d up to a universal constant. Indeed, in Examples 34 and 35 (Appendix B), we prove that

$$
\lim_{\varepsilon \searrow 0} \sup_{\substack{X:d\text{-dimensional} \\ p_{2d,X} < \varepsilon}} \frac{N_X p_{2d,X}}{2d} \geq \frac{1}{4}
$$

holds for each positive integer d. In contrast, the other inequality $p_{n,X} \leq \binom{n}{m} p_{m,X}$ is indeed very loose and drastically improved in Proposition 7.

In Propositions 4 and 5, we have never used the information of dimension except for observing $p_{d+1,X} > 0$ in Proposition 4. However, when the distribution of X has a certain regularity, there already exists a strong result that reflects the dimensionality.

Theorem 6 [44] *When the distribution of X is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$,*

$$p_{n,X} \leq 1 - \frac{1}{2^{n-1}} \sum_{i=0}^{d-1} \binom{n-1}{i} = \frac{1}{2^{n-1}} \sum_{i=0}^{n-d-1} \binom{n-1}{i} \tag{3}$$

holds for each $n \geq d + 1$. The equality is attained if and only if the distribution is balanced, i.e., $\mathbb{P}(\langle c, X \rangle \leq 0) = 1/2$ holds for all the unit vectors $c \in \mathbb{R}^d$.

The authors of [44] derived this result by showing the existence of a nonnegative continuous function h_X on $[0, 1]$ such that $h_X(t) = h_X(1 - t)$, $h_X(t) \leq \frac{d+1}{2} \min\{t^d, (1-t)^d\}$ and

$$p_{n,X} = 2\binom{n}{d+1} \int_0^1 t^{n-d-1} h_X(t) \, dt. \tag{4}$$

We shall provide an intuitive description of the function h_X. Let us consider a one-dimensional i.i.d. sequence $Y_1, Y_2, \ldots$ (also independent from $X_1, X_2, \ldots$), where each Y_i follows the uniform distribution over $(0, 1)$. If we consider the $(d + 1)$-dimensional random vectors $\tilde{X}_i := (X_i, Y_i)$, then, for each n, $0 \in \mathrm{conv}\{X_1, \ldots, X_n\} \subset \mathbb{R}^d$ is obviously equivalent to the condition that the $(d + 1)$-th coordinate axis (denoted by ℓ) intersects the convex set $\tilde{C}_n := \mathrm{conv}\{\tilde{X}_1, \ldots, \tilde{X}_n\} \subset \mathbb{R}^{d+1}$.

Under a certain regularity condition, there are exactly two facets (a d-dimensional face of C_n) respectively composed of a $(d+1)$-point subset of $\{\tilde{X}_1, \ldots, \tilde{X}_n\}$ that intersects ℓ. Let us call them *top* and *bottom*, where the top is the facet whose intersection with ℓ has the bigger $(d + 1)$-th coordinate. Let us define another random variable H as

- 0 if ℓ does not intersect $\mathrm{conv}\{\tilde{X}_1, \ldots, \tilde{X}_{d+1}\}$,
- otherwise the probability that 0 and $\tilde{X}_{d+2}$ are on the same side of the hyperplane supporting $\mathrm{conv}\{\tilde{X}_1, \ldots, \tilde{X}_{d+1}\}$ (conditioned by $\tilde{X}_1, \ldots, \tilde{X}_{d+1}$).

Then, for a given realization of $\{\tilde{X}_1, \ldots, \tilde{X}_n\}$, the probability that $\mathrm{conv}\{\tilde{X}_1, \ldots, \tilde{X}_{d+1}\}$ becomes the top of $\tilde{C}_n$ is H^{n-d-1}. As there are $\binom{n}{d+1}$ choice of (equally) possible "top," we can conclude that

$$p_{n,X} = \mathbb{P}\left(\ell \text{ intersects } \tilde{C}_n\right) = \binom{n}{d+1} \mathbb{P}\left(\{X_1, \ldots, X_{d+1}\} \text{ is the top of } \tilde{C}_n\right)$$

$$= \binom{n}{d+1} \mathbb{E}\left[H^{n-d-1}\right].$$

A similar observation shows $p_{n,X} = \binom{n}{d+1}\mathbb{E}[(1-H)^{n-d-1}, \ H > 0]$, and so we can understand h_X as the density of a half mixture of H and $1 - H$ over $\{H > 0\}$. This has been a simplified explanation of h_X. For more rigorous arguments and proofs, see [44].

By using this "densit" function, we can prove the following interesting relationship.

Proposition 7 *Let X be an $\mathbb{R}^d$-valued random variable with an absolutely continuous distribution. Then, for any integers $n \geq m \geq d + 1$, we have*

$$\frac{1}{2^{n-m}}\frac{n(n-1)\cdots(n-d)}{m(m-1)\cdots(m-d)}p_{m,X} \leq p_{n,X} \leq \frac{n(n-1)\cdots(n-d)}{m(m-1)\cdots(m-d)}p_{m,X}. \quad (5)$$

Proof The right inequality is clear from (4). For the left inequality, by using $h_X(t) = h_X(1-t)$, we can rewrite (4) as

$$p_{n,X} = \binom{n}{d+1}\int_0^1 t^{n-d-1}(h_X(t) + h_X(1-t))\, dt$$

$$= \binom{n}{d+1}\int_0^1 (t^{n-d-1} + (1-t)^{n-d-1})h_X(t)\, dt.$$

We can prove for $a \geq b \geq 0$ that $\frac{t^a+(1-t)^a}{t^b+(1-t)^b}$ attains its minimum at $t = 1/2$, e.g., by using the method of Lagrange multipliers. Accordingly, we obtain

$$\frac{p_{n,X}}{\binom{n}{d+1}} = \int_0^1 (t^{n-d-1} + (1-t)^{n-d-1})h_X(t)\, dt$$

$$\geq 2^{m-n}\int_0^1 (t^{m-d-1} + (1-t)^{m-d-1})h_X(t)\, dt = 2^{m-n}\frac{p_{m,X}}{\binom{m}{d+1}},$$

which is equivalent to the inequality to prove.   □

Remark 2 The left inequality has nothing to say when n and m are large so 2^{n-m} is faster than $(n/m)^d$. However, for small n and m, it works as a nice estimate. Consider the case $n = 2d$ and $m = d + 1$. Then, the proposition and the usual estimate for central binomial coefficients yield

$$p_{2d,X} \geq \frac{1}{2^{d-1}}\binom{2d}{d+1}p_{d+1,X} \geq \frac{1}{2^{d-1}}\left(\frac{d}{d+1}\frac{2^{2d}}{2\sqrt{d}}\right)p_{d+1,X} = \frac{2^d\sqrt{d}}{d+1}p_{d+1,X}.$$

This is comparable to the symmetric case, where $p_{d+1,X} = 1/2^d$ and $p_{2d,X} = 1/2$ hold.

The right inequality is an obvious improvement of the dimension-free estimate given in Proposition 5.

We next generalize these results to general distributions including discrete ones such as empirical measures. However, at least we have to assume $p_{d,X} = 0$. Note that

it is weaker than the condition that X has an absolutely continuous distribution, as it is satisfied with usual empirical measures (see Proposition 9).

From smoothing arguments, we obtain the following generalization of inequalities (3) and (5).

Theorem 8 *Let X be an arbitrary d-dimensional random vector with $p_{d,X} = 0$. Then, for any $n \geq m \geq d + 1$, inequalities*

$$p_{n,X} \leq 1 - \frac{1}{2^{n-1}} \sum_{i=0}^{d-1} \binom{n-1}{i}, \qquad \frac{1}{2^{n-m}} \frac{\binom{n}{d+1}}{\binom{m}{d+1}} p_{m,X} \leq p_{n,X} \leq \frac{\binom{n}{d+1}}{\binom{m}{d+1}} p_{m,X}$$

hold.

Proof Let U be a uniform random variable over the unit ball of $\mathbb{R}^d$ which is independent from X. Let also $U_1, U_2 \ldots$ be independent copies of U, which is independent from $X_1, X_2, \ldots$. We shall prove that $\lim_{\varepsilon \searrow 0} p_{n,X+\varepsilon U} = p_{n,X}$ for each n. Note that the distribution of $X + \varepsilon U$ has the probability density function

$$f(x) = \frac{1}{V \varepsilon^d} \mathbb{P}(\|X - x\|_2 \leq \varepsilon),$$

where V denotes the volume of the unit ball. Therefore, once we establish the limit $\lim_{\varepsilon \searrow 0} p_{n,X+\varepsilon U} = p_{n,X}$ the statement of the theorem is clear.

From $p_{d,X} = 0$, we know that

$$q_X(\delta) := \mathbb{P}\left(\inf_{y \in \mathrm{conv}\{X_i\}_{i=1}^d} \|y\| \leq \delta\right) \to 0, \qquad \delta \searrow 0. \tag{6}$$

For each $n \geq d + 1$, consider the event $A_n := \{0 \in \mathrm{conv}\{X_1, \ldots, X_n\}\}$. If the closed ε-ball centered at 0 is included in $\mathrm{conv}\{X_1, \ldots, X_n\}$, then 0 is also contained in $\mathrm{conv}\{X_i + \varepsilon U_i\}_{i=1}^n$ as $\|\varepsilon U_i\| \leq \varepsilon$ for all i (more precisely, we can prove this by using the separating hyperplane theorem). Therefore, by considering the facets of the convex hull, we have

$$\mathbb{P}\left(A_n \cap \bigcap_{\substack{J \subset \{1,\ldots,n\} \\ |J|=d}} \left\{\inf_{y \in \mathrm{conv}\{X_i\}_{i \in J}} \|y\| \geq \varepsilon\right\}\right) \leq \mathbb{P}\left(0 \in \mathrm{conv}\{X_i + \varepsilon U_i\}_{i=1}^n\right) = p_{n,X+\varepsilon U}.$$

By using (6), we have

$$p_{n,X+\varepsilon U} \geq \mathbb{P}(A_n) - \mathbb{P}\left(\bigcup_{\substack{J \subset \{1,\ldots,n\} \\ |J|=d}} \left\{\inf_{y \in \mathrm{conv}\{X_i\}_{i \in J}} \|y\| < \varepsilon\right\}\right)$$

$$\geq p_{n,X} - \binom{n}{d} q_X(\varepsilon) \to p_{n,X} \qquad (\varepsilon \searrow 0),$$

and so we obtain $\liminf_{\varepsilon \searrow 0} p_{n, X + \varepsilon U} \geq p_{n, X}$.

On the other hand, if we have $0 \in \text{conv}\{X_i + \varepsilon U_i\}_{i=1}^n$ and $0 \notin \text{conv}\{X_i\}_{i=1}^n$ at the same time, then there exsits $J \subset \{1, \ldots, n\}$ such that $|J| = d$ and $\inf_{y \in \text{conv}\{X_i\}_{i \in J}} \|y\| \leq \varepsilon$. Indeed, we can write 0 as a convex combination $\sum_{i=1}^n \lambda_i (X_i + \varepsilon U_i) = 0$, so

$$\left\| \sum_{i=1}^n \lambda_i X_i \right\| = \left\| \varepsilon \sum_{i=1}^n \lambda_i U_i \right\| \leq \varepsilon \sum_{i=1}^n \lambda_i \|U_i\| \leq \varepsilon.$$

As $0 \notin \text{conv}\{X_i\}_{i=1}^n$, there is a facet within ε-distance from 0. Therefore, we obtain

$$\mathbb{P}\left(0 \in \text{conv}\{X_i + \varepsilon U_i\}_{i=1}^n \right) \leq \mathbb{P}\left(A_n \cup \bigcup_{\substack{J \subset \{1, \ldots, n\} \\ |J| = d}} \left\{ \inf_{y \in \text{conv}\{X_i\}_{i \in J}} \|y\| \leq \varepsilon \right\} \right),$$

and similarly it follows that

$$p_{n, X + \varepsilon U} \leq p_{n, X} + \binom{n}{d} q_X(\varepsilon) \quad and \quad \limsup_{\varepsilon \searrow 0} p_{n, X + \varepsilon U} \leq p_{n, X}.$$

Thus we finally obtain $\lim_{\varepsilon \searrow 0} p_{n, X + \varepsilon U} = p_{n, X}$. □

We should remark that $p_{d, X} = 0$ is naturally satisfied with (centered) empirical measures.

Proposition 9 *Let μ be an absolutely continuous probability distribution on $\mathbb{R}^d$ and $Y_1, Y_2, \ldots$ be an i.i.d. samplings from μ. Then, with probability one, for each $M \geq d + 1$, distributions*

$$\mu_M := \frac{1}{M} \sum_{i=1}^M \delta_{Y_i} \quad and \quad \tilde{\mu}_M := \frac{1}{M} \sum_{i=1}^M \delta_{Y_i - \frac{1}{M} \sum_{j=1}^M Y_j}$$

satisfy $p_{d, \mu_M} = p_{d, \tilde{\mu}_M} = 0$. $p_{d, \mu_M} = 0$ also holds for $1 \leq M \leq d$ and requires only $p_{d, \mu} = 0$.

Proof For μ_M, it suffices to prove that with probability one there are no $J \subset \{1, \ldots, M\}$ with $|J| = d$ such that $0 \in \text{conv}\{Y_i\}_{i \in J}$. This readily follows from the absolute continuity of the original measure μ. The extension to the case μ satisfies only $p_{d, \mu} = 0$ is immediate.

For the centered version $\tilde{\mu}_M$, what to prove is that with probability one there are no $J \subset \{1, \ldots, M\}$ with $|J| = d$ such that $\frac{1}{M} \sum_{i=1}^M Y_j \in \text{conv}\{Y_i\}_{i \in J}$. Suppose this occurs for some J. Then, we have that $\frac{1}{M-d} \sum_{i \notin J} Y_i$ is on the affine hull of $\{Y_i\}_{i \in J}$. However, as $\{Y_i\}_{i \notin J}$ is independent from $\{Y_i\}_{i \in J}$ for a fixed J, this probability is zero again from the absolute continuity of μ. Therefore, we have the desired conclusion. □

3 Uniform bounds of $p^{\varepsilon}_{n,X}$ via the relaxed Tukey depth

We have not used any quantitative assumption on the distribution of X in the previous section. In this section, however, we shall evaluate $p_{n,X}$ and its ε-approximation version by using the Tukey depth and its relaxation. We shall fix an arbitrarily real inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^d$, and use the induced norm $\| \cdot \|$ and the notation $\mathrm{dist}(x, A) := \inf_{a \in A} \|x - a\|$ for an $x \in \mathbb{R}^d$ and $A \subset \mathbb{R}^d$.

For a d-dimensional random vector X and $\theta \in \mathbb{R}^d$, define an ε-relaxation version of the Tukey depth by

$$\alpha^{\varepsilon}_X(\theta) := \inf_{\|c\|=1} \mathbb{P}(\langle c, X - \theta \rangle \le \varepsilon).$$

We also define, for a positive integer n,

$$p^{\varepsilon}_{n,X}(\theta) := \mathbb{P}(\mathrm{dist}(\theta, \mathrm{conv}\{X_1, \ldots, X_n\}) \le \varepsilon),$$

where $X_1, \ldots, X_n$ are independent copies of X. Note that $p_{n,X} = p^0_{n,X}$. Although we regard them as functions of θ in Sect. 5, we only treat the case $\theta = 0$ and omit the argument θ in this section.

Proposition 10 *Let X be a d-dimensional random vector with an absolutely continuous distribution with respect to the Lebesgue measure. Then, for each $\varepsilon \ge 0$ and positive integer $n \ge d + 1$, we have*

$$1 - p^{\varepsilon}_{n,X} \le \frac{n(1 - \alpha^{\varepsilon}_X)}{n - d}(1 - p^{\varepsilon}_{n-1,X}).$$

Before going into details of quantitative results, we note the following equivalence of the positivity of α^{ε}_X and $p^{\varepsilon}_{n,X}$ which immediately follows from this assertion.

Proposition 11 *Let X be an arbitrary d-dimensional random vector and let $\varepsilon \ge 0$. Then, $p^{\varepsilon}_{n,X} > 0$ for some $n \ge 1$ implies $\alpha^{\varepsilon}_X > 0$. Reciprocally, $\alpha^{\varepsilon}_X > 0$ implies $p^{\varepsilon}_{n,X} > 0$ for all $n \ge d + 1$.*

Proof If $\mathrm{dist}(0, \mathrm{conv}\{X_i\}^n_{i=1}) \le \varepsilon$, there exists a point $x \in \mathrm{conv}\{X_i\}^n_{i=1}$ with $\|x\| \le \varepsilon$. Then, for each $c \in \mathbb{R}^d$ with $\|c\| = 1$, we have $\langle c, x \rangle \le \varepsilon$ and so $\langle c, X_i \rangle \le \varepsilon$ for at least one $i \in \{1, \ldots, n\}$. Hence we have a uniform evaluation

$$\mathbb{P}(\langle c, X \rangle \le \varepsilon) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}(\langle c, X_i \rangle \le \varepsilon) \ge \frac{1}{n} \mathbb{P}\left(\bigcup_{i=1}^{n} \{\langle c, X_i \rangle \le \varepsilon\} \right)$$

$$\ge \frac{1}{n} \mathbb{P}\left(\mathrm{dist}(0, \mathrm{conv}\{X_i\}^n_{i=1}) \le \varepsilon \right),$$

and the first assertion follows.

For the latter, if α^{ε}_X is positive, we have $p^{\varepsilon}_{n,X} > 0$ for a sufficiently large n from Proposition 10. Finally, Carathéodory's theorem yields the positivity for all $n \ge d + 1$. $\qquad \square$

Let us prove Proposition 10.

Proof of Proposition 10 Let $m \geq d$ be an integer. We first consider the quantity $q_m := 1 - p_{n,X}^{\varepsilon}$. Let A_m be the event given by

$$\text{dist}(0, \text{conv}\{X_i\}_{i=1}^m) > \varepsilon.$$

Also, let B_m be the event that $\{X_1, \ldots, X_m\}$ is in general position. Then, we have $\mathbb{P}(B_m) = 1$ and $q_m = \mathbb{P}(A_m \cap B_m)$.

Under the event $A_m \cap B_m$, we have a unique point $h_m \in \text{conv}\{X_i\}_{i=1}^m$ that minimizes $\|h_m\|$. Let H_m be the open halfspace defined by $H_m := \{x \in \mathbb{R}^d \mid \langle x - h_m, h_m \rangle > 0\}$. Then, the boundary ∂H_m is the hyperplane going through h_m and perpendicular to h_m. From the general-position assumption, there are at most d points on ∂H_m. Let I_m be the set of indices i satisfying ∂H_m, then I_m is a random subset of $\{1, \ldots, m\}$ with $1 \leq |I_m| \leq d$ under the event $A_m \cap B_m$. Note also that $X_i \in H$ for each $i \in \{1, \ldots, d\} \setminus I_m$. For simplicity, define $I_m = \emptyset$ for the event $(A_m \cap B_m)^c$.

As I_m is a random set determined uniquely, we can decompose the probability $\mathbb{P}(A_m \cap B_m)$ as follows by symmetry:

$$q_m = \mathbb{P}(A_m \cap B_m) = \sum_{k=1}^d \binom{m}{k} \mathbb{P}(I_m = \{1, \ldots, k\}).$$

Hence, we want to evaluate the probability $\mathbb{P}(I_m = \{1, \ldots, k\})$. Note that we can similarly define h_k as the unique point in $\text{conv}\{X_i\}_{i=1}^k$ that minimizes the distance from the origin. Then, H_k is the open halfspace $H_k = \{x \in \mathbb{R}^d \mid \langle x - h_{m,k}, h_{m,k} \rangle > 0\}$. Then, we have

$$\mathbb{P}(I_m = \{1, \ldots, k\})$$

$$= \mathbb{E}\left[\mathbb{1}_{\{\|h_k\| > \varepsilon, \ \text{conv}\{X_i\}_{i=1}^k \subset \partial H_k\}} \prod_{j=k+1}^m \mathbb{P}\left(X_j \in H_k \mid \text{conv}\{X_i\}_{i=1}^k\right) \right]$$

$$= \mathbb{E}\left[\mathbb{1}_{\{\|h_k\| > \varepsilon, \ \text{conv}\{X_i\}_{i=1}^k \subset \partial H_k\}} \mathbb{P}\left(X' \in H_k \mid \text{conv}\{X_i\}_{i=1}^k\right)^{m-k} \right],$$

where X' is a copy of X independent from $X_1, X_2, \ldots$. As $\mathbb{P}\left(X' \in H_k \mid \text{conv}\{X_i\}_{i=1}^k\right) \leq 1 - \alpha_X^{\varepsilon}$ under the event $\{\|h_k\| > \varepsilon, \ \text{conv}\{X_i\}_{i=1}^k \subset \partial H_k\}$, we have

$$\mathbb{P}(I_{m+1} = \{1, \ldots, k\}) = \mathbb{E}\left[\mathbb{1}_{\{\|h_k\| > \varepsilon, \ \text{conv}\{X_i\}_{i=1}^k \subset \partial H_k\}} \mathbb{P}\left(X' \in H_k \mid \text{conv}\{X_i\}_{i=1}^k\right)^{m+1-k} \right]$$

$$\leq (1 - \alpha_X^{\varepsilon}) \mathbb{P}(I_m = \{1, \ldots, k\}).$$

Therefore, we have

$$q_{m+1} = \sum_{k=1}^{d} \binom{m+1}{k} \mathbb{P}(I_{m+1} = \{1, \ldots, k\})$$

$$= \sum_{k=1}^{d} \frac{m+1}{m+1-k} \binom{m}{k} (1 - \alpha_X^\varepsilon) \mathbb{P}(I_m = \{1, \ldots, k\})$$

$$\leq \frac{(m+1)(1 - \alpha_X^\varepsilon)}{m+1-d} q_m.$$

By letting $n = m + 1$, we obtain the conclusion. $\qquad\square$

If we define $g_{d,n}(\alpha)$ by $g_{d,n} := 1$ for $n = 1, \ldots, d$ and

$$g_{d,n}(\alpha) := \min\left\{1, \frac{n(1-\alpha)}{n-d} g_{d,n-1}(\alpha)\right\} \tag{7}$$

for $n = d + 1, d + 2, \ldots$, we clearly have $1 - p_{n,X}^\varepsilon \leq g_{d,n}(\alpha_X^\varepsilon)$ from Proposition 10 for a d-dimensional X having density. We can actually generalize this to a general X.

Lemma 12 *Let X be an arbitrary d-dimensional random vector. Then, for each $\varepsilon \geq 0$ and positive integer n, we have $1 - p_{n,X}^\varepsilon \leq g_{d,n}(\alpha_X^\varepsilon)$.*

Proof Note first that $g_{d,n}(\alpha)$ is non-increasing with respect to $\alpha \in [0, 1]$. Let $\tilde{X}$ be a d-dimensional random vector such that $\|X - \tilde{X}\| \leq \delta$ for some $\delta > 0$. Then, for an arbitrary $c \in \mathbb{R}^d$ with $\|c\| = 1$, we have

$$\langle c, \tilde{X} \rangle \leq \langle c, X \rangle + \delta,$$

so $\mathbb{P}(\langle c, X \rangle \leq \varepsilon) \leq \mathbb{P}(\langle c, \tilde{X} \rangle \leq \varepsilon + \delta)$. Hence we have $\alpha_X^\varepsilon \leq \alpha_{\tilde{X}}^{\varepsilon+\delta}$.

Consider generating i.i.d. pairs $(X_1, \tilde{X}_1), \ldots, (X_n, \tilde{X}_n)$ that are copies of $(X, \tilde{X})$. Then, for each $x \in \text{conv}\{X_i\}_{i=1}^n$, there is a convex combination such that $x = \sum_{i=1}^n \lambda_i X_i$ with $\lambda_i \geq 0$ and $\sum_{i=1}^n \lambda_i = 1$. Then, we have

$$\left\| x - \sum_{i=1}^n \lambda_i \tilde{X}_i \right\| \leq \sum_{i=1}^n \lambda_i \|X_i - \tilde{X}_i\| \leq \delta.$$

It means that $\inf_{y \in \text{conv}\{\tilde{X}_i\}_{i=1}^n} \|x - y\| \leq \delta$ holds for every $x \in \text{conv}\{X_i\}_{i=1}^n$, and we can deduce that $p_{n,X}^{\varepsilon+2\delta} \geq p_{n,\tilde{X}}^{\varepsilon+\delta}$ holds.

In particular, we can choose $\tilde{X}$ having density, so that we have $1 - p_{n,X}^{\varepsilon+\delta} \leq g_{d,n}(\alpha_{\tilde{X}}^{\varepsilon+\delta})$. Therefore, from the monotonicity of $g_{d,n}$, we have

$$1 - p_{n,X}^{\varepsilon+2\delta} \leq 1 - p_{n,\tilde{X}}^{\varepsilon+\delta} \leq g_{d,n}(\alpha_{\tilde{X}}^{\varepsilon+\delta}) \leq g_{d,n}(\alpha_X^\varepsilon).$$

As $\delta > 0$ can be taken arbitrarily, we finally obtain

$$1 - p_{n,X}^{\varepsilon} \le g_{d,n}(\alpha_X^{\varepsilon})$$

by letting $\delta \to 0$. The δ-relaxation technique used in this proof is a big advantage of introducing $p_{n,X}^{\varepsilon}$ extending $p_{n,X}$.   □

From this lemma, we obtain the following general bound.

Proposition 13 *Let X be an arbitrary d-dimensional random vector. Then, for each $\varepsilon \ge 0$ and positive integer $n \ge d/\alpha_X^{\varepsilon}$, we have*

$$1 - p_{n,X}^{\varepsilon} \le \left(\frac{n\alpha_X^{\varepsilon}}{d} \exp\left\{ \left(\frac{1}{\alpha_X^{\varepsilon}} \log \frac{1}{1 - \alpha_X^{\varepsilon}} \right) \left(1 + \alpha_X^{\varepsilon} - \frac{n\alpha_X^{\varepsilon}}{d} \right) \right\} \right)^d .$$

Proof From Lemma 12, it suffices to prove that

$$g_{d,n}(\alpha) \le \left(\frac{n\alpha}{d} \exp\left\{ \left(\frac{1}{\alpha} \log \frac{1}{1 - \alpha} \right) \left(1 + \alpha - \frac{n\alpha}{d} \right) \right\} \right)^d \tag{8}$$

holds for each $\alpha \in (0, 1)$ and $n \ge d/\alpha$. From the definition of $g_{d,n}$ (see (7)), if we set $n_0 := \lceil d/\alpha \rceil$, then we have

$$
\begin{aligned}
g_{d,n}(\alpha) &\le \frac{n(n-1)\cdots n_0}{(n-d)(n-d-1)\cdots(n_0-d)} (1-\alpha)^{n-n_0+1} g_{d,n_0-1}(\alpha) \\
&\le \frac{n(n-1)\cdots(n-d+1)}{(n_0-1)(n_0-2)\cdots(n_0-d)} (1-\alpha)^{n-n_0+1} \\
&\le \left(\frac{n}{n_0-d} \right)^d (1-\alpha)^{n-n_0+1}.
\end{aligned}
$$

As we know $d/\alpha \le n_0 < d/\alpha + 1$ by definition, we have

$$g_{d,n}(\alpha) \le \left(\frac{n}{d/\alpha - d} \right)^d (1-\alpha)^{n-\frac{d}{\alpha}} = \left(\frac{n\alpha}{d} \right)^d (1-\alpha)^{n-\frac{d}{\alpha}-d}.$$

This is indeed the desired inequality (8).   □

Remark 3 As $\frac{1}{\alpha} \log \frac{1}{1-\alpha} \ge 1$ holds on $(0, 1)$ for $n \ge \frac{(1+\alpha)d}{\alpha}$, the bound (8) yields a looser but more understandable variant

$$g_{d,n}(\alpha) \le \left(\frac{n\alpha}{d} \exp\left(1 + \alpha - \frac{n\alpha}{d} \right) \right)^d .$$

Note that we have a trivial lower bound of $1 - p_{n,X}^{\varepsilon} \ge (1 - \alpha_X^{\varepsilon})^n$, which is proven by fixing a separating hyperplane between the origin and sample points.

For a special choice $n = \lceil 3d/\alpha \rceil$, the following is readily available:

Theorem 14 *Let X be an arbitrary d-dimensional random vector. Then, for each $\varepsilon \geq 0$ and positive integer $n \geq 3d/\alpha_X^\varepsilon$, we have*

$$p_{n,X}^\varepsilon > 1 - \frac{1}{2d}.$$

Proof From Proposition 13, it suffices to prove

$$3 \exp\left\{\left(\frac{1}{\alpha} \log \frac{1}{1-\alpha}\right)(\alpha - 2)\right\} < \frac{1}{2} \tag{9}$$

for all $\alpha \in (0, 1)$. If we let $f(x) = \frac{x-2}{x} \log \frac{1}{1-x}$ for $x \in (0, 1)$, then we have

$$f'(x) = \frac{1}{x^2}\left(2 \log \frac{1}{1-x} - \frac{x(2-x)}{1-x}\right) = \frac{1}{x^2}\left(2 \log \frac{1}{1-x} + (1-x) - \frac{1}{1-x}\right).$$

If we set $t := \log \frac{1}{1-x}$, t takes positive reals and we have

$$2 \log \frac{1}{1-x} + (1-x) - \frac{1}{1-x} = 2t + e^{-t} - e^t = 2(t - \sinh t) < 0.$$

Therefore, it suffices to consider the limit $\alpha \searrow 0$. In this limit, the left-hand side of (9) is equal to $3e^{-2}$, which is smaller than $1/2$ since $e > \sqrt{6}$ holds. $\qquad\square$

We complete this section with a stronger version of Proposition 10 only for $\varepsilon = 0$. Indeed, by summing up the following inequality, we can immediately obtain the $\varepsilon = 0$ case in Proposition 10.

Proposition 15 *Let X be a d-dimensional random vector with an absolutely continuous distribution with respect to the Lebesgue measure. Then,*

$$p_{n+1,X} - p_{n,X} \leq \frac{n(1 - \alpha_X)}{n - d}(p_{n,X} - p_{n-1,X})$$

holds for all $n \geq d + 1$.

Proof First, observe that $p_{n+1,X} - p_{n,X} = \mathbb{P}(0) \in \mathrm{conv}\{X_1, \ldots, X_{n+1}\} \backslash \mathrm{conv}\{X_1, \ldots, X_n\}$ for $n \geq d + 1$ and independent copies $X_1, X_2, \ldots$ of X. Assume $0 \in \mathrm{conv}\{X_1, \ldots, X_{n+1}\} \backslash \mathrm{conv}\{X_1, \ldots, X_n\}$ holds and no $d + 1$ points of $\{0, X_1, \ldots, X_{n+1}\}$ lie on the same hyperplane (the latter is satisfied almost surely as X is absolutely continuous). Then, there exists an expression such that

$$0 = \sum_{i=1}^{n+1} \lambda_i X_i, \quad \sum_{i=1}^{n+1} \lambda_i = 1, \quad \lambda_i \geq 0.$$

Here $0 < \lambda_{n+1} < 1$ must hold as $0 \notin \mathrm{conv}\{X_1, \dots, X_n\}$ and $X_{n+1} \neq 0$. Therefore, we can rewrite

$$\frac{1}{1 - \lambda_{n+1}} \sum_{i=1}^{n} \lambda_i X_i = -\frac{\lambda_{n+1}}{1 - \lambda_{n+1}} X_{n+1}$$

and this left-hand side is a convex combination of $\{X_1, \dots, X_n\}$. Therefore, the line ℓ passing through X_{n+1} and 0 intersects $\mathrm{conv}\{X_1, \dots, X_n\}$ after 0 (if directed from X_{n+1} to 0). Also, ℓ never intersects $\mathrm{conv}\{X_1, \dots, X_n\}$ before 0. Indeed, if $\lambda X_{n+1} \in \mathrm{conv}\{X_1, \dots, X_n\}$ for some $\lambda > 0$, then $0 \in \mathrm{conv}\{\lambda X_{n+1}, -\frac{\lambda_{n+1}}{1-\lambda_{n+1}} X_{n+1}\} \subset \mathrm{conv}\{X_1, \dots, X_n\}$ holds and it contradicts the assumption.

Hence, we can define the first hitting point of ℓ and $\mathrm{conv}\{X_1, \dots, X_n\}$ after 0. More formally, let P be the minimum-normed point in $\ell \cap \mathrm{conv}\{X_1, \dots, X_n\}$. Then, by the general-position assumption, there exists a unique $J \subset \{1, \dots, n\}$ with $|J| = d$ such that $P \in \mathrm{conv}\{X_i\}_{i \in J}$ (more strongly, P is in the relative interior of $\mathrm{conv}\{X_i\}_{i \in J}$). In other words, $\mathrm{conv}\{X_i\}_{i \in J}$ is the unique facet which intersects ℓ first. Then, there exists a unique normal vector c_J that defines the hyperplane supporting $\{X_i\}_{i \in J}$, i.e., $\langle c_J, X_i \rangle = 1$ for each $i \in J$. Since $\langle c_J, P \rangle = 1$ also holds, we have $\langle c_J, X_{n+1} \rangle < 0$. We can also prove $\langle c_J, X_i \rangle > 1$ for each $i \in \{1, \dots, n\} \backslash J$. Indeed, if we have $\langle c_J, X_j \rangle < 1$ for some $j \in \{1, \dots, n\} \backslash J$, then there are interior points of $\mathrm{conv}\{X_i\}_{i \in J \cup \{j\}}$ that belongs to ℓ and this contradicts the minimality of the norm of P.

Therefore, for a fixed $J \subset \{1, \dots, n\}$ with $|J| = d$, the probability that $0 \in \mathrm{conv}\{X_1, \dots, X_{n+1}\} \backslash \mathrm{conv}\{X_1, \dots, X_n\}$ holds and $\mathrm{conv}\{X_i\}_{i \in J}$ becomes the first facet intersecting ℓ after 0 is, from the independence,

$$\mathbb{E}\left[\mathbb{P}\big(0 \in \mathrm{conv}\{X_i\}_{i \in J \cup \{n+1\}} \mid \{X_i\}_{i \in J}\big) \prod_{j \in \{1,\dots,n\} \backslash J} \mathbb{P}\big(\langle c_J, X_j \rangle > 1 \mid \{X_i\}_{i \in J}\big) \right]$$
$$= \mathbb{E}\left[\mathbb{P}\big(0 \in \mathrm{conv}\{X_i\}_{i \in J \cup \{n+1\}} \mid \{X_i\}_{i \in J}\big) \mathbb{P}\big(\langle c_J, X' \rangle > 1 \mid \{X_i\}_{i \in J}\big)^{n-d} \right],$$

where X' is a copy of X independent from $\{X_i\}_{i \geq 1}$. By symmetry, this J is chosen with equal probability given $0 \in \mathrm{conv}\{X_1, \dots, X_{n+1}\} \backslash \mathrm{conv}\{X_1, \dots, X_n\}$ (almost surely without overlapping). Hence, we obtain

$$p_{n+1,X} - p_{n,X}$$
$$= \binom{n}{d} \mathbb{E}\left[\mathbb{P}(0 \in \mathrm{conv}\{X_1, \dots, X_{d+1}\} \mid \{X_i\}_{i \in I}) \mathbb{P}\big(\langle c_I, X' \rangle > 1 \mid \{X_i\}_{i \in I}\big)^{n-d} \right],$$

where $I = \{1, \dots, d\}$. Observe that this representation is still valid for $n = d$. From the definition of α_X, we have $\mathbb{P}\big(\langle c_I, X' \rangle > 1 \mid \{X_i\}_{i \in I}\big) \leq 1 - \alpha_X$, so finally obtain, for $n \geq d + 1$,

$$p_{n+1,X} - p_{n,X}$$

$$= \binom{n}{d} \mathbb{E}\left[\mathbb{P}(0 \in \text{conv}\{X_1, \ldots, X_{d+1}\} \mid \{X_i\}_{i \in I}) \, \mathbb{P}(\langle c_I, X' \rangle > 1 \mid \{X_i\}_{i \in I})^{n-d} \right]$$

$$\leq (1 - \alpha_X) \binom{n}{d} \mathbb{E}\left[\mathbb{P}(0 \in \text{conv}\{X_1, \ldots, X_{d+1}\} \mid \{X_i\}_{i \in I}) \, \mathbb{P}(\langle c_I, X' \rangle > 1 \mid \{X_i\}_{i \in I})^{n-1-d} \right]$$

$$= (1 - \alpha_X) \frac{\binom{n}{d}}{\binom{n-1}{d}} (p_{n,X} - p_{n-1,X})$$

$$= \frac{n(1 - \alpha_X)}{n - d} (p_{n,X} - p_{n-1,X}).$$

This is the desired inequality. □

4 Bounds of N_X via Berry–Esseen theorem

In this section, we discuss upper bounds of N_X for a centered X, which are of particular interest from the randomized measure reduction (see Sect. 1.1).

We know the following assertion as a consequence of Theorem 14.

Theorem 16 *Let X be an arbitrary d-dimensional random vector. Then, we have*

$$\frac{1}{2\alpha_X} \leq N_X \leq \left\lceil \frac{3d}{\alpha_X} \right\rceil.$$

Proof The right inequality is an immediate consequence of Theorem 14. To prove the left one, let n be a positive integer satisfying $\frac{1}{2n} > \alpha_X$. Then, there exists a vector $c \in \mathbb{R}^d \setminus \{0\}$ such that $\mathbb{P}(c^\top X \leq 0) < \frac{1}{2n}$. Then, for $X_1, X_2, \ldots, X_n$ (i.i.d. copies of X), we have

$$p_{n,X} = \mathbb{P}(0 \in \text{conv}\{X_1, \ldots, X_n\}) \leq \mathbb{P}\left(\bigcup_{i=1}^n \{c^\top X_i \leq 0\} \right) \leq n\mathbb{P}(c^\top X \leq 0) < \frac{1}{2}.$$

Therefore, N_X must satisfy $\frac{1}{2N_X} \leq \alpha_X$. □

Remark 4 The above theorem states that $1/2 \leq \alpha_X N_X \leq 3d + 1$. This evaluation for $\alpha_X N_X$ is indeed tight up to a universal constant. For example, if X is a d-dimensional standard Gaussian, we have $\alpha_X = \frac{1}{2}$ and $N_X = 2d$, so $\alpha_X N_X = d$. Moreover, for a small $\varepsilon \in (0, 1)$, if we consider $X = (X^1, \ldots, X^d)$ such that

- $\mathbb{P}(X^d = 1) = \varepsilon$ and $\mathbb{P}(X^d = -1) = 1 - \varepsilon$,
- $(X^1, \ldots, X^{d-1})|_{X^d=1}$ is a standard Gaussian,
- $X^1 = \cdots = X^{d-1} = 0$ if $X^d = -1$,

then we can see $\alpha_X = \varepsilon/2$ and $N_X = \Omega((d-1)/\varepsilon)$ as $(0, \ldots, 0, 1)$ has to be in the convex hull of samples to include the origin in it. Hence the bound $\alpha_X N_X = \mathcal{O}(d)$ is sharp even for a small α_X.

On the contrary,

$$\inf_{X:d\text{-dimensional}} \alpha_X N_X \leq 2$$

holds (even when requiring $p_{d,X} = 0$) for each positive integer d from Example 34 and Example 35 in the appendix (Sect. B).

Although Theorem 16 has strong generality, in many situations we have little information about the Tukey depth α_X. Indeed, approximately computing the Tukey depth itself is an important and difficult problem [9, 47]. However, if we limit the argument to a centered X, we can obtain various moment-based bounds as shown below. In this section, we use the usual Euclidean norm $\|\cdot\|_2$ given by $\|x\|_2 = \sqrt{x^\top x}$ for simplicity.

Let X be a d-dimensional centered random vector whose covariance matrix $V := \mathbb{E}[XX^\top]$ is nonsingular. We also define $V^{-1/2}$ as the positive-definite square root of V^{-1}. Then, for each unit vector $c \in \mathbb{R}^d$ (namely $\|c\|_2 = 1$), we have

$$\mathbb{E}\left[(c^\top V^{-1/2}X)^2\right] = \mathbb{E}\left[c^\top V^{-1/2}XX^\top V^{-1/2}c\right] = \mathbb{E}\left[c^\top c\right] = 1, \qquad (10)$$

We have the following simple result for a bounded X.

Proposition 17 *Let X be a centered d-dimensional random vector with nonsingular covariance matrix V. If $\|V^{-1/2}X\|_2 \leq B$ holds almost surely for a positive constant B, then we have*

$$\alpha_X \geq \frac{1}{2B^2}, \qquad N_X \leq \lceil 6dB^2 \rceil.$$

Proof For a one-dimensional random variable Y with $\mathbb{E}[Y] = 0$, $\mathbb{E}[Y^2] = 1$ and $|Y| \leq B$, we have

$$B\mathbb{P}(Y \leq 0) \geq \mathbb{E}[-\min\{Y, 0\}] = \frac{1}{2}\mathbb{E}[|Y|]$$

and so

$$\mathbb{P}(Y \leq 0) \geq \frac{\mathbb{E}[|Y|]}{2B} \geq \frac{\mathbb{E}\left[|Y|^2\right]}{2B^2} = \frac{1}{2B^2}.$$

By observing this inequality for each $Y = c^\top V^{-1/2}X$ with $\|c\|_2 = 1$, we obtain the bound of α_X. The latter bound then follows from Theorem 16. □

Let us consider the unbounded case. The Berry–Esseen theorem evaluates the speed of convergence in the central limit theorem [4, 12]. The following is a recent result with an explicit small constant.

Theorem 18 [21] *Let Y be a random variable with $\mathbb{E}[Y] = 0$, $\mathbb{E}[Y^2] = 1$, and $\mathbb{E}[|Y|^3] < \infty$, and let $Y_1, Y_2, \ldots$ be independent copies of Y. Also let Z be one-dimensional standard Gaussian. Then, we have*

$$\left| \mathbb{P}\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \leq x \right) - \mathbb{P}(Z \leq x) \right| \leq \frac{0.4784 \, \mathbb{E}[|Y|^3]}{\sqrt{n}}$$

for arbitrary $x \in \mathbb{R}$ and $n \geq 1$.

We can apply the Berry–Esseen theorem for evaluating the probability $\mathbb{P}(c^\top S_n \leq 0)$ from (10), where S_n is the normalized i.i.d. sum $\frac{1}{\sqrt{n}} V^{-1/2}(X_1 + \cdots + X_n)$. By elaborating this idea, we obtain the following bound of N_X.

Theorem 19 *Let X be a centered d-dimensional random vector with nonsingular covariance matrix V. Then,*

$$N_X \leq 17d \left(1 + \frac{9}{4} \sup_{c \in \mathbb{R}^d, \|c\|_2 = 1} \mathbb{E}\left[\left| c^\top V^{-1/2} X \right|^3 \right]^2 \right)$$

holds.

Proof Let n be an integer satisfying

$$n \geq \frac{9}{4} \sup_{c \in \mathbb{R}^d, \|c\|_2 = 1} \mathbb{E}\left[\left| c^\top V^{-1/2} X \right|^3 \right]^2.$$

Then, for an arbitrary $\|c\|_2 = 1$, from Theorem 18, we have

$$\mathbb{P}\left(\frac{c^\top V^{-1/2}(X_1 + \cdots + X_n)}{n} \leq 0 \right) = \mathbb{P}\left(\frac{c^\top V^{-1/2}(X_1 + \cdots + X_n)}{\sqrt{n}} \leq 0 \right)$$

$$\geq \frac{1}{2} - \frac{2}{3} \cdot 0.48 = \frac{9}{50},$$

where $X_1, X_2, \ldots$ are independent copies of X. Hence $\alpha_{n^{-1}(X_1 + \cdots + X_n)} \geq 9/50$ holds. Then we can use Theorem 16 to obtain

$$N_{n^{-1}(X_1 + \cdots + X_n)} \leq \left\lceil \frac{50}{9} \cdot 3d \right\rceil \leq 17d.$$

Since $N_X \leq n N_{n^{-1}(X_1 + \cdots + X_n)}$ holds, we have

$$N_X \leq 17d \left(1 + \frac{9}{4} \sup_{c \in \mathbb{R}^d, \|c\|_2 = 1} \mathbb{E}\left[\left| c^\top V^{-1/2} X \right|^3 \right]^2 \right),$$

which is the desired conclusion. □

Remark 5 The bound in Theorem 19 is sharp up to constant as a uniform bound in terms of $\mathbb{E}\left[\left| c^\top V^{-1/2} X \right|^3 \right]$. Indeed, if X is d-dimensional standard Gaussian, then $\mathbb{E}\left[\left| c^\top V^{-1/2} X \right|^3 \right] = \frac{2\sqrt{2}}{\sqrt{\pi}}$ holds for all $\|c\|_2 = 1$ while $N_X = 2d$, so we have

$$\sup_{c\in\mathbb{R}^d, \|c\|_2=1} \mathbb{E}\left[\left|c^\top V^{-1/2}X\right|^3\right]^{-2} N_X = \frac{\pi}{4}d.$$

From Theorem 19, we can also obtain several looser but more tractable bounds.

Corollary 20 *Let X be a centered d-dimensional random vector with nonsingular covariance matrix V. N_X can be bounded as*

$$N_X \leq 17d \left(1 + \frac{9}{4}\min\left\{\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]^2, \mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^4\right]\right\}\right).$$

Proof From Theorem 19, it suffices to prove

$$\mathbb{E}\left[\left|c^\top V^{-1/2}X\right|^3\right]^2 \leq \mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]^2, \ \mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^4\right]$$

for each unit vector $c \in \mathbb{R}^d$. The first bound is clear from

$$\left|c^\top V^{-1/2}X\right| \leq \|c\|_2 \left\|V^{-1/2}X\right\|_2 = \left\|V^{-1/2}X\right\|_2.$$

The second bound can also be derived as

$$\mathbb{E}\left[\left|c^\top V^{-1/2}X\right|^3\right]^2 \leq \mathbb{E}\left[\left|c^\top V^{-1/2}X\right|^2\right]\mathbb{E}\left[\left|c^\top V^{-1/2}X\right|^4\right]$$

$$= \mathbb{E}\left[\left|c^\top V^{-1/2}X\right|^4\right] \leq \mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^4\right],$$

where we have used the Cauchy–Schwarz inequality. □

Remark 6 In the order notation, the first bound in this corollary states

$$N_X = \mathcal{O}\left(d\, \mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]^2\right).$$

This estimate is also sharp up to $\mathcal{O}(d)$ factor in the sense that we can prove

$$\sup\left\{\left.\frac{N_X}{\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]^2} \right| \begin{array}{l} X\text{is }d\text{-dimensional, } \mathbb{E}[X]=0,\\ V=\mathbb{E}[XX^\top]\text{ is nonsingular, } \mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]<\infty \end{array}\right\} \geq \frac{1}{2}$$

for each positive integer d. For the proof of this fact, see Example 34 and Example 35 in the appendix (Sect. B).

We finally remark that there are multivariate versions of the Berry–Esseen theorem [35, 46] and we can use them to derive a bound of N_X in a different approach which does not use α_X. However, their bounds only gives the estimate

$$N_X = \mathcal{O}\left(d^{7/2}\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]^2\right), \tag{11}$$

which is far worse than the bounds obtained in Theorem 19 and Corollary 20. However, it is notable that this approach from multidimensional Berry–Esseen formulas is applicable to *non-identical* X_i's if the second and third moments are uniformly bounded, while the combinatorial approach based on α_X seems to be fully exploiting the i.i.d. assumption. Therefore, we provide the details of this alternative approach in the appendix (Sect. A).

5 Deterministic interior body of random polytopes

For each $\alpha > 0$, define a deterministic set defined by the level sets of Tukey depth

$$K^\alpha(X) := \{\theta \in \mathbb{R}^d \mid \alpha_X(\theta) \geq \alpha\}.$$

This set is known to be compact and convex [36]. We can also naturally generalize this set for the ε-relaxation of Tukey depth, and the generalization also satisfies the following:

Proposition 21 *Let X be a d-dimensional random vector. Then, for each $\varepsilon \geq 0$ and $\alpha > 0$, the set $\{\theta \in \mathbb{R}^d \mid \alpha_X^\varepsilon(\theta) \geq \alpha\}$ is compact and convex, and satisfies*

$$\{\theta \in \mathbb{R}^d \mid \alpha_X^\varepsilon(\theta) \geq \alpha\} \supset \{\theta \in \mathbb{R}^d \mid \mathrm{dist}(\theta, K^\alpha(X)) \leq \varepsilon\}.$$

Proof We fix α and denote

$$K_\varepsilon = \{\theta \in \mathbb{R}^d \mid \alpha_X^\varepsilon(\theta) \geq \alpha\}.$$

Note that $K_0 = K^\alpha(X)$. Let $c \in \mathbb{R}^d$ satisfy $\|c\| = 1$. Define $t(c)$ by

$$t(c) := \inf\{t \in \mathbb{R} \mid \mathbb{P}(\langle c, X \rangle \leq t) \geq \alpha\}. \tag{12}$$

If $t(c) = \infty$, i.e., the right-hand set is empty for some c, then each set K_ε is empty. $t(c) > -\infty$ is clear from $\alpha > 0$. Suppose $t(c) \in \mathbb{R}$ for all c. From the continuity of probability, the infimum can actually be replaced by minimum, so we have

$$\mathbb{P}(\langle c, X - \theta \rangle \leq \varepsilon) \geq \alpha \iff \langle c, \theta \rangle + \varepsilon \geq t(c)$$

for each $\theta \in \mathbb{R}^d$. Hence, if $\theta_0 \in K_0$ and $\|\theta - \theta_0\| \leq \varepsilon$, then we have $\theta \in K_\varepsilon$, so we obtain the inclusion statement.

Let us prove that K_ε is compact and convex. Define $H_\varepsilon(c) := \{\theta \in \mathbb{R}^d \mid \langle c, \theta \rangle \geq t(c) - \varepsilon\}$ for each $c \in \mathbb{R}^d$ with $\|c\| = 1$. From (12), we have $K_\varepsilon = \bigcap_{\|c\|=1} H_\varepsilon(c)$. As $H_\varepsilon(c)$ is closed and convex, K_ε is also closed and convex. To prove compactness, we shall prove K_ε is bounded. As X is a random vector, there is an $R > 0$ such that $\mathbb{P}(\|X\| \geq R) < \alpha$. Then, for each $\theta \in \mathbb{R}^d$ satisfying $\|\theta\| \geq R + \varepsilon$, we have

$$\mathbb{P}\left(\left\langle -\frac{\theta}{\|\theta\|}, X - \theta \right\rangle \leq \varepsilon \right) = \mathbb{P}\left(\left\langle -\frac{\theta}{\|\theta\|}, X \right\rangle \leq \varepsilon - \|\theta\| \right) \leq \mathbb{P}(\|X\| \geq R) < \alpha.$$

Therefore, we have $\|\theta\| < R + \varepsilon$ for each $\theta \in K_\varepsilon$ and so K_ε is bounded. □

Remark 7 Note that the inclusion stated in Proposition 21 can be strict. For example, if X is a d-dimensional standard Gaussian, $K^\alpha(X)$ is empty for each $\alpha > 1/2$, but the ε-relaxation of Tukey depth can be greater than $1/2$ for $\varepsilon > 0$.

From this proposition, we can naturally generalize the arguments given in this section to the ε-relaxation case; natural interior bodies of ε-neighborhood of $\text{conv}\{X_1, \ldots, X_n\}$ are given by the ε-relaxation of Tukey depth. However, to keep the notation simple, we only treat $K^\alpha(X)$ the interior body of usual convex hull in the following.

We next prove that the polar body $\left(\tilde{K}^\alpha(X)\right)^\circ$ used in [15], which we have introduced in Sect. 1.3, is essentially the same as $K^\alpha(X)$ in their setting, i.e., when X is symmetric. Recall that $\tilde{K}^\alpha(X)$ is defined as

$$\tilde{K}^\alpha(X) = \{s \in \mathbb{R}^d \mid \mathbb{P}(\langle s, X \rangle \geq 1) \leq \alpha\}.$$

Note that the following proposition is not surprising if we go back to the original background of $\tilde{K}^\alpha$ [37], where X is uniform from some deterministic convex set, and recent reseaches on its deep relation to the Tukey depth [32].

Proposition 22 *Let X be a d-dimensional symmetric random vector. Then, for each $\alpha \in (0, 1/2)$, we have*

$$\{\theta \in \mathbb{R}^d \mid \alpha_X(\theta) > \alpha\} \subset \left(\tilde{K}^\alpha(X)\right)^\circ \subset K^\alpha(X).$$

Proof Consider the set

$$A^\alpha := \{s \in \mathbb{R}^d \mid \mathbb{P}(\langle s, X \rangle \geq 1) < \alpha\}.$$

Then, we clearly have $A^\alpha \subset \tilde{K}^\alpha(X)$ and so $(A^\alpha)^\circ \supset \left(\tilde{K}^\alpha(X)\right)^\circ$. We first prove that $(A^\alpha)^\circ = K^\alpha(X)$ actually holds. From the definition of a polar, $\theta \in (A^\alpha)^\circ$ if and only if

$$\mathbb{P}(\langle s, X \rangle \geq 1) < \alpha \quad \Longrightarrow \quad \langle s, \theta \rangle \leq 1$$

holds for each $s \in \mathbb{R}^d \backslash \{0\}$. If we represent $s = r^{-1}c$ by $r > 0$ and $c \in \mathbb{R}^d$ with $\|c\| = 1$, this is equivalent to

$$\mathbb{P}(\langle c, X \rangle \geq r) < \alpha \implies \langle c, \theta \rangle \leq r \tag{13}$$

for each $r > 0$ and $\|c\| = 1$. As we have assumed that X is symmetric and $\alpha < 1/2$, (13) is still equivalent even if we allow r to take all reals.

We shall prove that, for a fixed c, (13) is equivalent to $\mathbb{P}(\langle c, X - \theta \rangle \geq 0) \geq \alpha$. Indeed, if

$$\mathbb{P}(\langle c, X - \theta \rangle \geq 0) = \mathbb{P}(\langle c, X \rangle \geq \langle c, \theta \rangle) < \alpha$$

holds, there exists a $\delta > 0$ such that $\mathbb{P}(\langle c, X \rangle \geq \langle c, \theta \rangle - \delta) < \alpha$. Then, we have the negation of (13) by letting $r = \langle c, \theta \rangle - \delta$. For the opposite direction, if we assume $\mathbb{P}(\langle c, X \rangle \geq \langle c, \theta \rangle) \geq \alpha$, we have $\mathbb{P}(\langle c, X \rangle \geq r) \geq \alpha$ for all $r < \langle c, \theta \rangle$ and so (13) is true. Therefore, we obtain $(A^\alpha)^\circ = K^\alpha(X)$.

For each $\beta \in (\alpha, 1/2)$, we clearly have $\tilde{K}^\alpha(X) \subset A^\beta$. Therefore, we have

$$\bigcup_{\alpha < \beta < 1/2} K^\beta(X) \subset \left(\tilde{K}^\alpha(X) \right)^\circ \subset K^\alpha(X),$$

which is the desired assertion. □

We are going to prove the extension of Theorem 3 by finding a finite set of points whose convex hull approximates $K^\alpha(X)$. The following statement is essentially well-known [2, 34], but we give the precise statement and a brief proof for completeness.

Proposition 23 *Let K be a compact and convex subset of $\mathbb{R}^d$ such that $K = -K$. Then, for each $\varepsilon \in (0, 1)$, there is a finite set $A \subset \mathbb{R}^d$ such that*

$$(1 - \varepsilon)K \subset \operatorname{conv} A \subset K, \qquad |A| \leq \left(1 + \frac{2}{\varepsilon} \right)^d.$$

Proof We can only consider the case K has full dimension, i.e., K has a nonempty interior. Then, the Minkowski functional of K (e.g., see [7, IV.1.14])

$$\|\|x\|\| := \inf\{t \mid t \geq 0, \ x \in tK\}$$

defines a norm on $\mathbb{R}^d$ (note that all norms are equivalent on $\mathbb{R}^d$). For this norm, it is known that there is a finite subset $A \subset S$ such that $\min_{y \in A} \|\|x - y\|\| \leq \varepsilon$ for all $x \in B$ and $|A| \leq (1 + 2/\varepsilon)^d$ [34, Lemma 4.10]. It suffices to prove $(1 - \varepsilon)K \subset \operatorname{conv} A$. Assume the contrary, i.e., let x_0 be a point such that $\|\|x\|\| \leq 1 - \varepsilon$ and $x_0 \notin \operatorname{conv} A$. Then, there exists a $(d - 1)$-dimensional hyperplane $H \subset \mathbb{R}^d$ such that $x_0 \in H$ and all the points in A lie (strictly) on the same side as the origin with respect to H. Let $y \in \operatorname{argmin}_{x \in H} \|\|x\|\|$. Then, we have $\|\|y\|\| \leq 1 - \varepsilon$, and $z := \|\|y\|\|^{-1} y$ satisfies $\min_{x \in H} \|\|z - x\|\| \geq \varepsilon$. Hence, we have $\min_{x \in A} \|\|z - x\|\| > \varepsilon$ and it contradicts the assumption for A. □

Theorem 24 *Let X be an arbitrary symmetric d-dimensional random vector, and let $\alpha, \delta, \varepsilon \in (0, 1)$. If a positive integer n satisfies*

$$n \geq \frac{2d}{\alpha} \max \left\{ \frac{\log(1/\delta)}{d} + \log \frac{1}{\varepsilon}, \ 6 \right\},$$

then we have, with probability at least $1 - \delta$,

$$\operatorname{conv}\{X_1, \ldots, X_n\} \supset (1 - \varepsilon) K^\alpha(X),$$

where $X_1, X_2, \ldots$ are independent copies of X.

Proof As $K^\alpha(X)$ is symmetric and convex, there is a set $A \subset K^\alpha(X)$ with cardinality at most $(1 + 2/\varepsilon)^d$ such that $(1 - \varepsilon)K^\alpha(X) \subset \operatorname{conv} A$ from Proposition 23. We shall evaluate the probability of $A \subset \operatorname{conv}\{X_i\}_{i=1}^n$. As each point $\theta \in A$ satisfies $\alpha_X(\theta) \geq \alpha$, from Remark 3, we have

$$1 - p_{n,X}(\theta) \leq \left(\frac{n\alpha}{d} \exp\left(1 + \alpha - \frac{n\alpha}{d}\right) \right)^d \tag{14}$$

for each $\theta \in A$. Hence, it suffices to prove the right-hand side of (14) is bounded by $(1 + 2/\varepsilon)^{-d} \delta$. By taking the logarithm, it is equivalent to showing

$$\frac{n\alpha}{d} - \log \frac{n\alpha}{d} \geq 1 + \alpha + \frac{\log(1/\delta)}{d} + \log\left(1 + \frac{2}{\varepsilon}\right).$$

Let us denote $x := n\alpha/d$. For $x \geq 12$, as $x/2 - \log x$ is increasing, we have

$$\frac{x}{2} - \log x \geq 6 - \log 6 \geq 2 + \log 3 \geq 1 + \alpha + \log 3$$

by a simple computation. Therefore, from $\log(1 + 2/\varepsilon) \leq \log 3 + \log(1/\varepsilon)$ and the assumption for n, we obtain the inequality (14). $\qquad \square$

Remark 8 Although the bound given in Theorem 24 requires $n \geq 12d/\alpha$, it can be loosened for moderate δ and ε. For example, if we want to obtain a bound for the case $\delta = \varepsilon = 1/2$, then we can prove $n \geq 5d/\alpha$ to be sufficient by using the bound in Proposition 13. Moreover, we should note that we have used the assumption that X is symmetric only for assuring that $K^\alpha(X)$ is symmetric so that we can use Proposition 23. If we take a symmetric convex subset $K \subset K^\alpha(X)$, we can prove a similar inclusion statement for K even for a nonsymmetric X.

If we want a generalized version of Theorem 3, we can prove the following:

Corollary 25 *Let X be an arbitrary d-dimensional symmetric random vector. Let $\beta \in (0, 1)$ and set $\alpha = (en/d)^{-\beta}$. Then, there exists an absolute constant $c > 0.45$ such that, for each integer n satisfying $n \geq (12e^\beta)^{1/(1-\beta)}d$, we have*

$$\operatorname{conv}\{X_1, \ldots, X_n\} \supset \frac{1}{2} K^\alpha(X)$$

with probability at least $1 - \exp(-ce^{-\beta} n^{1-\beta} d^\beta)$, where $X_1, X_2, \ldots$ are independent copies of X.

Proof For $\alpha = (en/d)^{-\beta}$, we have

$$\frac{\alpha}{12d} n = \frac{1}{12e^\beta} \left(\frac{n}{d}\right)^{1-\beta},$$

so $n \geq 12d/\alpha$ is equivalent to $n \geq (12e^\beta)^{1/(1-\beta)}d$. Hence, from Theorem 24, it suffices to determine how small δ can be taken so as to satisfy

$$n \geq \frac{2d}{\alpha} \left(\frac{\log(1/\delta)}{d} + \log 2\right).$$

As $n \geq 12d$ holds for all β, for $a := \frac{\log 2}{6} < 0.1$, we have $an \geq \frac{2d}{\alpha} \log 2$. Therefore, we can take δ as small as

$$\log(1/\delta) = \frac{\alpha}{2}(1-a)n = \frac{1-a}{2} e^{-\beta} n^{1-\beta} d^\beta.$$

Therefore, we can take $c = \frac{1-a}{2} > 0.45$ as desired. $\square$

6 Application

We discuss implications of the results of this paper in two parts. The first part discusses the use of the bounds we gave on $p_{n,X}$, while the second part gives implication of N_X's bounds on the randomized cubature construction.

6.1 Bounds of $p_{n,X}$

Firstly, the inequality between $p_{n,X}$ and $p_{m,X}$ given in Proposition 7 provides the inequality

$$p_{2d,X} \geq \frac{2^d \sqrt{d}}{d+1} p_{d+1,X} \tag{15}$$

as it is mentioned in Remark 2.

Measure reduction Consider a discrete (probability) measure $\mu = \sum_{x \in \mathcal{X}} w_x \delta_x$ for a finite subset of $\mathcal{X} \subset \mathbb{R}^d$. In [8], randomized algorithms for constructing a convex combination satisfying $\mathbb{E}_{X \sim \mu}[X] = \sum_{i=1}^{d+1} \lambda_i x_i$ $(x_i \in \mathcal{X})$, whose existence is assured by Tchakaloff's theorem [3, 40], are considered. As a basic algorithm, the authors consider the following scheme:

(a.1) Randomly choose d points $A = \{x_1, \ldots, x_d\}$ from $\mathcal{X}$.
(a.2) For each $x \in \mathcal{X} \subset A$, determine if $\mathbb{E}_{X \sim \mu}[X] \in \text{conv}(A \cup \{x\})$ or not, and finish the algorithm and return $A \cup \{x\}$ if it holds.
(a.3) Go back to (a.1).

Although we can execute the decision for each x in (a.2) with $\mathcal{O}(d^2)$ computational cost with an $\mathcal{O}(d^3)$ preprocessing for a fixed A, the overall expected computational cost

until the end of the algorithm is at least $\Omega\left(d^2/p_{d+1,X}\right)$ under some natural assumption on μ (see Proposition 9).

However, we can also consdier the following naive procedure:

(b.1) Randomly choose $2d$ points $B = \{x_1, \ldots, x_{2d}\}$ from $\mathcal{X}$.
(b.2) Return B if $\mathbb{E}_{X \sim \mu}[X] \in \text{conv } B$, and go back to (b.1) if not.

By using an LP solver with the simplex method we can execute (b.2) in (empirically) $\mathcal{O}(d^3)$ time [33, 38]. Hence the overall computational cost can be heuristically bounded above by $\mathcal{O}(d^3/p_{2d,X})$, which is faster than the former by $\Omega\left(d^{-3/2}2^d\right)$ from the evaluation in (15). Note also that we have rigorously polynomial bounds via other LP methods (e.g., an infeasible-interior-point method [30]), and so the latter scheme is preferable even in worst-case when the dimension d becomes large.

Relation between two depths We can also deduce an inequality between two depth concepts in statistics. As is mentioned in Introduction, for a random vector $X \in \mathbb{R}^d$, $p_{d+1,X}$ is called the simplicial depth whereas α_X is the Tukey depth of the origin with respect to X.

Naively, we have $\alpha_X \geq \frac{p_{n,X}}{n}$ for each n, so $\alpha_X \geq \frac{p_{d+1,X}}{d+1}$ holds. However, by using (15) here, we obtain a sharper estimate

$$\alpha_X \geq \frac{p_{2d,X}}{2d} \geq \frac{1}{2d}\frac{2^d\sqrt{d}}{d+1}p_{d+1,X} \geq \frac{2^{d-1}}{\sqrt{d}(d+1)}p_{d+1,X}.$$

In contrast, deriving a nontrivial upper bound of α_X in terms of $p_{d+1,X}$ still seems difficult.

6.2 Bounds of N_X

Secondly, we give applications of the bounds of N_X given in Sect. 4.
Random trigonometric cubature Consider a d-dimensional random vector

$$X = (\cos\theta, \ldots, \cos d\theta)^\top \in \mathbb{R}^d$$

for a positive integer d, where θ is a uniform random variable over $(-\pi, \pi)$. Then, from an easy computation, we have $V := \mathbb{E}\left[XX^\top\right] = \frac{1}{2}I_d$, and so we obtain

$$\|V^{-1/2}X\|^2 \leq 2d$$

almost surely. Therefore, from Proposition 17, we have

$$N_X \leq 1 + 12d^2.$$

This example is equivalent to a random construction of the so-called Gauss–Chebyshev quadrature [28, Chapter 8]. Although we can bound as above the number of observations required in a random construction, concrete constructions with fewer points are already known.

Deriving a bound for random construction of cubature without any know deterministic construction, such as cubature on Wiener space [18, 25], which is more important, is still unsolved and left for future work.

Beyond naive cubature construction Recall the cubature construction problem described in Sect. 1.1. We consider a random variable of the form $X = \boldsymbol{f}(Y)$, where Y is a random variable on some topological space $\mathcal{X}$ and $\boldsymbol{f} = (f_1, \ldots, f_d)^\top : \mathcal{X} \to \mathbb{R}^d$ is a d-dimensional vector valued integrable function. Our aim is to find points $y_1, \ldots, y_{d+1} \in \mathcal{X}$ and weights $w_1, \ldots, w_{d+1} \geq 0$ whose total is one such that

$$\mathbb{E}[\boldsymbol{f}(Y)] = \sum_{j=1}^{d+1} w_j \boldsymbol{f}(y_j). \tag{16}$$

A naive algorithm proposed by [17] was to generate independent copies $Y_1, Y_2, \ldots$ of Y and choose y_j from these random samples. Without any knowledge of N_X, the algorithm would be of the form

(c.1) Take $k = 2d$.
(c.2) Randomly generate Y_i up to $i = k$ and determine if (16) can be satisfied with $y_j \in \{Y_i\}_{i=1}^k$ by using an LP solver.
(c.3) If we find a solution, stop the algorithm. Otherwise, go to (c.2) after replacing k by $2k$.

This procedure ends at $k \leq 2N_X(\mathbb{E}[X])$ with probability more than half. We can then heuristically estimate the computational cost by $\Theta(C(d, N_X(\mathbb{E}[X])))$, where we denote by $C(d, n)$ the computational complexity of a linear programming problem finding the solution of (16) from n sample points. Empirically, this is estimated as $\Omega(d^2 n)$ or more when we use the simplex method [38].

However, our analysis on N_X via the Berry–Esseen bound tells us the possibility of an alternative (Algorithm 1).

Although the pseudocode may seem a little long, this is just uses ℓd random vectors of the form $n^{-1}(X_1 + \cdots + X_n)$ as the possible vertices of the convex combination, which is used for deriving bounds of N_X in Sect. 4. After executing Algorithm 1, we can use any algorithm for deterministic measures (typically called recombination; [22, 26, 41]) to obtain an actual $d + 1$ points cubature rule, whose time complexity is rigorously bounded by $\mathcal{O}(kd^3 + 2^k d^2)$ by using the final value of k in the above algorithm.

As we can carry out Algorithm 1 within $\mathcal{O}(2^k \ell d^2 + kC(d, \ell d))$, the overall computational cost is $\mathcal{O}(kC(d, \ell d) + 2^k \ell d^2)$. Then we heuristically have the bound $\mathcal{O}(k\ell d^3 + 2^k \ell d^2)$ for a small ℓ. By using the number $N = 2^k \ell d$, which is the number of randomly generated copies of Y, this cost is rewritten as

$$\mathcal{O}\left(\log(N/\ell d)\ell d^3 + Nd\right).$$

As our bound for $N_X(\mathbb{E}[X])$ in Theorem 19 is applicable for this N because of the use of Berry–Esseen type estimate ($\ell = 17$ is used in the proof), we can also give an estimate for this alternative algorithm. If the N is not as large as $\Omega(dN_X(\mathbb{E}[X]))$ for

---

Algorithm 1 Randomized cubature construction for recombination

---

Input: An integer $\ell \geq 2$
Output: $(w_1, y_1), \ldots, (w_n, y_n) \in \mathbb{R}_{\geq 0} \times \mathcal{X}$ satisfying $\sum_{j=1}^{n} w_j = 1$ and $\mathbb{E}[X] = \sum_{j=1}^{n} w_j f(y_j)$
1: **Initialize:**
2:   $x_1, \ldots, x_{\ell d}, z_1, \ldots, z_{\ell d}$: vectors in $\mathbb{R}^d, k \leftarrow 0$
3: **for** $i = 1, \ldots, \ell d$ **do**
4:   Sample Y_i
5:   $x_i \leftarrow f(Y_i)$
6: **end for**
7: **while** $\mathbb{E}[X] \notin \text{conv}\{x_1, \ldots, x_{\ell d}\}$ **do**
8:   **for** $i = 1, \ldots, \ell d$ **do**
9:     $z_i \leftarrow 0$ (as an $\mathbb{R}^d$ vector)
10:   **end for**
11:   **for** $j = 2^k, \ldots, 2^{k+1} - 1$ **do**
12:     **for** $i = 1, \ldots, \ell d$ **do**
13:       Sample $Y_{j\ell d+i}$
14:       $z_i \leftarrow z_i + 2^{-k} f(Y_{j\ell d+i})$
15:     **end for**
16:   **end for**
17:   **for** $i = 1, \ldots, \ell d$ **do**
18:     $x_i \leftarrow (x_i + z_i)/2$
19:   **end for**
20:   $k \leftarrow k + 1$
21: **end while**
22: Take $x_{i_1}, \ldots, x_{i_{d+1}}$ and $\lambda_1, \ldots, \lambda_{d+1}$ such that $\mathbb{E}[X] = \sum_{m=1}^{d+1} \lambda_m x_{i_m}$ by solving an LP
23: **Return** $(2^{-k} \lambda_m, Y_{j\ell d+i_m})$ for $(j, m) \in \{0, \ldots, 2^k - 1\} \times \{1, \ldots, d+1\}$

---

an appropriate choice of ℓ, we indeed have a better scheme, though the comparison itself may be a nontrivial problem in general. In any event, the fact that we can avoid solving a large LP problem is an obvious advantage.

7 Concluding remarks

In this paper, we have investigated inequalities regarding $p_{n,X}$, N_X and α_X, which is motivated from the fields of numerical analysis, data science, statistics and random matrix. We generalized the existing inequalities for $p_{n,X}$ in Sect. 2. After pointing out that the convergence rate of $p_{n,X}$ is determined by α_X in Sect. 3 with introduction of ε-relaxation of both quantities, we proved that N_X and $1/\alpha_X$ are of the same magnitude up to an $\mathcal{O}(d)$ factor in Theorem 16. We also gave estimates of N_X based on the moments of X in Sect. 4 by using Berry–Esseen type bounds. Although arguments have been based on whether a given vector is included in the random convex polytope $\text{conv}\{X_1, \ldots, X_n\}$, in Sect. 4, we extended our results to the analysis of deterministic convex bodies included in the random convex hull, which immediately led to a technical improvement on a result from the random matrix community. We finally discussed several implications of our results on application in Sect. 6.

A Bounds of N_X via Multivariate Berry–Esseen theorem

In this section, we provide two different estimates of N_X. Although we can prove that the first bound (Sect. A.2) is strictly stronger than the second one (Sect. A.3), we also give the proof of the second as there seems to be more room for improvement in the second approach than in the first.

The following first bound is the one mentioned in (11). The proof is given in Sect. A.2.

Theorem 26 *Let X be an $\mathbb{R}^d$-valued random vector which is centered and of nonsingular covariance matrix V. Then,*

$$N_X \leq 8d \left(1 + 36d^2 (42d^{1/4} + 16)^2 \mathbb{E}\left[\left\| V^{-1/2} X \right\|_2^3 \right]^2 \right)$$

holds.

Note that

$$\mathbb{E}\left[\left\| V^{-1/2} X \right\|_2^3 \right]^2 \geq \mathbb{E}\left[\left\| V^{-1/2} X \right\|_2^2 \right]^3 = d^3$$

holds so we can ignore the $\mathcal{O}(d)$ term. In the case $\sup \left\| V^{-1/2} X \right\|_2 < \infty$, we have

$$\mathbb{E}\left[\left\| V^{-1/2} X \right\|_2^3 \right]^2 \leq \mathbb{E}\left[\left\| V^{-1/2} X \right\|_2^2 \sup \left\| V^{-1/2} X \right\|_2 \right]^2 = d^2 \sup \left\| V^{-1/2} X \right\|_2^2 .$$

Therefore, the following proposition, which only states $N_X = \tilde{\mathcal{O}}\left(d^{15/2} \sup \left\| V^{-1/2} X \right\|_2^2 \right)$, is weaker than Theorem 26. However, the approach of proofs is different and there seems to remain some room for improvement in the proof of Proposition 27, so we give the proof in Sect. 1.

Proposition 27 *Let X be an $\mathbb{R}^d$-valued random vector which is centered, bounded and of nonsingular covariance matrix V. Then, for all n satisfying*

$$\frac{n}{(1 + \log n)^2} \leq 2^{16} 100 d^{13/2} \sup \left\| V^{-1/2} X \right\|_2^2 ,$$

$N_X \leq 6dn$ holds.

A.1 Multivariate Berry–Esseen bounds

Before proceeding to the evaluation of N_X, we briefly review multivariate Berry–Esseen type theorems. The following theorem should be the best known bound with explicit constants and dependence with respect to the dimension.

Theorem 28 [35] *Let $Y_1, \ldots, Y_n$ be i.i.d. D-dimensional independent random vectors with mean zero and covariance I_D. For any convex measurable set $A \subset \mathbb{R}^D$, it holds*

$$\left| \mathbb{P}\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \in A \right) - \mathbb{P}(Z \in A) \right| \leq \frac{(42D^{1/4} + 16)\mathbb{E}\big[\|Y_1\|_2^3\big]}{\sqrt{n}},$$

where Z is a D-dimensional standard Gaussian.

Note that the original statement is not limited to the i.i.d. case. However, similarly to the other existing Berry–Esseen type bounds, Theorem 28 only gives information about *convex* measurable sets. Thus we cannot use this result directly. However, Sect. A.2 gives a creative use of Theorem 28.

Unlike the usual Berry–Esseen results, the next theorem can be used for nonconvex case with reasonable dependence on dimension. We denote by $\mathcal{W}_2(\mu, \nu)$ the Wasserstein-2 distribution between two probability measures μ and ν on the same domain. This is defined formally as

$$\mathcal{W}_2(\mu, \nu) := \inf_{Y \sim \mu, Z \sim \nu} \mathbb{E}\Big[\|Y - Z\|_2^2\Big],$$

where the infimum is taken for all the joint distribution (Y, Z) with the marginal satisfying $Y \sim \mu$ and $Z \sim \nu$. Although it is an abuse of notation, we also write $\mathcal{W}_2(Y, Z)$ to represent $\mathcal{W}_2(\mu, \nu)$ when $Y \sim \mu$ and $Z \sim \nu$ for some random variables Y and Z.

Theorem 29 [46] *Let $Y_1, \ldots, Y_n$ be D-dimensional independent random vectors with mean zero, covariance Σ, and $\|Y_i\|_2 \leq B$ almost surely for each i. If we let Z be a Gaussian with covariance Σ, then we have*

$$\mathcal{W}_2\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}, Z \right) \leq \frac{5\sqrt{D}B(1 + \log n)}{\sqrt{n}}.$$

For a set $A \subset \mathbb{R}^D$ and an $\varepsilon > 0$, define

$$A^\varepsilon := \left\{ x \in \mathbb{R}^D \,\middle|\, \inf_{y \in A} \|x - y\|_2 \leq \varepsilon \right\}, \qquad A^{-\varepsilon} := \left\{ x \in \mathbb{R}^D \,\middle|\, \inf_{y \in A^c} \|x - y\|_2 \geq \varepsilon \right\}.$$

By combining the following assertion with Theorem 29, we derive another bound of N_X in Sect. 1.

Proposition 30 *Let Y, Z be D-dimensional random vectors. Then, for any measurable set $A \subset \mathbb{R}^d$ and any $\varepsilon > 0$, the following estimates hold:*

$$\mathbb{P}(Y \in A) \leq \mathbb{P}(Z \in A^\varepsilon) + \frac{\mathcal{W}_2(Y, Z)^2}{\varepsilon^2},$$

$$\mathbb{P}(Y \in A) \geq \mathbb{P}(Z \in A^{-\varepsilon}) - \frac{\mathcal{W}_2(Y, Z)^2}{\varepsilon^2}.$$

Proof This proof is essentially the same as the argument given in the proof of [46, Proposition 1.4]. Let (Y', Z') be an arbitrary couple of random variables such that $Y' \sim Y$ and $Z' \sim Z$. Then, we have

$$\begin{aligned}
\mathbb{P}(Y' \in A) &= \mathbb{P}(\|Y' - Z'\|_2 < \varepsilon, \ Y' \in A) + \mathbb{P}(\|Y' - Z'\|_2 \geq \varepsilon, \ Y \in A) \\
&\leq \mathbb{P}(Z' \in A^\varepsilon) + \mathbb{P}(\|Y' - Z'\|_2 \geq \varepsilon) \\
&\leq \mathbb{P}(Z' \in A^\varepsilon) + \frac{1}{\varepsilon^2}\mathbb{E}\left[\|Y' - Z'\|_2^2\right]. \quad \text{(by Chebyshev's inequality)}
\end{aligned}$$

By taking the infimum of the right-hand side with respect to all the possible couples (Y', Z'), we obtain the former result. The latter can also be derived by evaluating

$$\begin{aligned}
\mathbb{P}(Z' \in A^{-\varepsilon}) &= \mathbb{P}(\|Y' - Z'\|_2 < \varepsilon, \ Z' \in A^{-\varepsilon}) + \mathbb{P}(\|Y' - Z'\|_2 \geq \varepsilon, \ Z \in A^{-\varepsilon}) \\
&\leq \mathbb{P}(Y' \in A) + \mathbb{P}(\|Y' - Z'\|_2 \geq \varepsilon) \\
&\leq \mathbb{P}(Y' \in A) + \frac{1}{\varepsilon^2}\mathbb{E}\left[\|Y' - Z'\|_2^2\right]
\end{aligned}$$

and again taking the infimum. $\qquad\square$

A.2 The first bound

In this section, we prove Theorem 26. We shall set $D = d$ and make use of Theorem 28.

First, fix a set $S \subset \mathbb{R}^d$ and consider the set $C(S) := \{x \in \mathbb{R}^d \mid 0 \in \text{conv}(S \cup \{x\})\}$. We can prove this set is convex for any S. Indeed, if $0 \in \text{conv}\, S$, then clearly $C(S) = \mathbb{R}^d$. Otherwise, $x \in C(S)$ is equivalent to the existence of some $k \geq 0$ and $x_1, \ldots, x_k \in S$, $\lambda > 0$, $\lambda_1, \ldots, \lambda_k \geq 0$ such that

$$\lambda + \lambda_1 + \cdots + \lambda_k = 1, \qquad \lambda x + \lambda_1 x_1 + \cdots + \lambda_k x_k = 0.$$

Here, $\lambda > 0$ comes from the assumption $0 \notin \text{conv}\, S$. This occurs if and only if x is contained in the negative cone of S, i.e., $C(S) = \{\sum_{i=1}^k \tilde{\lambda}_i x_i \mid k \geq 0, \ \tilde{\lambda}_i \leq 0, \ x_i \in S\}$. In both cases $C(S)$ is convex, so S_0 is always convex (and of course measurable).

Let X be an $\mathbb{R}^d$-valued random vector with mean 0 and nonsingular covariance V. Suppose $\mathbb{E}\left[\|V^{-1/2}X\|_2^3\right] < \infty$. Let $X_1, X_2, \ldots$ be independent copies of X, and for

a fixed positive integer n, define

$$W_i := \frac{V^{-1/2}X_{(i-1)n+1} + \cdots + V^{-1/2}X_{in}}{\sqrt{n}}$$

for $i = 1, \ldots, 2d$. We also let $Z_1, \ldots, Z_{2d}$ be independent d-dimensional standard Gaussian which is also independent from $X_1, X_2, \ldots$. Then, by using Theorem 28 and the above-mentioned convexity of $C(S)$, we have

$$\mathbb{P}(0 \in \{W_1, \ldots, W_{2d}\}) = \mathbb{P}(W_1 \in C(\{W_2, \ldots, W_{2d}\}))$$

$$\geq \mathbb{P}(Z_1 \in C(\{W_2, \ldots, W_{2d}\})) - \frac{(42d^{1/4} + 16)\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]}{\sqrt{n}}$$

$$= \mathbb{P}(0 \in \mathrm{conv}\{Z_1, W_2, \ldots, W_{2d}\}) - \frac{(42d^{1/4} + 16)\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]}{\sqrt{n}}.$$

By repeating similar evaluations, we obtain

$$\mathbb{P}(0 \in \mathrm{conv}\{W_1, \ldots, W_{2d}\})$$

$$\geq \mathbb{P}(0 \in \mathrm{conv}\{Z_1, W_2, \ldots, W_{2d}\}) - \frac{(42d^{1/4} + 16)\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]}{\sqrt{n}}$$

$$\geq \mathbb{P}(0 \in \mathrm{conv}\{Z_1, Z_2, W_3, \ldots, W_{2d}\}) - \frac{2(42d^{1/4} + 16)\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]}{\sqrt{n}}$$

$$\vdots$$

$$\geq \mathbb{P}(0 \in \mathrm{conv}\{Z_1, \ldots, Z_i, W_{i+1}, \ldots, W_{2d}\}) - \frac{i(42d^{1/4} + 16)\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]}{\sqrt{n}}$$

$$\vdots$$

$$\geq \mathbb{P}(0 \in \mathrm{conv}\{Z_1, \ldots, Z_{2d}\}) - \frac{2d(42d^{1/4} + 16)\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]}{\sqrt{n}}$$

$$= \frac{1}{2} - \frac{2d(42d^{1/4} + 16)\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]}{\sqrt{n}}.$$

Therefore, by letting

$$n = \left\lceil 36d^2(42d^{1/4} + 16)^2\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]^2 \right\rceil,$$

we have $\mathbb{P}(0 \in \mathrm{conv}\{X_1, \ldots, X_{2dn}\}) \geq 1/6$. Since $(1 - 1/6)^4 < 1/2$ holds, we finally obtain $N_X \leq 8dn$.

A.3 The second bound

In this section, we provide a proof of Sect. 27 in a different manner from the one given in the previous section. We set $D = 2d^2$ and define $A_d \subset \mathbb{R}^D$ as follows:

$$A_d := \{x = (x_1, \ldots, x_{2d}) \in (\mathbb{R}^d)^{2d} \simeq \mathbb{R}^D \mid 0 \in \mathrm{conv}\{x_1, \ldots, x_{2d}\} \subset \mathbb{R}^d\}.$$

Then, it suffices to find a suitable upper bound of $\mathbb{P}(Z \in A_d \backslash A_d^{-\varepsilon})$ for a D-dimensional standard Gaussian Z for our purpose. For an $\varepsilon > 0$, $B_{d,\varepsilon} := A_d \backslash A_d^{-\varepsilon}$ can be explicitly written as

$$B_{d,\varepsilon} = \left\{ x = (x_1, \ldots, x_{2d}) \in \mathbb{R}^D \; \middle| \; \begin{array}{l} 0 \in \mathrm{conv}\{x_i\}_{i=1}^{2d}, \\ \exists \tilde{x} = (\tilde{x}_i)_{i=1}^{2d} \in \mathbb{R}^D \text{ s.t.} \|x - \tilde{x}\|_2 < \varepsilon, \; 0 \notin \mathrm{conv}\{\tilde{x}_i\}_{i=1}^{2d} \end{array} \right\}. \tag{17}$$

For a (finite) set $S = \{v_1, \ldots, v_j\} \subset \mathbb{R}^d$, define the negative box $N(S) \subset \mathbb{R}^d$ by

$$N(S) := \{a_1 v_1 + \cdots + a_j v_j \mid a_i \in [-1, 0]\}.$$

$N(S)$ is obviously a convex set.

Lemma 31 *For an arbitrary $x = (x_1, \ldots, x_{2d}) \in B_{d,\varepsilon}$, there exists an index $k \in \{1, \ldots, 2d\}$ such that $x_k \in N(\{x_i \mid i \neq k\}) \backslash N(\{x_i \mid i \neq k\})^{-\varepsilon\sqrt{2d}}$.*

Proof As $0 \in \mathrm{conv}\{x_i\}_{i=1}^{2d}$, there exist nonnegative weights $\lambda_1, \ldots, \lambda_{2d}$ such that $\lambda_1 x_1 + \cdots + \lambda_{2d} x_{2d} = 0$ with the total weight one. Let k be an index such that w_k is the maximum weight. Then, λ_k is clearly positive and we have $x_k = \sum_{i \neq k} -\frac{\lambda_i}{\lambda_k} x_i$. Therefore, we obtain $x_k \in N(\{x_i \mid i \neq k\})$.

By (17), there exists an $\tilde{x} = (\tilde{x}_i)_{i=1}^{2d} \in \mathbb{R}^D$ such that $\sum_{i=1}^{2d} \|x_i - \tilde{x}_i\|_2^2 < \varepsilon^2$ and $0 \notin \mathrm{conv}\{\tilde{x}_i\}_{i=1}^{2d}$. We can prove that $\tilde{x}_k \notin N(\{\tilde{x}_i \mid i \neq k\})$. Indeed, if we can write $\tilde{x}_k = -\sum_{i \neq k} a_i \tilde{x}_i$ with $a_i \in [0, 1]$, then

$$\left(1 + \sum_{i \neq k} a_i\right)^{-1} \left(\tilde{x}_k + \sum_{i \neq k} a_i \tilde{x}_i\right) = 0$$

is a convex combination and it contradicts the assumption $0 \notin \mathrm{conv}\{\tilde{x}_i\}_{i=1}^{2d}$. Therefore, we can take a unit vector $c \in \mathbb{R}^d$ such that

$$c^\top \tilde{x}_k > \max\{c^\top y \mid y \in N(\{\tilde{x}_i \mid i \neq k\})\}. \tag{18}$$

Let us assume the closed ball with center x_k and radius δ is included in $N(\{x_i \mid i \neq k\})$ for a $\delta > 0$. Then, if $\delta > \|x_k - \tilde{x}_k\|_2$, the closed ball with center $\tilde{x}_k$ and radius

$\delta' := \delta - \|x_k - \tilde{x}_k\|_2$ is included in $N(\{x_i \mid i \neq k\})$. In particular, we have some coefficients $a_i \in [-1, 0]$ such that $\tilde{x}_k + \delta'c = \sum_{i \neq k} a_i x_i$. By the inequality (18), we have

$$c^\top \tilde{x}_k > c^\top \sum_{i \neq k} a_i \tilde{x}_i = c^\top \left(\tilde{x}_k + \delta'c + \sum_{i \neq k} a_i (\tilde{x}_i - x_i) \right),$$

so by arranging

$$\delta' < \sum_{i \neq k} a_i c^\top (x_i - \tilde{x}_i) \leq \sum_{i \neq k} \|x_i - \tilde{x}_i\|_2.$$

Therefore, from the definition of δ', we obtain

$$\delta < \sum_{i=1}^{2d} \|x_i - \tilde{x}_i\|_2 \leq \left(2d \sum_{i=1}^{2d} \|x_i - \tilde{x}_i\|_2 \right)^{1/2} \leq \varepsilon \sqrt{2d}$$

by Cauchy-Schwarz and the assumption. It immediately implies the desired assertion.
$\square$

Proposition 32 $\mathbb{P}(Z \in B_{d,\varepsilon}) \leq 8\sqrt{2}d^{7/4}\varepsilon$ holds.

Proof By Lemma 31, we have $B_{d,\varepsilon} \subset \bigcup_{k=1}^{2d}\{x \mid x_k \in N(\{x_i \mid i \neq k\})\backslash N(\{x_i \mid i \neq k\})^{-\varepsilon\sqrt{2d}}\}$. Therefore, letting $Z = (Z_1, \ldots, Z_{2d})$ be a standard Gaussian in $\mathbb{R}^D$ (where each Z_i is a independent standard Gaussian in $\mathbb{R}^d$), we can evaluate

$$\mathbb{P}(Z \in B_{d,\varepsilon}) \leq \sum_{k=1}^{2d} \mathbb{P}\left(Z_k \in N(\{Z_i \mid i \neq k\})\backslash N(\{Z_i \mid i \neq k\})^{-\varepsilon\sqrt{2d}}\} \right).$$

For each k, Z_k is independent from the random convex set $N(\{Z_i \mid i \neq k\})$. Therefore, we can use the result of [1] to deduce $\mathbb{P}\left(Z_k \in N(\{Z_i \mid i \neq k\})\backslash N(\{Z_i \mid i \neq k\})^{-\varepsilon\sqrt{2d}}\} \right)$
$\leq 4d^{1/4} \cdot \varepsilon\sqrt{2d}$. Therefore, we finally obtain

$$\mathbb{P}(Z \in B_{d,\varepsilon}) \leq 2d \cdot 4d^{1/4} \cdot \varepsilon\sqrt{2d} = 8\sqrt{2}d^{7/4}\varepsilon.$$

$\square$

By letting $\varepsilon = 2^{-13/2}d^{-7/4}$, we have $\mathbb{P}(Z \in B_{d,\varepsilon}) \leq 1/8$. Under this value of ε, if we let n satisfy

$$\frac{n}{(1 + \log n)^2} \geq \frac{8 \cdot 25DB^2}{\varepsilon^2} = 400d^2B^2 \cdot 2^{13}d^{7/2} = 2^{15}100B^2d^{11/2}, \qquad (19)$$

for a constant B, then we have

$$\left(\frac{5\sqrt{D}B(1+\log n)}{\sqrt{n}}\right)^2 \le \frac{\varepsilon^2}{8}.$$

Now consider a bounded and centered $\mathbb{R}^d$-valued random vector X with $V = \mathbb{E}[XX^\top]$ nonsingular. Then $B' := \sup\|V^{-1/2}X\|_2$ is finite. Let $X_1, X_2, \ldots$ be independent copies of X. Define $\mathbb{R}^D$-valued random vectors $Y_1, Y_2, \ldots$ by $Y_i := (V^{-1/2}X_{(2i-1)d+1}, \ldots, V^{-1/2}X_{2id})^\top$ for each i. Then, note that $\|Y_i\|_2 \le \sqrt{2d}B'$. By taking $B = \sqrt{2d}B'$ in (19), we have from Theorem 29 that (for $\varepsilon = 2^{-13/2}d^{-7/4}$)

$$\mathbb{P}(Z \in B_{d,\varepsilon}) \le \frac{1}{8}, \qquad \frac{1}{\varepsilon^2}\mathcal{W}_2\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}, Z\right) \le \frac{1}{8}.$$

From Proposition 30, we obtain

$$\mathbb{P}\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \in A_d\right) \ge \mathbb{P}(Z \in A_d) - \mathbb{P}(Z \in B_{d,\varepsilon})$$
$$-\frac{1}{\varepsilon^2}\mathcal{W}_2\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}, Z\right) \ge \frac{1}{4}.$$

Therefore, 0 is contained in the convex hull of $\{X_1, \ldots, X_{2dn}\}$ with probability at least $1/4$. Since $(1-1/4)^3 < 1/2$, $N_X \le 6dn$ holds. Therefore, our proof of Proposition 27 is complete.

B Extreme examples

Before treating concrete examples, we prove a proposition which is useful for evaluating N_X.

Lemma 33 *For a random vector X and its independent copies $X_1, X_2, \ldots$, define $\tilde{N}_X$ as the minimum index n satisfying $0 \in \mathrm{conv}\{X_1, \ldots, X_n\}$. Then, we have*

$$\frac{1}{2}\mathbb{E}[\tilde{N}_X] \le N_X \le 2\mathbb{E}[\tilde{N}_X].$$

Proof From the definition of N_X, $\mathbb{P}(0 \in \{X_1, \ldots, X_{N_X-1}\}) < 1/2$ holds. Thus $\mathbb{P}(\tilde{N}_X \ge N_X) \ge 1/2$, and so we obtain $\mathbb{E}[\tilde{N}_X] \ge \frac{1}{2}N_X$.

For the other inequality, we use the evaluation $\mathbb{P}(\tilde{N}_X \ge kN_X) \le 2^{-k}$ for each nonnegative integer k. As $\tilde{N}_X$ is a nonnegative discrete random variable, we have

$$\mathbb{E}[\tilde{N}_X] = \sum_{n=1}^\infty \mathbb{P}(\tilde{N}_X \ge n) \le \sum_{k=0}^\infty N_X \mathbb{P}(\tilde{N}_X \ge kN_X) \le 2N_X.$$

$\square$

Note that all the examples given below satisfy $p_{d,X} = 0$. They are given as one of the worst-case examples for uniform estimates of N_X in Proposition 5 or Theorem 26. Let us start with the simplest extreme case.

Example 34 Let $d = 1$. For an $\varepsilon \in (0, 1)$, let X be a random variable such that $\mathbb{P}(X = 1/\varepsilon) = \varepsilon$ and $\mathbb{P}(X = -1/(1 - \varepsilon)) = 1 - \varepsilon$. Then $\mathbb{E}[X] = 0$.

In this example, we can explicitly calculate $p_{n,X}$ as

$$p_{n,X} = 1 - \varepsilon^n - (1 - \varepsilon)^n.$$

In particular, $p_{2,X} = 2\varepsilon - 2\varepsilon^2$. We have $\lim_{\varepsilon \searrow 0}(1 - \varepsilon)^{1/2\varepsilon} = e^{-1/2} = 0.60\ldots$, so $p_{\lceil 1/2\varepsilon \rceil, X} < 1/2$ holds for a sufficiently small ε. For such an ε, we have

$$N_X \geq \frac{1}{2\varepsilon} = \frac{1 - \varepsilon}{2} \frac{2}{p_{2,X}}, \tag{20}$$

and so $N_X \leq \frac{2}{p_{2,X}}$ in Proposition 5 is sharp up to constant.

For $\varepsilon \in (0, 1/2)$, N_X can also be evaluated above as $N_X \leq 2\mathbb{E}[\tilde{N}_X] \leq 2\left(\frac{1}{\varepsilon} + \frac{1}{(1-\varepsilon)}\right)$ by using Proposition 33. We also have $\alpha_X = \varepsilon$ for $\varepsilon \in (0, 1/2)$, so

$$\inf_{X:\text{1-dimensional}} \alpha_X N_X \leq 2 + \frac{2\varepsilon}{1 - \varepsilon} \to 2 \quad (\varepsilon \to 0).$$

As the variance is $V = \mathbb{E}[X^2] = \frac{1}{\varepsilon} + \frac{1}{1-\varepsilon} = \frac{1}{\varepsilon(1-\varepsilon)}$, we have

$$\mathbb{E}\left[\left|V^{-1/2}X\right|^3\right]^2 = V^{-3}\left(\frac{1}{\varepsilon^2} + \frac{1}{(1 - \varepsilon)^2}\right)^2$$

$$= \varepsilon^3(1 - \varepsilon)^3\left(\frac{1}{\varepsilon^4} + \frac{2}{\varepsilon^2(1 - \varepsilon)^2} + \frac{1}{(1 - \varepsilon)^4}\right)$$

$$= \frac{1}{\varepsilon} + \mathcal{O}(1).$$

Therefore, from (20), we obtain

$$\sup\left\{\mathbb{E}\left[\left|V^{-1/2}X\right|^3\right]^{-2} N_X \,\middle|\, \begin{array}{l} X \text{ is 1-dimensional, } \mathbb{E}[X] = 0, \\ V = \mathbb{E}[X^2] \in (0, \infty), \ \mathbb{E}\left[\left|V^{-1/2}X\right|^3\right] < \infty \end{array}\right\} \geq \frac{1}{2},$$

which is what is mentioned in Remark 6 when $d = 1$.

The next example is a multi-dimensional version of the previous one.

Example 35 Let $d \geq 2$. Let $\{e_1, \ldots, e_d\} \subset \mathbb{R}^d$ be the standard basis of $\mathbb{R}^d$. Let us first consider, for an arbitrary $\varepsilon \in (0, 1)$, a random vector X given by

$$X = Y \left(\sum_{i=1}^{d-1} Z^i e_i - \frac{1}{1-\varepsilon} e_d \right) + \frac{1}{\varepsilon}(1 - Y)e_d,$$

where $\mathbb{P}(Y = 1) = 1 - \varepsilon$, $\mathbb{P}(Y = 0) = \varepsilon$ and $Z^1, \ldots, Z^{d-1}$ are independent uniform random variables over $[-1, 1]$. (also independent from Y). Namely, X is $\varepsilon^{-1} e_d$ with probability ε and a $(d-1)$-dimensional uniform vector over a box on the hyperplane $\{x \in \mathbb{R}^d \mid e_d^\top x = -(1-\varepsilon)^{-1}\}$ otherwise. $\mathbb{E}[X] = 0$ also holds.

Let us estimate $p_{d+1,X}$, $p_{2d,X}$ and N_X for this X. To contain the origin in the convex hull, we have to observe at least one X_i with $Y = 0$. Therefore, for an $\varepsilon \ll 1/d$, we have

$$p_{d+1,X} = (d+1)\varepsilon(1-\varepsilon)^d 2^{-(d-1)} = \frac{d+1}{2^{d-1}} \varepsilon \left(1 + \mathcal{O}\left(d^2\varepsilon^2\right) \right)$$

$$p_{2d,X} = \sum_{k=1}^{d} \binom{2d}{k} \varepsilon^k (1-\varepsilon)^{2d-k} p_{2d-k,X'}$$

$$= 2d\varepsilon \, p_{2d-1,X'} + \mathcal{O}\left(d^2\varepsilon^2\right) = d \left(1 + \frac{1}{2^{2d-2}} \binom{2d-2}{d-1} \right) \varepsilon + \mathcal{O}\left(d^2\varepsilon^2\right)$$

$$\geq d \left(1 + \frac{1}{2\sqrt{d-1}} \right) \varepsilon + \mathcal{O}\left(d^2\varepsilon^2\right),$$

where X' represents a $(d-1)$-dimensional uniform random vector over the box $[-1, 1]^{d-1}$. We can see that $p_{2d,X} \gtrsim 2^{d-1} p_{d+1,X}$ holds for a small ε as Remark 2 suggests.

For the calculation of N_X, we can exploit Proposition 33. We first bound the expectation of $\tilde{N}_X$. For independent copies $X_1, X_2, \ldots$ of X, let N_1 be the minimum integer n satisfying $X_n = \varepsilon^{-1} e_d$. We also define N_2 as the minimum integer n satisfying $-(1-\varepsilon)^{-1} e_d \in \text{conv}\{X_1, \ldots, X_n\}$. Then, $\tilde{N}_X = \max\{N_1, N_2\}$ holds. Thus we have $N_1 \leq \tilde{N}_X \leq N_1 + N_2$. $\mathbb{E}[N_1] = 1/\varepsilon$ clearly holds. For N_2, we can evaluate (again using X') as

$$\mathbb{E}[N_2] = \frac{1}{1-\varepsilon} \mathbb{E}\left[\tilde{N}_{X'}\right] \leq \frac{2N_{X'}}{1-\varepsilon} = \frac{4(d-1)}{1-\varepsilon},$$

where we have used Proposition 33 for the inequality. Therefore, from Proposition 33, we obtain

$$\frac{1}{2\varepsilon} \leq \frac{1}{2}\mathbb{E}\left[\tilde{N}_X\right] \leq N_X \leq 2\mathbb{E}\left[\tilde{N}_X\right] \leq \frac{2}{\varepsilon} + \frac{8(d-1)}{1-\varepsilon}. \tag{21}$$

We finally compare the naive general estimate $N_X \leq \frac{n}{p_{n,X}}$ in Proposition 5 with this example. From (21), we have

$$\frac{N_X p_{2d,X}}{2d} \geq \frac{p_{2d,X}}{4d\varepsilon} \geq \frac{1}{4} + \frac{1}{8\sqrt{d-1}} + \mathcal{O}(d\varepsilon).$$

Therefore, the evaluation $N_X \leq \frac{2d}{p_{2d,X}}$ is sharp even for small $p_{2d,X}$ up to constant in the sense that

$$\lim_{\varepsilon \to 0} \sup_{\substack{X:d\text{-dimensional} \\ p_{2d,X}<\varepsilon}} \frac{N_X p_{2d,X}}{2d} \geq \frac{1}{4} + \frac{1}{8\sqrt{d-1}}$$

holds.

Also in this example, we have $\alpha_X = \varepsilon$ for $\varepsilon \in (0, 1/3)$. Hence, combined with (21), we have

$$\alpha_X N_X \leq \varepsilon \left(\frac{2}{\varepsilon} + \frac{8(d-1)}{1-\varepsilon} \right) = 2 + \frac{8(d-1)\varepsilon}{1-\varepsilon} \to 2 \quad (\varepsilon \to 0).$$

Therefore, we have $\inf_{X:d\text{-dim}} \alpha_X N_X \leq 2$.

We next evaluate the value of $\mathbb{E}\left[\left\| V^{-1/2} X \right\|_2^3 \right]$, where $V = (V^{ij})$ is the covariance matrix of X with respect to the basis $\{e_1, \ldots, e_d\}$. Then, for $(i, j) \in \{1, \ldots, d-1\}^2$, we obtain

$$V^{ij} = \mathbb{E}\left[Y^2 Z^i Z^j\right] = \mathbb{E}\left[Y^2\right] \mathbb{E}\left[Z^i Z^j\right] = \frac{1-\varepsilon}{2}\delta^{ij}, \qquad (\delta^{ij}: \text{Kronecker's delta})$$

$$V^{id} = \mathbb{E}\left[YZ^i \left(-\frac{Y}{1-\varepsilon} + \frac{1-Y}{\varepsilon}\right)\right] = \mathbb{E}\left[Z^i\right] \mathbb{E}\left[Y\left(-\frac{Y}{1-\varepsilon} + \frac{1-Y}{\varepsilon}\right)\right] = 0$$

by using the independence of $Y, Z_1, \ldots, Z_{d-1}$. For the V^{dd}, we have

$$V^{dd} = \frac{1}{1-\varepsilon} + \frac{1}{\varepsilon} = \frac{1}{\varepsilon(1-\varepsilon)}.$$

Therefore, $V^{-1/2} X$ can be explicitly written as

$$V^{-1/2} X = Y \left(\sqrt{\frac{2}{1-\varepsilon}} \sum_{i=1}^{d-1} Z^i e_i - \sqrt{\frac{\varepsilon}{1-\varepsilon}} e_d \right) + \sqrt{\frac{1-\varepsilon}{\varepsilon}}(1-Y)e_d.$$

Thus we have

$$\left\| V^{-1/2} X \right\|_2^2 \leq Y\frac{2(d-1)+\varepsilon}{1-\varepsilon} + (1-Y)\frac{1-\varepsilon}{\varepsilon},$$

and so

$$\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right] \leq \frac{(2(d-1)+\varepsilon)^{3/2}}{\sqrt{1-\varepsilon}} + \frac{(1-\varepsilon)^{3/2}}{\sqrt{\varepsilon}} \leq 4d^{3/2} + \varepsilon^{-1/2}$$

holds when $0 < \varepsilon < 1/2$. By using (21), we obtain

$$\frac{N_X}{\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]^2} \geq \frac{1}{2\varepsilon(4d^{3/2}+\varepsilon^{-1/2})^2} = \frac{1}{2(4d^{3/2}\varepsilon^{1/2}+1)^2}.$$

Therefore, by taking $\varepsilon \to 0$, we finally obtain the estimate

$$\sup\left\{\frac{N_X}{\mathbb{E}\left[\left\|V^{-1/2}X\right\|_2^3\right]^2} \;\middle|\; \begin{array}{l} X \text{ is } d\text{-dimensional, } \mathbb{E}[X]=0, \\ V = \mathbb{E}[X^2] \text{ is nonsingular, } \mathbb{E}\left[\left\|V^{-1/2}X\right\|^3\right] < \infty \end{array}\right\} \geq \frac{1}{2}$$

as mentioned in Remark 6.

References

1. Ball, K.: The reverse isoperimetric problem for Gaussian measure. Discrete Comput. Geom. **10**(4), 411–420 (1993)
2. Barvinok, A.: Thrifty approximations of convex bodies by polytopes. Int. Math. Res. Not. **2014**(16), 4341–4356 (2014)
3. Bayer, C., Teichmann, J.: The proof of Tchakaloff's theorem. Proc. Am. Math. Soc. **134**(10), 3035–3040 (2006)
4. Berry, A.C.: The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Am. Math. Soc. **49**(1), 122–136 (1941)
5. Cascos, I.: Depth functions based on a number of observations of a random vector. Working paper 07–07, Statistics and Econometrics Series, Universidad Carlos III de Madrid (2007) https://hdl.handle.net/10016/700
6. Combettes, C.W., Pokutta, S.: Revisiting the approximate Carathéodory problem via the Frank–Wolfe algorithm. Math. Program. (2021). https://doi.org/10.1007/s10107-021-01735-x
7. Conway, J.B.: A Course in Functional Analysis. Springer, Berlin (2007)
8. Cosentino, F., Oberhauser, H., Abate, A.: A randomized algorithm to reduce the support of discrete measures. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems, vol. 33, pp. 15100–15110, Curran Associates, Inc (2020)
9. Cuesta-Albertos, J.A., Nieto-Reyes, A.: The random Tukey depth. Comput. Stat. Data Anal. **52**(11), 4979–4988 (2008)
10. Dafnis, N., Giannopoulos, A., Tsolomitis, A.: Asymptotic shape of a random polytope in a convex body. J. Funct. Anal. **257**(9), 2820–2839 (2009)
11. Donoho, D.L., Gasko, M.: Breakdown properties of location estimates based on halfspace depth and projected outlyingness. Ann. Stat. **20**(4), 1803–1827 (1992)
12. Esseen, C.G.: On the Liapunoff limit of error in the theory of probability. Arkiv for Matematik, Astronomi och Fysik, A 1–19 (1942)
13. Giannopoulos, A., Hartzoulaki, M.: Random spaces generated by vertices of the cube. Discrete Comput. Geom. **28**(2), 255–273 (2002)
14. Gluskin, E.D.: Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces. Math. USSR-Sbornik **64**(1), 85–96 (1989)
15. Guédon, O., Krahmer, F., Kümmerle, C., Mendelson, S., Rauhut, H.: On the geometry of polytopes generated by heavy-tailed random vectors. Commun. Contemp. Math. **24**(03), 2150056 (2022)
16. Guédon, O., Litvak, A.E., Tatarko, K.: Random polytopes obtained by matrices with heavy-tailed entries. Commun. Contemp. Math. **22**(04), 1950027 (2020)

17. Hayakawa, S.: Monte Carlo cubature construction. Jpn. J. Ind. Appl. Math. **38**, 561–577 (2021)
18. Hayakawa, S., Tanaka, K.: Monte Carlo construction of cubature on Wiener space. Jpn. J. Ind. Appl. Math. **39**(2), 543–571 (2022)
19. Hug, D.: Random polytopes. In: Spodarev, E. (ed.) Stochastic Geometry, Spatial Statistics and Random Fields, pp. 205–238. Springer, Berlin (2013)
20. Kabluchko, Z., Zaporozhets, D.: Absorption probabilities for Gaussian polytopes and regular spherical simplices. Adv. Appl. Probab. **52**(2), 588–616 (2020)
21. Korolev, V., Shevtsova, I.: An improvement of the Berry-Esseen inequality with applications to Poisson and mixed Poisson random sums. Scand. Actuar. J. **2012**(2), 81–105 (2012)
22. Litterer, C., Lyons, T.: High order recombination and an application to cubature on Wiener space. Ann. Appl. Probab. **22**(4), 1301–1327 (2012)
23. Litvak, A., Pajor, A., Rudelson, M., Tomczak-Jaegermann, N.: Smallest singular value of random matrices and geometry of random polytopes. Adv. Math. **195**(2), 491–523 (2005)
24. Liu, R.Y.: On a notion of data depth based on random simplices. Ann. Stat. **18**(1), 405–414 (1990)
25. Lyons, T., Victoir, N.: Cubature on Wiener space. Proc. R. Soc. Lond. Ser. A **460**, 169–198 (2004)
26. Maalouf, A., Jubran, I., Feldman, D.: Fast and accurate least-mean-squares solvers. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32. Curran Associates Inc., Red Hook (2019)
27. Majumdar, S.N., Comtet, A., Randon-Furling, J.: Random convex hulls and extreme value statistics. J. Stat. Phys. **138**(6), 955–1009 (2010)
28. Mason, J.C., Handscomb, D.C.: Chebyshev Polynomials. CRC Press, Boca Raton (2002)
29. Mirrokni, V., Leme, R.P., Vladu, A., Wong, S.C.W.: Tight bounds for approximate Carathéodory and beyond. In: International Conference on Machine Learning, pp. 2440–2448. PMLR (2017)
30. Mizuno, S.: Polynomiality of infeasible-interior-point algorithms for linear programming. Math. Program. **67**(1–3), 109–119 (1994)
31. Mosler, K.: Depth statistics. In: Becker, C., Fried, R., Kuhnt, S. (eds.) Robustness and Complex Data Structures, pp. 17–34. Springer, Berlin (2013)
32. Nagy, S., Schütt, C., Werner, E.M.: Halfspace depth and floating body. Stat. Surv. **13**(1), 52–118 (2019)
33. Pan, V.: On the complexity of a pivot step of the revised simplex algorithm. Comput. Math. Appl. **11**(11), 1127–1140 (1985)
34. Pisier, G. (ed.): The Volume of Convex Bodies and Banach Space Geometry, vol. 94. Cambridge University Press, Cambridge (1999)
35. Raič, M.: A multivariate Berry–Esseen theorem with explicit constants. Bernoulli **25**(4A), 2824–2853 (2019)
36. Rousseeuw, P.J., Ruts, I.: The depth function of a population distribution. Metrika **49**(3), 213–244 (1999)
37. Schütt, C., Werner, E.: The convex floating body. Math. Scand. **66**(2), 275–290 (1990)
38. Shamir, R.: The efficiency of the simplex method: a survey. Manag. Sci. **33**(3), 301–334 (1987)
39. Stroud, A.H.: Approximate Calculation of Multiple Integrals. Prentice-Hall, Hoboken (1971)
40. Tchakaloff, V.: Formules de cubature mécanique à coefficients non négatifs. Bulletin des Sciences Mathématiques **81**, 123–134 (1957)
41. Tchernychova, M.: Carathéodory cubature measures. Ph.D. thesis, University of Oxford (2015)
42. Tukey, J.W.: Mathematics and the picturing of data. In: Proceedings of the International Congress of Mathematicians, Vancouver, 1975, vol. 2, pp. 523–531 (1975)
43. Vershynin, R.: High-Dimensional Probability: An Introduction with Applications in Data Science, vol. 47. Cambridge University Press, Cambridge (2018)
44. Wagner, U., Welzl, E.: A continuous analogue of the upper bound theorem. Discrete Comput. Geom. **26**(2), 205–219 (2001)
45. Wendel, J.G.: A problem in geometric probability. Math. Scand. **11**(1), 109–111 (1963)
46. Zhai, A.: A high-dimensional CLT in $\mathcal{W}_2$ distance with near optimal convergence rate. Probab. Theory Relat. Fields **170**(3–4), 821–845 (2018)
47. Zuo, Y.: A new approach for the computation of halfspace depth in high dimensions. Commun. Stat. Simul. Comput. **48**(3), 900–921 (2019)