

On the geometry of polytopes generated by heavy-tailed random vectors

Olivier Guédon* Felix Krahmer† Christian Kümmeler†
Shahar Mendelson‡ Holger Rauhut§

July 16, 2019

Abstract

We study the geometry of centrally-symmetric random polytopes, generated by N independent copies of a random vector X taking values in \mathbb{R}^n . We show that under minimal assumptions on X , for $N \gtrsim n$ and with high probability, the polytope contains a deterministic set that is naturally associated with the random vector—namely, the polar of a certain floating body. This solves the long-standing question on whether such a random polytope contains a canonical body. Moreover, by identifying the floating bodies associated with various random vectors we recover the estimates that have been obtained previously, and thanks to the minimal assumptions on X we derive estimates in cases that had been out of reach, involving random polytopes generated by heavy-tailed random vectors (e.g., when X is q -stable or when X has an unconditional structure). Finally, the structural results are used for the study of a fundamental question in compressive sensing—noise blind sparse recovery.

1 Introduction

Let X be a symmetric random vector in \mathbb{R}^n and let X_1, \dots, X_N be independent copies of X . The goal of this article is to study the geometry of the random polytope $\text{absconv}(X_1, \dots, X_N)$, that is, the convex hull of the points $\pm X_1, \dots, \pm X_N$. Various aspects of the geometry of such random polytopes have been the subject of extensive study for many years. As a starting point, let us formulate two notable results in the direction we are interested in, and to that end, denote by B_p^n the unit ball in ℓ_p^n .

Theorem 1.1. [18] *Let X be the standard Gaussian random vector in \mathbb{R}^n , set $0 < \alpha < 1$ and consider $N \geq c_0(\alpha)n$. Then*

$$c_1(\alpha)\sqrt{\log(eN/n)}B_2^n \subset \text{absconv}(X_1, \dots, X_N) \quad (1.1)$$

with probability at least $1 - 2\exp(-c_2N^{1-\alpha}n^\alpha)$. Here c_0 and c_1 are constants that depend on α and c_2 is an absolute constant.

Theorem 1.1 can be extended beyond the Gaussian case, to random polytopes generated by a random vector $X = (\xi_1, \dots, \xi_n)$ where the ξ_i 's are independent copies of a mean-zero, variance

*Université Paris-Est, Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), UPEM, UPEC, CNRS, F-77454, Marne-la-Vallée, France (olivier.guedon@u-pem.fr)

†Department of Mathematics, Technical University of Munich, 85748 Garching bei München, Germany (felix.krahmer@tum.de, c.kuemmerle@tum.de)

‡LPSM, Sorbonne University, Paris, France and Mathematical Sciences Institute, The Australian National University, Canberra, Australia. (shahar.mendelson@anu.edu.au)

§Chair for Mathematics of Information Processing, RWTH Aachen University, 52056 Aachen, Germany (rauhut@mathc.rwth-aachen.de)

1 random variable ξ that is L -subgaussian¹. This class of random vectors includes, in particular, the Rademacher vector X , which is uniformly distributed in $\{-1, 1\}^n$.

A version of Theorem 1.1 for the Rademacher vector was established in [17] (with a slightly suboptimal dependence of N on the dimension n). The optimal estimate when ξ is an arbitrary subgaussian random variable is the following special case of a result in [26].

Theorem 1.2. [26] *Let ξ be a mean-zero random variable that has variance 1 and is L -subgaussian, and set $X = (\xi_i)_{i=1}^n$ as above. Let $0 < \alpha < 1$ and set $N \geq c_0(\alpha, L)n$. Then there exists an absolute constant c_1 such that with probability at least $1 - 2 \exp(-c_1 N^{1-\alpha} n^\alpha)$*

$$c_2(\alpha, L)(B_\infty^n \cap \sqrt{\log(eN/n)}B_2^n) \subset \text{absconv}(X_1, \dots, X_N). \quad (1.2)$$

In both cases, the typical random polytope $\text{absconv}(X_1, \dots, X_N)$ contains a large regular convex body: a multiple of the Euclidean unit ball when X is the standard Gaussian random vector, and an intersection body of two ℓ_p balls when X is L -subgaussian and has i.i.d. coordinates. As we explain in what follows, the fact that the bodies that are contained in $\text{absconv}(X_1, \dots, X_N)$ are different in these two examples is not a coincidence. Rather, it reflects the fact that a subgaussian random vector may in general generate a different geometry than the Gaussian one.

Motivated by these two facts, we study the following questions:

Question 1.3. (1) *Is it possible to find a set K that is naturally associated with X and is contained in $\text{absconv}(X_1, \dots, X_N)$ with high probability?*

(2) *If the answer to (1) is yes, when does K contain large (intersections of) ℓ_p balls as, for example, in Theorems 1.1 and 1.2?*

Both Theorem 1.1 and Theorem 1.2, as well as the numerous other results in this direction, can be explained by a general principle stated in our main result, Theorem 1.6, that answers part (1) of Question 1.3. The geometric features of X that are significant in this context are reflected by the natural *floating bodies* associated with X . Part (2) of Question 1.3 will be answered in Section 3 by identifying those floating bodies for a variety of choices of X —thus recovering, and at times improving, previously known results, as well as establishing new estimates in cases that were out of reach before.

Definition 1.4. *Let X be a symmetric random vector in \mathbb{R}^n . For $p \geq 1$, we define the associated floating body*

$$K_p(X) := \{t \in \mathbb{R}^n : \mathbb{P}(\langle X, t \rangle \geq 1) \leq \exp(-p)\}.$$

The notion of floating bodies plays a crucial role in the study of approximation of convex bodies by polytopes, see, e.g., [36, 32, 3], where X is distributed according to the uniform probability measure on the given convex body. It is known how to identify the floating bodies associated to Gaussian or Rademacher random vectors, see below.

In order to continue we require the following notation. Given sets $A, B \subset \mathbb{R}^n$, $A \sim B$ denotes that there are absolute constants c_1 and c_2 such that $c_1 A \subset B \subset c_2 B$. We write $A \sim_\kappa B$ if the constants c_1 and c_2 depend on the parameter κ . Identifying each $t \in \mathbb{R}^n$ with the linear functional $\langle \cdot, t \rangle$, we define, for $p > 0$, the $L_p(X)$ (quasi-) norm of $t \in \mathbb{R}^n$ to be $\|\langle X, t \rangle\|_{L_p} = (\mathbb{E} |\langle X, t \rangle|^p)^{1/p}$, and denote its unit ball by

$$B(L_p(X)) := \{t \in \mathbb{R}^n : \|\langle X, t \rangle\|_{L_p} \leq 1\}.$$

For $1 \leq q < \infty$ and $t = (t_1, \dots, t_n)$, let

$$\|t\|_q = \left(\sum_{i=1}^n |t_i|^q \right)^{1/q} \quad \text{and} \quad \|t\|_\infty = \max_{i=1, \dots, n} |t_i|.$$

¹Recall that a centered random variable is L -subgaussian if for every $p \geq 2$, $(\mathbb{E}|\xi|^p)^{1/p} \leq L\sqrt{p}$.

For $1 \leq q \leq \infty$, let $B_q^n = \{t \in \mathbb{R}^n : \|t\|_q \leq 1\}$ be the unit ball of the normed space ℓ_q^n , and set q' to be the conjugate index of q ; that is, $\frac{1}{q} + \frac{1}{q'} = 1$. Finally, for $T \subset \mathbb{R}^n$ let

$$T^\circ = \{x \in \mathbb{R}^n : \langle t, x \rangle \leq 1 \text{ for every } t \in T\};$$

the set T° is the *polar body* of T , which is a convex, centrally symmetric subset of \mathbb{R}^n if T is centrally symmetric.

With this notation in place, consider the following examples:

- Let $X = G$ be the standard Gaussian random vector in \mathbb{R}^n . Then for every $p \geq 1$,

$$K_p(G) \sim (1/\sqrt{p})B_2^n, \quad (1.3)$$

which can be shown by a direct calculation using the rotation invariance of G . Thus, the polar body of $K_p(G)$ satisfies $(K_p(G))^\circ \sim \sqrt{p}B_2^n$.

- Let $X = \mathcal{E}$ be the Rademacher random vector in \mathbb{R}^n (i.e., X is uniformly distributed in $\{-1, 1\}^n$). Results in [34] imply that

$$K_p(\mathcal{E}) \sim \text{conv}(B_1^n \cup (1/\sqrt{p})B_2^n),$$

and in particular, $(K_p(\mathcal{E}))^\circ \sim B_\infty^n \cap \sqrt{p}B_2^n$.

Thus, the assertions of Theorem 1.1 for $X = G$ and of Theorem 1.2 for $X = \mathcal{E}$ can be formulated in a unified way: with high probability, it holds that

$$\text{absconv}(X_1, \dots, X_N) \supset c_1 (K_p(X))^\circ,$$

for $p = c_2 \log(eN/n)$, where c_1 and c_2 are suitable constants. Our main result shows that this phenomenon holds under minimal assumptions on X , which we explain in the following.

Let $\|\cdot\|$ be a norm on \mathbb{R}^n and set

$$\mathcal{B} = \mathcal{B}_{\|\cdot\|} = \{x \in \mathbb{R}^n : \|x\| \leq 1\} \quad \text{and} \quad \mathcal{S} = \mathcal{S}_{\|\cdot\|} = \{x \in \mathbb{R}^n : \|x\| = 1\}.$$

The random vector X is said to satisfy a small-ball condition with respect to the norm $\|\cdot\|$ with constants γ and δ if for every $t \in \mathbb{R}^n$,

$$\mathbb{P}(|\langle X, t \rangle| \geq \gamma \|t\|) \geq \delta. \quad (1.4)$$

Also, for some $r > 0$, X is said to satisfy an L_r condition with respect to the norm $\|\cdot\|$ and with constant L if for every $t \in \mathbb{R}^n$,

$$(\mathbb{E}|\langle X, t \rangle|^r)^{1/r} \leq L \|t\|. \quad (1.5)$$

Assumption 1.5. *We assume that X satisfies a small-ball condition with constants $\gamma > 0$ and $\delta > 0$, and an L_r condition with constant L for some $r > 0$ with respect to the same norm $\|\cdot\|$.*

Assumption 1.5 implies that the random vector X is not degenerate: the small-ball condition (1.4) means that marginals of X do not have ‘too much’ mass at 0, and the L_r condition (1.5) leads to some minimal uniform control on the tail decay of each marginal. Also, it is straightforward to verify from (1.4) and (1.5) that

$$\frac{1}{L}\mathcal{B} \subset B(L_r(X)) \subset \frac{1}{\gamma\delta^{1/r}}\mathcal{B}.$$

Our answer to the first part of Question 1.3 is that under Assumption 1.5, a typical realization of $\text{absconv}(X_1, \dots, X_N)$ contains a constant multiple of $(K_p(X))^\circ$ for $p \sim \log(eN/n)$.

Theorem 1.6. *Let X be a symmetric random vector that satisfies Assumption 1.5 with respect to a norm $\|\cdot\|$ and some $\delta, \gamma, r, L > 0$. Let $0 < \alpha < 1$ and set $p = \alpha \log(eN/n)$ and assume that $N \geq c_0 n$ for a constant $c_0 = c_0(\alpha, \delta, r, L/\gamma)$. Let X_1, \dots, X_N be independent copies of X then with probability at least $1 - 2 \exp(-c_1 N^{1-\alpha} n^\alpha)$,*

$$\text{absconv}(X_1, \dots, X_N) \supset \frac{1}{2} (K_p(X))^\circ, \quad (1.6)$$

where c_1 is an absolute constant.

Remark 1.7. *Theorem 1.6 still holds – even with the same constants and the same proof – when $\text{absconv}(X_1, \dots, X_N)$ is replaced by the standard convex hull $\text{conv}(X_1, \dots, X_N)$, see also Remark 2.6.*

Assumption 1.5 is weaker than any of the assumptions in all previous results on the inner structure of $\text{absconv}(X_1, \dots, X_N)$. In particular, we allow heavy-tailed distributions and do not require independence of the entries of X . The freedom of choosing the norm $\|\cdot\|$ makes the method very flexible. Observe that (1.6) does not depend on the specific choice of $\|\cdot\|$, but the constant $c_0 = c_0(\alpha, \delta, r, L/\gamma)$ does. In fact, the constants L and γ may change when chaining the norm. So the art consists in choosing a norm such that quotient L/γ , and hence, the constant c_0 become as small as possible.

As applications of Theorem 1.6 we show in Section 3 how one can recover or improve the previous central results on the geometry of the random polytope $\text{absconv}(X_1, \dots, X_N)$ in this context. This is done by answering the second part of Question 1.3: we identify the floating bodied $K_p(X)$ in all those cases, for example, when X is the Gaussian vector ([18]); when X has i.i.d. subgaussian centered coordinates ([26]); when X is an isotropic, log-concave random vector ([10]); and when X has i.i.d. centered coordinates that satisfy a small-ball condition ([19]).

In addition, and thanks to the universality of Theorem 1.6, one may establish various new outcomes that were previously completely out of reach like when X is an unconditional random vector without necessarily independent entries, see Theorem 3.9. The main applications we present in this introduction are two results that we found to be particularly surprising: firstly, an answer to Question 1.3 when X has i.i.d. q -stable coordinates for $1 \leq q < 2$ (e.g., a Cauchy random vector); and secondly, an answer to a fundamental question on sparse recovery.

1.1 Stable random vectors

Consider standard q -stable random vectors for $1 \leq q < 2$ (a 2-stable random vector is just a Gaussian), that is, vectors that have i.i.d. standard q -stable random variables as coordinates. Recall that a random variable ξ is standard q -stable if its characteristic function satisfies $\mathbb{E}[\exp(itX)] = \exp(-|t|^q/2)$ for every $t \in \mathbb{R}$ (we consider only the symmetric case). The following features of a standard q -stable random variable ξ are of significance here:

- ξ belongs to the weak- L_q space; i.e., $\sup_{u>0} u^q \mathbb{P}(|\xi| > u) \leq C_q$, and for large values of u , $\mathbb{P}(|\xi| > u) \geq c_q/u^q$.
- the stability property: if ξ_1, \dots, ξ_n are i.i.d. copies of ξ and $t \in \mathbb{R}^n$ then $\sum_{i=1}^n t_i \xi_i$ has the same distribution as $\|t\|_q \xi$.

For a more comprehensive discussion on q -stable random variables see, e.g., [25, Chapter 5]. Note that for $q < 2$, ξ does not have a finite second moment, which makes the analysis of the structure of the random polytope $\text{absconv}(X_1, \dots, X_N)$ more challenging.

The answer to Question 1.3 for a q -stable random vector is as follows:

Theorem 1.8. *Let ξ be a standard, q -stable random variable for some $1 \leq q < 2$. Let ξ_1, \dots, ξ_n be independent copies of ξ and set $X = (\xi_i)_{i=1}^n$. Then for $0 < \alpha < 1$ and $N \geq c_0(\alpha, q)n$, with probability at least $1 - 2 \exp(-c_1 N^{1-\alpha} n^\alpha)$,*

$$\text{absconv}(X_1, \dots, X_N) \supset c_2(q) \left(\frac{N}{n}\right)^{\alpha/q} B_q^n$$

where $1/q + 1/q' = 1$.

In particular, if ξ is a standard Cauchy random variable (corresponding to $q = 1$) then with probability at least $1 - 2\exp(-c_1 N^{1-\alpha} n^\alpha)$

$$\text{absconv}(X_1, \dots, X_N) \supset c_3 \left(\frac{N}{n}\right)^\alpha B_\infty^n.$$

Observe that a typical realization of $\text{absconv}(X_1, \dots, X_N)$ is much larger than, say, the typical realization of the random polytope generated by the Gaussian random vector. Indeed, the latter only contains $\sqrt{\log(eN/n)} B_2^n$, which is a much smaller set than $c(N/n)^{\alpha/q} B_q^n$. The intuitive reason behind this phenomenon is that for $q < 2$, a q -stable random variable is more ‘heavy-tailed’ than the Gaussian random variable: its tail decay is of the order of u^{-q} rather than $\exp(-u^2/2)$ and that difference leads to the polynomial growth of the ‘inner radius’ of $\text{absconv}(X_1, \dots, X_N)$. At the same time, the difference in the canonical body contained in $\text{absconv}(X_1, \dots, X_N)$ is due to the natural metric associated with X : each marginal $\langle t, X \rangle$ is distributed as $\|t\|_q \xi$ rather than as $\|t\|_2 \xi$.

The proof of Theorem 1.8 is presented in Section 3.1.1.

1.2 Relation to Compressive Sensing

The second surprising outcome of Theorem 1.6 is related to a fundamental question in the area of *compressive sensing*²: can sparse signals be recovered efficiently when the given data consist of a few measurements that are noisy, *but the ‘noise level’ is not known*.

Suppose one would like to recover an unknown vector (signal) $x \in \mathbb{R}^N$ from an underdetermined set of a linear measurements, i.e., from $y = Ax \in \mathbb{R}^n$, where $A \in \mathbb{R}^{n \times N}$ with n much smaller than N . While this is impossible in general, the theory of compressive sensing studies when such recovery is possible by efficient methods for (*s*-)sparse vectors, i.e., vectors in \mathbb{R}^N that satisfy $\|x\|_0 = |\{\ell : x_\ell \neq 0\}| \leq s \ll n$.

One of the main achievement of compressive sensing was the discovery that a computationally efficient recovery procedure can be used to recover the signal. Indeed, if x^\sharp is the solution of the ℓ_1 -minimization problem

$$\min_{z \in \mathbb{R}^N} \|z\|_1 \quad \text{subject to } Az = y, \quad (1.7)$$

then for a well-chosen $\sim s \log(eN/s)$ measurements, x^\sharp coincides with the original s -sparse x . This upper estimate on the required number of measurements is optimal, and it is attained by a wide variety of random measurement ensembles—for example, if the measurements are $(\langle G_i, x \rangle)_{i=1}^n$, i.e., A is a draw of a random matrix with independent, mean-zero, variance one, Gaussian entries.

Naturally, to be of value in real-life applications, recovery should be possible in the presence of noise. The additional appeal of ℓ_1 -minimization is that it can be modified to perform well even if the given measurements $(\langle Ax, e_i \rangle)_{i=1}^n$ are corrupted by noise, and if the signal x is not necessarily sparse but only approximately sparse in some appropriate sense. Indeed, assume that the data one is given is $\bar{y} = Ax + w$ for $x \in \mathbb{R}^N$ and a vector $w \in \mathbb{R}^n$ of perturbations (noise) *with a known noise level* $\|w\|_2 \leq \eta$. It is important to emphasize that unlike standard problems in statistics, here w is an arbitrary vector, rather than a random draw according to some statistical law.

One can show that for a variety of random matrices, a sample size of $n \sim s \log(eN/s)$ suffices to ensure that the minimizer x^\sharp of the modified ℓ_1 -minimization problem

$$\min_{z \in \mathbb{R}^N} \|z\|_1 \quad \text{subject to } \|Az - \bar{y}\|_2 \leq \eta \quad (1.8)$$

satisfies

$$\|x - x^\sharp\|_1 \lesssim \sigma_s(x)_1 + \eta \sqrt{\frac{s}{n}}, \quad (1.9)$$

²For more information on compressive sensing we refer the reader to [14, 9, 16], and for more a detailed explanation on the connections between the geometry of random polytopes and sparse recovery, see [12, 40, 9, 16, 15, 8].

where

$$\sigma_s(x)_1 = \inf_{z: \|z\|_0 \leq s} \|x - z\|_1$$

is the best ℓ_1 approximation error of x by an s -sparse vector; again, this is the best estimate one can hope for.

Unfortunately, the ℓ_1 -minimization procedure of (1.8) requires accurate information on the true noise level $\|w\|_2$, or at least a good upper estimate of it. However, in real world applications, this information is often not available. Getting the noise level wrong renders the estimate (1.9) useless: if the employed value of η is an underestimation of the true noise level then the error bound (1.9) need not be valid. On the other hand, if η is chosen to be significantly larger than the true noise level, the resulting error estimate (1.9) (involving the chosen η) is terribly loose.

As it happens, one can show that *noise blind recovery*, in which the noise level is not known, is possible if the measurement matrix A satisfies two conditions:

- (1) A version of the *null space property* (NSP), see (B.1). We refer the reader to [9, 16] for a detailed exposition on the NSP.

Identifying matrices that satisfy the null space property has been of considerable interest in recent years and many examples can be found, for example, in [2, 9, 16, 23, 31, 13]. From our perspective, and somewhat inaccurately put, it is important to note that the NSP is (almost) a necessary condition for sparse recovery in noise-free problems. Therefore, to have any hope of successful recovery in noisy problems, the measurement matrix has to satisfy some version of the NSP.

- (2) The second, and seemingly more restrictive condition is the so-called ℓ_1 quotient property [12]. The matrix A satisfies the ℓ_1 quotient property with respect to the norm $\|\cdot\|$ if for every $w \in \mathbb{R}^n$ there exists a vector $v \in \mathbb{R}^N$ such that $Av = w$ and

$$\|v\|_1 \leq D^{-1} \|w\|. \quad (1.10)$$

It follows that if A satisfies an appropriate null space property and the ℓ_1 -quotient property, then the solution x^\sharp of (1.7) for $y = Ax + w$ satisfies

$$\|x^\sharp - x\|_1 \lesssim \sigma_s(x)_1 + \|w\|; \quad (1.11)$$

in other words, the noise-blind recovery error depends on ‘how far’ x is from being sparse and on the norm $\|w\|$ of the noise vector. For the sake of completeness, an outline of the proof of (1.11) can be found in Appendix B.

Theorem 1.6 implies that contrary to prior belief, (2) is not restrictive at all; in fact, it is almost universal. Indeed, let $\|\cdot\|_p$ be the norm whose unit ball is the polar body $(K_p(X))^\circ$, i.e.,

$$\|x\|_p = \inf\{t > 0 : x \in t(K_p(X))^\circ\}.$$

Set $A = (X_1 | \dots | X_N)$ to be the random matrix whose columns are independent random draws of the random vector X . Then the inclusion from Theorem 1.6 implies that for each vector $w \in \mathbb{R}^n$ there exists a vector $v \in \mathbb{R}^N$ such that $Av = w$ and

$$\|v\|_1 \leq c_2^{-1} \|w\|_p, \quad (1.12)$$

which is precisely the ℓ_1 quotient property with respect to the norm $\|\cdot\|_p$.

Thanks to the study of the floating bodies $K_p(X)$ presented in Section 3, the norm $\|\cdot\|_p$ can be identified in a variety of cases, and in some of which the appropriate null space property has already been established – leading the error bound (1.11). These examples include some of the natural random ensembles that are used in sparse recovery, for example, when X has i.i.d. subgaussian or subexponential coordinates [2, 15]; when X is an isotropic, log-concave random vector [2]; and when X has independent coordinates that have $\log(N)$ finite moments [31, 13] (for example, when

the coordinates are distributed according to the Student- t distribution with $\sim \log N$ degrees of freedom).

Thanks to Theorem 1.6, the ℓ_1 quotient property can be established in those (and many other) cases, implying that noise-blind recovery is possible. To give a flavour of such a result, we present the example of the Student- t distribution in an appendix. More information and numerical experiments are given in [22].

2 Proof of the main result

For the proof of Theorem 1.6, we need some basic properties of the floating body

$$K_p(X) = \{t \in \mathbb{R}^n, \mathbb{P}(\langle X, t \rangle \geq 1) \leq e^{-p}\}.$$

Recall that a set K is star-shaped around 0 if for every $x \in K$ and any $0 \leq \lambda \leq 1$, $\lambda x \in K$.

Proposition 2.1. *Let X be a symmetric random vector on \mathbb{R}^n . Then*

- (1) *The set $K_p(X)$ is star-shaped and symmetric around 0. Moreover, for any $a > 0$,*

$$aK_p(X) = \{t \in \mathbb{R}^n, \mathbb{P}(\langle X, t \rangle \geq a) \leq e^{-p}\}.$$

- (2) *Let $\|\cdot\|$ be a norm on \mathbb{R}^n and denote its unit ball by $\mathcal{B} = \mathcal{B}_{\|\cdot\|} = \{t \in \mathbb{R}^n : \|t\| \leq 1\}$. If X satisfies the small-ball condition (1.4) with respect to the norm $\|\cdot\|$ with constants γ and δ , then for $p > \log(2/\delta)$,*

$$\gamma \operatorname{absconv}(K_p(X)) \subset \mathcal{B}. \quad (2.1)$$

- (3) *If X satisfies the L_r condition (1.5) with respect to the norm $\|\cdot\|$ with constant L , then*

$$\mathcal{B} \subset L \exp(p/r) K_p(X). \quad (2.2)$$

Proof. The first observation is straightforward. To prove (2.1) observe that by convexity, it is enough to show that $\gamma K_p(X) \subset \mathcal{B}$. But if $\|t\| \geq 1$ then the small-ball condition and the symmetry of X imply that

$$\mathbb{P}(\langle X, t \rangle \geq \gamma) \geq \mathbb{P}(\langle X, t \rangle \geq \gamma \|t\|) \geq \frac{\delta}{2} > \exp(-p),$$

provided that $\delta > 2 \exp(-p)$, as was assumed. Hence, $t \notin \gamma K_p(X)$.

As for (2.2), note that for $\|u\| \leq 1$, the L_r condition yields that $\mathbb{E}|\langle X, u \rangle|^r \leq L^r$ and thus by Markov's inequality

$$\mathbb{P}(\langle X, u \rangle \geq L \exp(p/r)) \leq \frac{\mathbb{E}|\langle X, u \rangle|^r}{L^r} \exp(-p) \leq \exp(-p),$$

hence, $u \in L \exp(p/r) K_p(X)$. ■

An outcome of Proposition 2.1 is that if X satisfies Assumption 1.5 and

$$p > \log(2/\delta), \quad (2.3)$$

then $K_p(X)$ is a centrally symmetric subset of \mathbb{R}^n that is star-shaped around 0 and for which

$$(1/L) \exp(-p/r) \mathcal{B} \subset K_p(X) \subset (1/\gamma) \mathcal{B}. \quad (2.4)$$

Let \mathcal{S} be the unit sphere of $(\mathbb{R}^n, \|\cdot\|)$. For $\theta \in \mathcal{S}$ set

$$r(\theta) = \sup\{\beta \geq 0 : \beta\theta \in K_p(X)\}$$

and note that by (2.4), $(1/L) \exp(-p/r) \leq r(\theta) \leq 1/\gamma$. With a possible abuse of notation, put

$$\partial K_p(X) = \{r(\theta)\theta : \theta \in \mathcal{S}\}. \quad (2.5)$$

Note that $\partial K_p(X)$ may not coincide with the topological boundary of $K_p(X)$ as $\theta \mapsto r(\theta)$ need not be continuous on \mathcal{S} for general X .

Corollary 2.2. For every $\theta \in \mathcal{S}$,

$$\mathbb{P}(\langle X, r(\theta)\theta \rangle \geq 1) \geq \exp(-p).$$

Proof. It follows from the definition of $r(\theta)$ that for any $\rho > 1$, $\rho r(\theta)\theta \notin K_p(X)$, and thus,

$$\mathbb{P}(\langle X, \rho r(\theta)\theta \rangle \geq 1) > \exp(-p).$$

Taking the intersection of these events for any $\rho > 1$ gives the result. \blacksquare

The proof of Theorem 1.6 follows the path set in the (much simpler) proof of Theorem 1.5 from [30]. The goal is to show that if X satisfies Assumption 1.5, and

$$N \geq c_0(\alpha, \delta, r, L/\gamma) n, \quad p = \alpha \log(eN/n),$$

then with probability at least

$$1 - 2 \exp(-c_1 N^{1-\alpha} n^\alpha),$$

one has that

$$\frac{1}{2}(K_p(X))^\circ \subset \text{absconv}(X_1, \dots, X_N). \quad (2.6)$$

For a symmetric convex body U with a nonempty interior, define its support function h_U by

$$h_U(t) = \sup_{u \in U} \langle u, t \rangle, \quad \text{for all } t \in \mathbb{R}^n.$$

The inclusion (2.4) ensures that $(K_p(X))^\circ$ has nonempty interior. Therefore, (2.6) is equivalent to

$$\frac{1}{2}h_{(K_p(X))^\circ}(t) \leq h_{\text{absconv}(X_1, \dots, X_N)}(t)$$

for every $t \in \mathbb{R}^n$; and the negation of this event is that there exists $t \in \mathbb{R}^n$ such that

$$\frac{1}{2} \sup_{u \in (K_p(X))^\circ} \langle u, t \rangle > \sup_{v \in \text{absconv}(X_1, \dots, X_N)} \langle v, t \rangle. \quad (2.7)$$

By homogeneity of (2.7) and since $\theta \mapsto r(\theta)$ is bounded away from 0 on \mathcal{S} , it suffices to show that there is $t \in \partial K_p(X)$ for which (2.7) holds. Denote by $\Gamma : \mathbb{R}^n \rightarrow \mathbb{R}^N$ the random matrix whose rows are X_1, \dots, X_N . Observe that $\text{absconv}(X_1, \dots, X_N) = \Gamma^* B_1^N$, and therefore

$$\sup_{u \in \Gamma^* B_1^N} \langle u, t \rangle = \sup_{x \in B_1^N} \langle x, \Gamma t \rangle = \|\Gamma t\|_\infty. \quad (2.8)$$

Moreover, for $t \in \partial K_p(X)$, the definition of polarity gives $\sup_{u \in (K_p(X))^\circ} \langle u, t \rangle \leq 1$. Hence, for the proof of Theorem 1.6 it remains to show that

$$\mathbb{P} \left(\inf_{t \in \partial K_p(X)} \|\Gamma t\|_\infty \leq 1/2 \right) \leq 2 \exp(-c_1 N^{1-\alpha} n^\alpha). \quad (2.9)$$

The proof of (2.9) is based on the small-ball method (see, for example, [29]). First, fix any $t \in \partial K_p(X)$ and recall that by Corollary 2.2,

$$\mathbb{P}(\langle X, t \rangle \geq 1) \geq \exp(-p).$$

Therefore, by independence of the X_i and Chernoff's inequality, with probability at least

$$1 - \exp(-N \exp(-p)/8), \quad (2.10)$$

it holds that

$$|\{i : \langle X_i, t \rangle \geq 1\}| \geq \frac{N}{2} \exp(-p). \quad (2.11)$$

Second, thanks to the high probability estimate (2.10), it follows from the union bound that if $T \subset \partial K_p(X)$ with

$$|T| \leq \exp(N \exp(-p)/16), \quad (2.12)$$

then

$$\inf_{t \in T} |\{i : \langle X_i, t \rangle \geq 1\}| \geq \frac{N}{2} \exp(-p). \quad (2.13)$$

with probability at least

$$1 - \exp(-N \exp(-p)/16).$$

The only restriction on the set $T \subset \partial K_p(X)$ is its cardinality. With this in mind, we will define T as a covering of $\partial K_p(X)$ with balls of appropriate radius associated to the norm $\|\cdot\|$.

Observe that by (2.1), $\partial K_p(X) \subset (1/\gamma)\mathcal{B}$ provided that $p > \log(2/\delta)$. By a standard volumetric estimate, see, e.g., [16, Proposition C.3], for every $\rho > 0$ there exists a η/γ -cover of $\partial K_p(X)$ with respect to the norm $\|\cdot\|$ of cardinality at most $(1 + 2/\eta)^n$. This η/γ -cover has the required cardinality (2.12) if

$$\eta \geq 2 \left(\exp \left(\frac{N}{16n} \exp(-p) \right) - 1 \right)^{-1}. \quad (2.14)$$

If

$$N \geq 16 \ln(2) \exp(p) n \quad (2.15)$$

then (2.14) is satisfied for the choice

$$\eta = 4 \exp \left(-\frac{N}{16n} \exp(-p) \right). \quad (2.16)$$

Denoting by \mathcal{A}_1 the event on which (2.13) holds for T that is a minimal η/γ -cover of $\partial K_p(X)$, it is evident that

$$\mathbb{P}(\mathcal{A}_1) \geq 1 - \exp(-N \exp(-p)/16).$$

Finally, for every $t \in \partial K_p(X)$ let $\pi t \in T$ be the nearest element to t in the (η/γ) -cover with respect to the norm $\|\cdot\|$. Consider the event \mathcal{A}_2 on which

$$\sup_{t \in T} |\{i : |\langle X_i, t - \pi t \rangle| \geq 1/2\}| \leq \frac{3N}{8} \exp(-p). \quad (2.17)$$

For each $t \in \partial K_p(X)$ consider the sets of indices

$$I_1(t) := \{i : \langle X_i, \pi t \rangle \geq 1\}, \quad I_2(t) := \{i : |\langle X_i, t - \pi t \rangle| \geq 1/2\}.$$

and observe that on the event $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$,

$$|I_1(t)| \geq \frac{N}{2} \exp(-p), \quad |I_2(t)| \leq \frac{3N}{8} \exp(-p).$$

Clearly,

$$|I_1(t)| + |I_2^c(t)| \geq \frac{N}{2} \exp(-p) + (N - \frac{3N}{8} \exp(-p)) = N + \frac{N}{8} \exp(-p),$$

and therefore

$$|I_1(t) \cap I_2^c(t)| \geq \frac{N}{8} \exp(-p).$$

For each $t \in I(t) := I_1(t) \cap I_2^c(t)$ the triangle inequality gives

$$\langle X_i, t \rangle \geq \langle X_i, \pi t \rangle - |\langle X_i, t - \pi t \rangle| \geq \frac{1}{2}.$$

In particular, on the event \mathcal{A} , it holds that $\inf_{t \in \partial K_p(X)} \|\Gamma t\|_\infty \geq \frac{1}{2}$ and

$$\mathbb{P} \left(\inf_{t \in \partial K_p(X)} \|\Gamma t\|_\infty \leq 1/2 \right) \leq \mathbb{P}(\mathcal{A}^c).$$

Finally, let us show that $\mathbb{P}(\mathcal{A}_2)$ is ‘large enough’ for the right choice of p . To that end, observe that for every $t \in \partial K_p(X)$, $\|t - \pi t\| \leq (\eta/\gamma)$, and therefore,

$$\begin{aligned} \sup_{t \in \partial K_p(X)} |\{i : |\langle X_i, t - \pi t \rangle| \geq 1/2\}| &\leq \sup_{u \in (\eta/\gamma)\mathcal{B}} |\{i : |\langle X_i, u \rangle| \geq 1/2\}| \\ &= \sup_{u \in (\eta/\gamma)\mathcal{B}} \sum_{i=1}^N \mathbb{1}_{\{|\langle X_i, u \rangle| \geq 1/2\}}, \end{aligned}$$

which is the supremum of an empirical process indexed by the class of indicator functions

$$\mathcal{F} = \left\{ \mathbb{1}_{\{|\langle \cdot, u \rangle| \geq 1/2\}} : u \in (\eta/\gamma)\mathcal{B} \right\}. \quad (2.18)$$

The wanted estimate on this supremum is based on an outcome of Talagrand’s concentration inequality for bounded empirical processes, in the special case in which the indexing class is binary-valued and has a finite Vapnik-Chervonenkis (VC) dimension (for a definition of the VC dimension, see, e.g., [39]).

Before stating this result, let us first recall the definition of VC dimension and a basic bound needed in our proof.

Definition 2.3. *Let \mathcal{F} be a class of $\{0, 1\}$ -valued functions on a space Ω . The class shatters $\{x_1, \dots, x_k\} \subset \Omega$, if for every $I \subset \{1, \dots, k\}$ there exists a function $f_I \in \mathcal{F}$ for which $f_I(x_i) = 1$ if $i \in I$ and $f_I(x_i) = 0$ if $i \notin I$. Let*

$$VC(\mathcal{F}) = \sup \{|A| : A \subset \Omega, A \text{ is shattered by } \mathcal{F}\}.$$

Lemma 2.4. *Let \mathcal{D} be a set of subsets of Ω such that the set of indicator functions $\mathcal{F} = \{\mathbb{1}_D : D \subset \mathcal{D}\}$ satisfies $VC(\mathcal{F}) = d$. If $\tilde{\mathcal{F}} = \{\mathbb{1}_{D \cup D'} : D, D' \in \mathcal{D}\}$ then $VC(\tilde{\mathcal{F}}) < 10d$.*

Proof. The statement is a special case of [4, Lemma 3.2.3], which treats the case of the class of k unions, i.e., $\mathcal{D}^k = \{D_1 \cup \dots \cup D_k : D_1, \dots, D_k \in \mathcal{D}\}$, and states that the VC dimension of the corresponding class \mathcal{F}^k of indicator functions satisfies $VC(\mathcal{F}^k) < 2dk \log_2(3k)$. For $k = 2$ one has $VC(\tilde{\mathcal{F}}) = cd$ with $c = 4 \log_2(6) \approx 10.34$. The slightly better constant 10 (or even 9.4) follows from an inspection of the proof, which shows that a strict upper bound for the VC dimension of $\tilde{\mathcal{F}}$ is any m such that $(em/d)^{2d} < 2^m$. An explicit calculation shows that $m = 10d$ is a valid choice. ■

Let us now state the outcome of Talagrand’s concentration inequality when the indexing set of functions is a VC class (see [37] and also [28, Lemma 3.7]).

Theorem 2.5. *Let \mathcal{F} be a class of $\{0, 1\}$ -valued functions for which $VC(\mathcal{F}) \leq d$ and $\sup_{f \in \mathcal{F}} \mathbb{E}f^2 \leq \sigma^2$. Set*

$$R := 64 \frac{d}{N} \log \left(\frac{c}{\sigma^2} \right) + 8\sigma \sqrt{\frac{d}{N} \log \left(\frac{c}{\sigma^2} \right)}, \quad (2.19)$$

where $c = 8e^2\sqrt{2} \approx 83.6$. Then for any $x > 0$,

$$\mathbb{P} \left(\sup_{f \in \mathcal{F}} \left| \frac{1}{N} \sum_{i=1}^N f(X_i) - \mathbb{E}f \right| \geq R + x \right) \leq \exp \left(-N \frac{x^2/2}{\sigma^2 + 2R + x/3} \right).$$

For the sake of completeness, we provide a sketch of the argument in Appendix A.

Let us return to the proof of Theorem 1.6 and consider \mathcal{F} as defined in (2.18). Each $f \in \mathcal{F}$ is the indicator of a union of two half spaces in \mathbb{R}^n . By Radon’s theorem, the VC dimension of the class of indicators of half spaces in \mathbb{R}^n is $n + 1$, see e.g. [33, Theorem 3.4]. It follows from Lemma 2.4 that $d := VC(\mathcal{F}) < 10(n + 1)$. Moreover, by the L_r condition from Assumption 1.5 and Markov’s inequality, for any $\theta \in [0, 1]$,

$$\sup_{f \in \mathcal{F}} \mathbb{E}f^2 = \sup_{u \in (\eta/\gamma)\mathcal{B}} \mathbb{P} \left(|\langle X, u \rangle| \geq \frac{1}{2} \right) \leq \min \left\{ 1, \left(\frac{2L\eta}{\gamma} \right)^r \right\} \leq \min \left\{ 1, \left(\frac{2L\eta}{\gamma} \right)^{\theta r} \right\}.$$

Hence, any $\sigma^2 \geq (2L\eta/\gamma)^{\theta r}$ is a valid choice in the context of Theorem 2.5. By our choice of η in (2.16), this requirement is fulfilled for

$$\sigma^2 = (\max\{1, 8L/\gamma\})^{\theta r} \exp\left(-\theta r \frac{N}{16n} \exp(-p)\right).$$

With that choice of σ and $p = \alpha \log(eN/n)$ the first term in the definition (2.19) of R can be bounded as

$$\begin{aligned} T_1 &:= \frac{64d}{N} \log\left(\frac{c}{\sigma}\right) < \frac{640(n+1)}{N} \left(\log(c) + \theta r \left(-\log(\max\{1, 8L/\gamma\}) + \frac{N}{16n} \exp(-p) \right) \right) \\ &\leq \frac{640 \log(c)(n+1)}{N} + \frac{40(n+1)}{n} \theta r \left(\frac{n}{eN}\right)^\alpha. \end{aligned}$$

Choosing $\theta = c_1 \min\{1, 1/r\}$ with $c_1 = 1/(16 \cdot 80)$ and assuming $N \geq c_2 n$ for a suitable constant $c_2 = c_2(\alpha)$, it is evident that $\frac{640 \log(c)(n+1)}{N} \leq \left(\frac{n}{eN}\right)^\alpha / 16$ and therefore,

$$T_1 \leq \frac{1}{8} \left(\frac{n}{eN}\right)^\alpha = \frac{1}{8} \exp(-p). \quad (2.20)$$

Also, under the same assumptions, the second term in the definition (2.19) of R can be estimated using (2.20) as

$$\begin{aligned} T_2 &:= 8\sigma \sqrt{\frac{d}{N} \log\left(\frac{c}{\sigma^2}\right)} \leq 8\sigma \sqrt{\frac{1}{8 \cdot 64} \exp(-p)} \\ &\leq (\max\{1, 8L/\gamma\})^{\theta r/2} \exp\left(-\theta r \frac{N}{32n} \exp(-p)\right) \sqrt{\frac{1}{8} \exp(-p)} \\ &= \frac{(\max\{1, 8L/\gamma\})^{c_1 \frac{\min\{1, r\}}{2}}}{\sqrt{8}} \exp\left(-c_1 \frac{\min\{1, r\}}{32e^\alpha} \left(\frac{N}{n}\right)^{1-\alpha} - \frac{\alpha}{2} \ln(eN/n)\right) \\ &\leq \frac{1}{8} \exp(-\alpha \ln(eN/n)) = \frac{1}{8} \exp(-p), \end{aligned}$$

provided that $N \geq c_3 n$ for some suitable $c_3 = c_3(\alpha, r, L/\gamma)$. Combining the two estimates, it follows that

$$R = T_1 + T_2 \leq \frac{1}{4} \exp(-p).$$

Moreover, with a similar argument we have that

$$\sigma \leq \frac{1}{16} \exp(-p)$$

provided that $N \geq c_4 n$ with $c_4 = c_4(\alpha, r, L/\gamma)$; furthermore,

$$\sup_{f \in \mathcal{F}} \mathbb{E}f \leq \sup_{f \in \mathcal{F}} (\mathbb{E}f^2)^{1/2} \leq \sigma \leq \frac{1}{16} \exp(-p).$$

Now, recall that we assumed (2.3), i.e., that $p > \log(2/\delta)$, which by definition of p is equivalent to $N > e(2/\delta)^{1/\alpha} n$. At the same time, the requirement (2.15) is equivalent to $N \geq (16 \ln(2)e^\alpha)^{1/(1-\alpha)} n$.

Summarizing, all required conditions on N are satisfied if $N \geq c_0 n$ with

$$c_0 = c_0(\alpha, r, L/\gamma, \delta) = \max\left\{c_2(\alpha), c_3(\alpha, r, L/\gamma), c_4(\alpha, r, L/\gamma), 3(2/\delta)^{1/\alpha}, (16 \ln(2)e^\alpha)^{1/(1-\alpha)}\right\}.$$

In this case, choosing $x = \exp(-p)/16$ in Theorem 2.5 and noting that

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{N} \sum_{i=1}^N f(X_i) \right| \leq \sup_{f \in \mathcal{F}} \left| \frac{1}{N} \sum_{i=1}^N f(X_i) - \mathbb{E}f \right| + \sigma,$$

it is evident that

$$\sup_{u \in (\eta/\gamma)\mathcal{B}} \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{\langle X_i, z \rangle \geq 1/2\}} \leq R + \sigma + x \leq \frac{3}{8} \exp(-p)$$

outside an event whose probability is at most

$$\begin{aligned} \exp\left(-N \frac{x^2/2}{\sigma^2 + 2R + x/3}\right) &\leq \exp\left(-N \frac{\exp(-p)^2/(2 \cdot 16^2)}{\exp(-p)^2/256 + \exp(-p)/2 + \exp(-p)/48}\right) \\ &\leq \exp(-c_6 N \exp(-p)) = \exp(-c_1 N^{1-\alpha} n^\alpha), \end{aligned}$$

where $c_6 = 3/806$ and $c_1 = c_6/e \approx 0.0014$. This completes the proof of (2.17) and, hence, of Theorem 1.6. \blacksquare

Remark 2.6. *The proof only needs very little adaptation if one replaces $\text{absconv}(X_1, \dots, X_N)$ by the standard convex hull $\text{conv}(X_1, \dots, X_N)$. In fact, $\text{conv}(X_1, \dots, X_N) = \Gamma^* \Delta_1^N$, where $\Delta_1^N = \{x \in \mathbb{R}^N : x_i \geq 0, \sum_{i=1}^N x_i \leq 1\}$ is the standard simplex. Then $\|\Gamma t\|_\infty$ in (2.8) and (2.9) is replaced by $\max_{i=1, \dots, N} (\Gamma t)_i = \max_{i=1, \dots, N} \langle X_i, t \rangle$. Now, (2.11) works without the absolute values around $\langle X_i, t \rangle$, anyway, so that the rest of the proof remains the same.*

3 The floating bodies for various random vectors

Although Theorem 1.6 is (almost) universal, it is unrealistic to expect that the second part of Question 1.3 can be addressed with a single result. Therefore, the identity of the sets $K_p(X)$ has to be studied on a case-by-case basis. Having said that, there are some general principles that can be used to identify, or at least approximate the sets $K_p(X)$. Firstly, as outlined in what follows, there are natural examples in which $K_p(X)$ can be identified directly—among them are the standard Gaussian vector $X = G$; the standard Rademacher vector $X = \mathcal{E}$; and when X is a q -stable random vector. Secondly, we show in Section 3.2 that if linear forms $\langle X, t \rangle$ have p -th moments and satisfy a weak regularity condition, then $K_p(X)$ is equivalent to $B(L_p(X))$. Perhaps, one could have actually expected a variant of Theorem 1.6 with $K_p(X)$ replaced by $B(L_p(X))$ in the first place, but clearly $B(L_p(X))$ does not work in heavy-tailed situations, where it may be trivial if $p = \alpha \log(eN/n) > r$. This observation, combined with Theorem 1.6 improves the main result from [10] which studies random polytopes generated by isotropic, log-concave random vectors. Then, in Section 3.3, we explain how stochastic domination can be translated to information on the structures of the floating bodies. That allows one to show that $\text{absconv}(X_1, \dots, X_N)$ contains large canonical sets for very general random vectors, even when X does not necessarily have independent entries.

3.1 Direct analysis of the floating body

The first two natural examples one should consider are $X = G$, the standard Gaussian random vector and $X = \mathcal{E}$, the standard Rademacher random vector. A direct computation shows that

$$c_1 \frac{1}{\sqrt{p}} B_2^n \subset K_p(G) \subset c_2 \frac{1}{\sqrt{p}} B_2^n,$$

and by [34],

$$c'_1 \text{conv}(B_1^n \cup (1/\sqrt{p})B_2^n) \subset K_p(\mathcal{E}) \subset c'_2 \text{conv}(B_1^n \cup (1/\sqrt{p})B_2^n),$$

where c_1, c'_1, c_2 and c'_2 are absolute constants. Therefore, in both cases, Theorem 1.6 implies that $\text{absconv}(X_1, \dots, X_N)$ contains a large canonical body. In particular, one recovers the estimates of Theorem 1.1 and of Theorem 1.2 for the Rademacher random vector \mathcal{E} stating that with high probability,

$$\text{absconv}(G_1, \dots, G_N) \supset c_2 \sqrt{\alpha \log(eN/n)} B_2^n$$

and

$$\text{absconv}(\mathcal{E}_1, \dots, \mathcal{E}_N) \supset c'_2(B_\infty^n \cap \sqrt{\alpha \log(eN/n)} B_2^n).$$

We explain how Theorem 1.2 can be recovered from Theorem 1.6 in full generality in Section 3.3.

Another, more surprising example in which $K_p(X)$ can be computed directly consists in the case that X is a standard q -stable random vector, a situation outlined in Theorem 1.8.

3.1.1 Proof of Theorem 1.8

Recall that for $1 \leq q < 2$, a random variable ξ is called standard q -stable if its characteristic function satisfies $\mathbb{E}[\exp(itX)] = \exp(-|t|^q/2)$ for every $t \in \mathbb{R}$ (we consider only the symmetric case). The proof of Theorem 1.8 is based on several well known facts, see, e.g., [25, Chapter 5].

- (F1) If ξ_1, \dots, ξ_n are independent copies of a standard q -stable random variable ξ , and $X = (\xi_i)_{i=1}^n$, then for any $t \in \mathbb{R}^n$, $\langle t, X \rangle$ has the same distribution as $\xi \|t\|_q$.
- (F2) While a standard q -stable random variable does not belong to L_q , it does belong to the weak- L_q space $L_{q,\infty}$, i.e., $\sup_{u>0} u^q \mathbb{P}(|\xi| \geq u) \leq C_q$ for some constant $C_q > 0$.
- (F3) The weak L_q behaviour of ξ is sharp: there exist constants $M_q, c_q > 0$ such that for any $u \geq M_q$, $\mathbb{P}(|\xi| \geq u) \geq c_q/u^q$.

From here on, let ξ be a standard q -stable random variable for some $1 \leq q < 2$. Let us first show that X satisfies Assumption 1.5, though obviously, due to the stability property (F1), *not* with respect to the Euclidean norm, but rather with respect to $\|\cdot\|_q$. By (F2), ξ has a bounded L_r (quasi)-norm for any $0 < r < q$. As a result, X satisfies the L_r condition (1.5) with respect to $\|\cdot\|_q$ for $r = q/2$ and constant $L = L_q$. At the same time, e.g., by a Paley-Zygmund argument (see e.g. [11, Chapter 3.3]), it is straightforward to verify that X satisfies the small-ball condition (1.4) with respect to $\|\cdot\|_q$ for constants $\gamma = \gamma_q$ and $\delta = \delta_q$ that depend only on q .

Therefore, invoking Theorem 1.6, a typical realization of $\text{absconv}(X_1, \dots, X_N)$ contains $c(K_p(X))^\circ$ for $p = \alpha \log(eN/n)$. It remains to identify the floating body $K_p(X)$. To this end, observe that

$$K_p(X) \subset c_2(q) \left(\frac{n}{N}\right)^{\alpha/q} B_q^n. \quad (3.1)$$

Indeed, let $t \in K_p(X)$. By (F1), $\langle X, t \rangle$ has the same distribution as $\xi \|t\|_q$ and

$$\mathbb{P}\left(\xi \geq \frac{1}{\|t\|_q}\right) = \mathbb{P}(\langle X, t \rangle \geq 1) \leq \exp(-p) = \left(\frac{n}{eN}\right)^\alpha.$$

Since N/n is ‘large enough’, it follows that for M_q as in (F3), $\|t\|_q \leq 1/M_q$; indeed, otherwise $\mathbb{P}(\xi \geq M_q) \leq (n/(eN))^\alpha$ which is impossible when N/n is larger than a suitable constant. Now, by (F3),

$$c_q (\|t\|_q)^q \leq \mathbb{P}\left(\xi \geq \frac{1}{\|t\|_q}\right) \leq \left(\frac{n}{eN}\right)^\alpha,$$

implying that

$$\|t\|_q \leq c_2 \left(\frac{n}{eN}\right)^{\alpha/q},$$

where $c_2 = c_2(q) = c_q^{1/q}$. This establishes (3.1) and completes the proof of Theorem 1.8 by taking the polar. \blacksquare

3.2 Floating bodies and the unit ball of $L_p(X)$.

In order to get a better intuition on the role of the sets $K_p(X)$, let us consider a case in which X is a ‘reasonably nice’ random vector, in the sense that each $\langle X, t \rangle$ has sufficiently many moments and exhibits a weak kind of regularity. As we show next, the sets $K_p(X)$ are then equivalent to

$$B(L_p(X)) = \left\{ t \in \mathbb{R}^n : (\mathbb{E} |\langle X, t \rangle|^p)^{1/p} \leq 1 \right\}, \quad p \geq 1.$$

The polar body

$$Z_p(X) := B(L_p(X))^\circ \tag{3.2}$$

is called the L_p -centroid body of X . The fact that there is a connection between $K_p(X)$ and $B(L_p(X))$ is an immediate outcome of Markov’s inequality:

$$\mathbb{P}(\langle X, t \rangle \geq e \| \langle X, t \rangle \|_{L_p}) \leq \mathbb{P}(|\langle X, t \rangle|^p \geq e^p \| \langle X, t \rangle \|_{L_p}^p) \leq \exp(-p).$$

Therefore, if $\| \langle X, t \rangle \|_{L_p} \leq 1/e$ then $t \in K_p(X)$, i.e.,

$$\frac{1}{e} B(L_p(X)) \subset K_p(X). \tag{3.3}$$

In order to prove a reverse inequality one requires an additional regularity condition on X .

Definition 3.1. *The random vector X satisfies a regularity condition with constant D if for every $q \geq 2$ and every $t \in \mathbb{R}^n$,*

$$\| \langle t, X \rangle \|_{L_{2q}} \leq D \| \langle t, X \rangle \|_{L_q}. \tag{3.4}$$

Lemma 3.2. *Let X be a symmetric random vector for which (3.4) holds. Then, for every $p \geq c_2$,*

$$K_p(X) \subset 2B(L_{c_1 p}(X)),$$

where $c_1 = 1/(4 \log(4D/3))$ and $c_2 = \max\{2c_1, 2 \log(2)\}$.

Proof. Fix $t \in \mathbb{R}^n$. By the symmetry of X ,

$$\mathbb{P}(\langle X, t \rangle \geq 1) = \frac{1}{2} \mathbb{P}(|\langle X, t \rangle| \geq 1),$$

and invoking the Paley-Zygmund inequality (see, e.g.[11, Chapter 3.3]) yields, for any $q \geq 2$,

$$\begin{aligned} \mathbb{P} \left(|\langle X, t \rangle| \geq \frac{1}{2} \| \langle X, t \rangle \|_{L_q} \right) &\geq \left((1 - (1/2)^q) \frac{\| \langle X, t \rangle \|_{L_q}}{\| \langle X, t \rangle \|_{L_{2q}}} \right)^{2q} \geq \left(\frac{3}{4D} \right)^{2q} \\ &= \exp(-2q \log(4D/3)). \end{aligned} \tag{3.5}$$

Hence, if $q = c_1 p$ with $c_1 = c_1(D) = 1/(4 \log(4D/3))$ and $p \geq 2 \log(2)$ then

$$\mathbb{P} \left(\langle X, t \rangle \geq \frac{1}{2} \| \langle X, t \rangle \|_{L_q} \right) \geq \frac{1}{2} \exp(-q/(2c_1)) \geq \exp(-p/2 - p/2) = \exp(-p).$$

Hence, if $t \in K_p(X)$ then $\| \langle X, t \rangle \|_{L_q} \leq 2$, as claimed. ■

Remark 3.3. *Note that in order to prove that $K_p(X) \subset 2B(L_p(X))$ for a fixed value of p it suffices that X satisfies that $\| \langle X, t \rangle \|_{L_{2q}} \leq D \| \langle X, t \rangle \|_{L_q}$ for $q = c'p$.*

Log-concave random vectors

Let us give one generic example in which (3.4) holds and $K_p(X)$ is equivalent to $B(L_p(X))$. There are many other natural examples of random vectors that satisfy (3.4) (e.g., the Rademacher vector \mathcal{E} , thanks to Borell's hypercontractivity inequality [5]), but since the focus of this note is on random polytopes generated by a heavy-tailed random vectors we will not pursue this direction further.

A random vector is log-concave if it has a density f satisfying that for every x, y in its support and any $0 < \lambda < 1$, $f(\lambda x + (1 - \lambda)y) \geq f^\lambda(x)f^{1-\lambda}(y)$. The L_p -centroid bodies $Z_p(X)$ defined in (3.2) play a crucial role in the study of log-concave measures [27, 35]. For more information on log-concave random vectors we refer the reader to [7, 20].

Let X be a symmetric log-concave random vector that is non-degenerate, i.e., whose support is not contained in a proper subspace of \mathbb{R}^n . It follows from Borell's inequality [5] (see e.g. [20, Proposition 5.16]) that for every $t \in \mathbb{R}^n$ and $1 \leq p \leq q < \infty$,

$$\|\langle X, t \rangle\|_{L_p} \leq \|\langle X, t \rangle\|_{L_q} \leq 12 \frac{q}{p} \|\langle X, t \rangle\|_{L_p}. \quad (3.6)$$

Therefore, X satisfies the weak regularity condition (3.4) with constant $D = 24$, implying that $B(L_p(X)) \sim K_p(X)$. Further, by (3.5) with $q = 2$, X satisfies a small-ball condition with respect to the norm $\|t\|_X := (\mathbb{E}|\langle X, t \rangle|^2)^{1/2} = \|\Sigma^{1/2}t\|_2$ with constants $\gamma = 1/2$ and $\delta = (1/32)^4$. Here $\Sigma = \mathbb{E}XX^T$ is the covariance matrix of X , which is nonsingular by the non-degenerateness assumption on X so that $\|\cdot\|_X$ is actually a norm. Moreover, (3.6) also implies that X satisfies the L_r -condition for $r = 4$ with respect to $\|t\|_X$ with $L = 24$. Theorem 1.6 then leads to the following result.

Theorem 3.4. *Let X be a symmetric, non-degenerate, log-concave random vector. Let $0 < \alpha < 1$, set $N \geq c_0(\alpha)n$ and put $p = \alpha \log(eN/n)$. Then, with probability at least $1 - 2 \exp(-c_1 N^{1-\alpha} n^\alpha)$,*

$$\text{absconv}(X_1, \dots, X_N) \supset c_2 Z_p(X) \quad (3.7)$$

where c_2 is a universal constant.

Theorem 3.4 improves the main result from [10], which states that if X is an isotropic (which means that its covariance matrix Σ is the identity), log-concave random vector and Γ is the random matrix whose rows are X_1, \dots, X_N , then with probability at least $1 - 2 \exp(-c_1(\alpha)N^{1-\alpha}n^\alpha) - \mathbb{P}(\|\Gamma : \ell_2^n \rightarrow \ell_2^N\| \geq c\sqrt{N})$,

$$\text{absconv}(X_1, \dots, X_N) \supset c_2(\alpha)Z_p(X).$$

Thanks to the progress made in [1] in the study of random matrices with i.i.d. isotropic log-concave rows, it is known that

$$\mathbb{P}(\|\Gamma : \ell_2^n \rightarrow \ell_2^N\| \geq c\sqrt{N}) \leq \exp(-c'\sqrt{n}).$$

Therefore, the probability bound of the result in [10] is weaker than the one Theorem 3.4.

3.3 Stochastic domination

Up to this point, the examples focused on random vectors X for which $K_p(X)$ can either be studied directly, or is equivalent to a natural convex body. One way of extending the scope of the analysis of the random polytopes $\text{absconv}(X_1, \dots, X_N)$ is by comparing the floating bodies $K_p(X)$ that are associated with different random vectors. As it happens, this comparison is simply a way of coding *stochastic domination*.

Definition 3.5. *Let X and Y be centered random vectors in \mathbb{R}^n . The random vector X dominates Y with constants λ_1 and λ_2 if for every $t \in \mathbb{R}^n$ and every $u > 0$,*

$$\mathbb{P}(\langle X, t \rangle \geq u) \geq \lambda_1 \mathbb{P}(\langle Y, t \rangle \geq \lambda_2 u).$$

This means that if X dominates Y with constants λ_1 and λ_2 then

$$K_p(X) \subset \lambda_2 K_{p'}(Y) \quad (3.8)$$

for $p' = p - \log(1/\lambda_1)$.

It is well known that this notion of domination is well-suited for the study of random vectors with i.i.d. coordinates because it is preserved under tensorization:

Theorem 3.6. [24] *There are absolute constants c_1 and c_2 for which the following holds. Let x and y be symmetric random variables and assume that for every $u > 0$, $\mathbb{P}(x > u) \geq \lambda_1 \mathbb{P}(y \geq \lambda_2 u)$. Let x_1, \dots, x_n be independent copies of X and set y_1, \dots, y_n to be independent copies of y . Then $X = (x_i)_{i=1}^n$ dominates $Y = (y_i)_{i=1}^n$ with constants $c_1 \lambda_1$ and $c_2 \lambda_2$.*

Theorem 3.6 leads to many structural results on $\text{absconv}(X_1, \dots, X_N)$ for vectors with i.i.d. coordinates, by comparing x to a canonical random variable like a Rademacher random variable (i.e., a symmetric, $\{-1, 1\}$ -valued random variable) or to the standard Gaussian random variable.

Observe that if x is a symmetric random variable that satisfies $\mathbb{P}(|x| \geq \gamma_0) \geq \delta_0$ then we have

$$\mathbb{P}(x \geq u) \geq \delta_0 \mathbb{P}(\varepsilon > u/\gamma_0),$$

where ε is a Rademacher random variable. Hence, from Theorem 3.6, we get that if x_1, \dots, x_n are independent copies of x and $X = (x_i)_{i=1}^n$, then X dominates the Rademacher vector \mathcal{E} with constants λ_1 and λ_2 that depend only on γ_0 and δ_0 . As a result, by (3.8),

$$K_p(X) \subset \lambda_2 K_{p'}(\mathcal{E}),$$

where $p' = p - \log(1/\lambda_1)$. Thanks to the characterization of $K_p(\mathcal{E})$ and Theorem 1.6 one immediately recovers Theorem 1.2 as well as the main result from [19].

Theorem 3.7. *Let x be a symmetric random variable that satisfies $\mathbb{E}x^2 = 1$ and set x_1, \dots, x_n to be independent copies of x and put $X = (x_i)_{i=1}^n$. If there are constants γ and δ such that $\mathbb{P}(|x| \geq \gamma) \geq \delta$, then for $N \geq c_0 n$, with probability at least $1 - 2 \exp(-c_1 N^{1-\alpha} n^\alpha)$,*

$$\text{absconv}(X_1, \dots, X_N) \supset c_2 (B_\infty^n \cap \sqrt{\alpha \log(eN/n)} B_2^n);$$

here c_0 depends on α, γ and δ , c_2 depends on γ and δ , and c_1 is an absolute constant.

The result can be pushed much further. The fact that X has i.i.d. coordinates can be relaxed to an unconditional assumption. Moreover, X need not have a covariance, as in fact, Assumption 1.5 suffices to get the desired conclusion.

Definition 3.8. *A random vector $X = (x_i)_{i=1}^n$ is unconditional if for every $(\varepsilon_i)_{i=1}^n \in \{-1, 1\}^n$, $(x_i)_{i=1}^n$ has the same distribution as $(\varepsilon_i x_i)_{i=1}^n$.*

Theorem 3.9. *For every $0 < \delta < 1$ there is a constant $c = c(\delta)$ such that the following holds. Let X be an unconditional random vector that satisfies the small-ball condition with constants γ and δ . Then, for any $p > c_0(\delta) = 4 \log(8/\delta) + \log(4)$,*

$$K_p(X) \subset \frac{c(\delta)}{\gamma} K_p(\mathcal{E}).$$

In particular, if X satisfies Assumption 1.5 and $N \geq c_0(\alpha, \delta, r, L/\gamma)n$, then with probability at least $1 - 2 \exp(-c_1 N^{1-\alpha} n^\alpha)$,

$$\text{absconv}(X_1, \dots, X_N) \supset \frac{1}{2} (K_p(X))^\circ \supset c'(\delta) \gamma (B_\infty^n \cap \sqrt{\alpha \log(eN/n)} B_2^n).$$

The proof of Theorem 3.9 is based on contraction inequalities for the Rademacher random vector (see, e.g. [25]): if $|a_i| \leq |b_i|$ for $1 \leq i \leq n$ then for every $p \geq 1$,

$$\left(\mathbb{E} \left| \sum_{i=1}^n \varepsilon_i a_i \right|^p \right)^{1/p} \leq \left(\mathbb{E} \left| \sum_{i=1}^n \varepsilon_i b_i \right|^p \right)^{1/p}, \quad (3.9)$$

and for every $u > 0$,

$$\mathbb{P} \left(\left| \sum_{i=1}^n \varepsilon_i a_i \right| \geq u \right) \leq 2 \mathbb{P} \left(\left| \sum_{i=1}^n \varepsilon_i b_i \right| \geq u \right). \quad (3.10)$$

We also require Borell's hypercontractivity inequality [5]: for every $t \in \mathbb{R}^n$ and $q > p > 1$,

$$\| \langle \mathcal{E}, t \rangle \|_{L_q} \leq \frac{q-1}{p-1} \| \langle \mathcal{E}, t \rangle \|_{L_p}. \quad (3.11)$$

Proof of Theorem 3.9. The second part of the theorem is an immediate outcome of the first part, Theorem 1.6, and the fact that X satisfies Assumption 1.5. To establish the first part, let us show that if $X = (x_1, \dots, x_n)$ is an unconditional random vector and there are $\gamma, \delta > 0$ such that for every $1 \leq i \leq n$,

$$\mathbb{P}(|x_i| \geq \gamma) \geq \delta, \quad (3.12)$$

then $K_p(X) \subset \frac{c(\delta)}{\gamma} K_p(\mathcal{E})$. Note that for this part of the theorem, X does not need to satisfy the small-ball condition 1.4 for every direction, but rather only for coordinate directions.

Let $t \in K_p(X)$. Since X is unconditional and symmetric, it holds that

$$\frac{1}{2} \mathbb{P}_{X \otimes \varepsilon} \left(\left| \sum_{i=1}^n \varepsilon_i |x_i| |t_i| \right| \geq 1 \right) = \mathbb{P}_X \left(\sum_{i=1}^n x_i t_i \geq 1 \right) \leq \exp(-p). \quad (3.13)$$

Let $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ be the truncation at level γ , that is,

$$\phi(z) = \begin{cases} |z| & \text{if } |z| \leq \gamma, \\ \gamma & \text{if } |z| > \gamma, \end{cases}$$

and set $Z_t = \sum_{i=1}^n \varepsilon_i \phi(x_i) |t_i|$. Since $\phi(z) \leq |z|$ the contraction principle (3.10) yields, for every $(x_i)_{i=1}^n \in \mathbb{R}^n$,

$$\begin{aligned} \mathbb{P}(|Z_t| \geq 1) &= \mathbb{E}_X \mathbb{P}_\varepsilon(|Z_t| \geq 1) \leq 2 \mathbb{E}_X \mathbb{P}_\varepsilon \left(\left| \sum_{i=1}^n \varepsilon_i |x_i| |t_i| \right| \geq 1 \right) = 4 \mathbb{P}_X \left(\sum_{i=1}^n x_i t_i \geq 1 \right) \\ &\leq 4 \exp(-p). \end{aligned} \quad (3.14)$$

Observe that for every $1 \leq i \leq n$,

$$\mathbb{E}_X \phi(x_i) \geq \gamma \mathbb{P}(|x_i| \geq \gamma) \geq \gamma \delta,$$

where the last inequality follows from the small ball assumption (3.12). This observation implies that for any $q > 1$,

$$\begin{aligned} (\mathbb{E} |Z_t|^{2q})^{1/2q} &\leq \gamma \left(\mathbb{E}_\varepsilon \left(\sum_{i=1}^n \varepsilon_i t_i \right)^{2q} \right)^{1/2q} \leq \frac{2q-1}{q-1} \gamma \left(\mathbb{E}_\varepsilon \left| \sum_{i=1}^n \varepsilon_i t_i \right|^q \right)^{1/q} \\ &\leq \frac{2q-1}{q-1} \delta^{-1} \left(\mathbb{E}_\varepsilon \left| \sum_{i=1}^n \varepsilon_i \mathbb{E}_X \phi(x_i) |t_i| \right|^q \right)^{1/q} \leq \frac{2q-1}{q-1} \delta^{-1} (\mathbb{E} |Z_t|^q)^{1/q}. \end{aligned} \quad (3.15)$$

Here, the first inequality used that $\phi(z) \leq |z|$ as well as the contraction principle (3.9), the second inequality is based on the hypercontractivity inequality for the Rademacher vector (3.11) and the

last inequality follows from Jensen's inequality. Therefore, by the Paley-Zygmund inequality (as in, e.g., [11, Chapter 3.3]), we have that

$$\begin{aligned} \mathbb{P}\left(|Z_t| \geq (\mathbb{E}|Z_t|^q/2)^{1/q}\right) &\geq \left(\frac{1}{2} \frac{(\mathbb{E}|Z_t|^q)^{1/q}}{(\mathbb{E}|Z_t|^{2q})^{1/2q}}\right)^{2q} \geq \left(\frac{\delta(q-1)}{2(2q-1)}\right)^{2q} \\ &= \exp\left(-2q \log\left(\frac{4q-2}{(q-1)\delta}\right)\right). \end{aligned} \quad (3.16)$$

If q is such that

$$2q \log\left(\frac{4q-2}{(q-1)\delta}\right) < p - \log(4), \quad (3.17)$$

then it follows that $\mathbb{E}|Z_t|^q < 2$ because otherwise (3.16) would be in contraction to (3.14). Before elaborating on the implication of $\mathbb{E}|Z_t|^q < 2$, let us discuss the particular choice

$$q = \frac{p - \log(4)}{2 \log(8/\delta)}.$$

Since $p > 4 \log(8/\delta) + \log(4)$ by assumption, it follows that $q > 2$ and $q - 1 > q/2$ so that $(4q - 2)/(q - 1) < 8 - 4/q < 8$ and

$$2q \log\left(\frac{4q-2}{(q-1)\delta}\right) < 2q \log(8/\delta) = p - \log(4),$$

so that (3.17) is satisfied. Note that since $q > 2$ and $p > 4 \log(8/\delta) + \log(4)$,

$$\begin{aligned} C_{p,q} &:= \frac{p-1}{q-1} < \frac{p-1}{q/2} = \frac{p-1}{p-\log(4)} 4 \log(8/\delta) = \left(1 + \frac{\log(4)-1}{p-\log(4)}\right) 4 \log(8/\delta) \\ &< 4 \log(8/\delta) + \log(4) - 1 =: C_\delta. \end{aligned}$$

By hypercontractivity combined with (3.15) (starting with the term after the second inequality in the first line) and the observation that $(\mathbb{E}|Z_t|^q)^{1/q} < 2^{1/q} < \sqrt{2}$, we obtain

$$\left(\mathbb{E}_\varepsilon \left| \sum_{i=1}^n \varepsilon_i t_i \right|^p\right)^{1/p} \leq \frac{p-1}{q-1} \left(\mathbb{E}_\varepsilon \left| \sum_{i=1}^n \varepsilon_i t_i \right|^q\right)^{1/q} \leq C_{p,q} \frac{1}{\gamma \delta} (\mathbb{E}|Z_t|^q)^{1/q} < C_\delta \frac{\sqrt{2}}{\gamma \delta} =: C(\delta, \gamma).$$

Markov's inequality gives

$$\mathbb{P}_\varepsilon \left(\sum_{i=1}^n \varepsilon_i t_i \geq eC(\delta, \gamma) \right) \leq \exp(-p).$$

Hence, for

$$c(\delta) = \frac{\sqrt{2}e(4 \log(8/\delta) + \log(4/e))}{\delta}$$

it holds that $c(\delta)/\gamma = eC(\delta, \gamma)$ and $K_p(X) \subset \frac{c(\delta)}{\gamma} K_p(\mathcal{E})$ as claimed. \blacksquare

A Concentration inequality for VC classes of functions

We prove Theorem 2.5 in this section, basically following [28] but with a simplification (avoiding the use of [28, Lemma 3.6] due to Talagrand [37]). The main tool is the following version of Talagrand's concentration inequality [38] due to Bousquet [6], see also [16, Theorem 8.42], which features explicit and small constants.

Theorem A.1. Let \mathcal{G} be a set of functions $g : \mathbb{R}^n \rightarrow \mathbb{R}$. Let X_1, \dots, X_N be independent random vectors in \mathbb{R}^n such that $\mathbb{E}g(X_\ell) = 0$ and $|g(X_\ell)| \leq K$ almost surely for all $\ell = 1, \dots, N$ and for all $g \in \mathcal{G}$ for some constant $K > 0$. Introduce

$$Z = \sup_{g \in \mathcal{G}} \left| \sum_{\ell=1}^N g(X_\ell) \right|. \quad (\text{A.1})$$

Let $\sigma_\ell^2 > 0$ such that $\mathbb{E}[g(X_\ell)^2] \leq \sigma_\ell^2$ for all $g \in \mathcal{G}$ and $\ell = 1, \dots, N$. Then, for all $t > 0$,

$$\mathbb{P}(Z \geq \mathbb{E}Z + t) \leq \exp\left(-\frac{t^2/2}{\sigma_{\mathcal{G}}^2 + 2K\mathbb{E}Z + tK/3}\right), \quad (\text{A.2})$$

where $\sigma_{\mathcal{G}}^2 = \sum_{\ell=1}^N \sigma_\ell^2$.

In the situation of Theorem 2.5, we consider $\mathcal{G} = \{g = f - \mathbb{E}[f(X)] : f \in \mathcal{F}\}$, so that $\mathbb{E}g(X) = 0$ and $|g(X)| \leq 1 =: K$ almost surely for all $g \in \mathcal{G}$. Moreover, $\sigma_\ell^2 \leq \sigma^2 = \sup_{f \in \mathcal{F}} \mathbb{E}[f(X_\ell)^2]$ so that $\sigma_{\mathcal{G}}^2 \leq N\sigma^2$. It remains to estimate $\mathbb{E}Z$.

Symmetrization, see e.g. [25, Lemma 6.3], and Dudley's inequality in the form of [16, Theorem 8.23] yield, for a Rademacher sequence $\varepsilon_1, \dots, \varepsilon_N$ independent of X_1, \dots, X_N ,

$$\begin{aligned} \mathbb{E}Z &= \mathbb{E} \sup_{f \in \mathcal{F}} \left| \sum_{j=1}^N f(X_j) - \mathbb{E}[f(X_j)] \right| \leq 2 \mathbb{E} \sup_{f \in \mathcal{F}} \left| \sum_{j=1}^N \varepsilon_j f(X_j) \right| \\ &\leq 8\sqrt{2N} \mathbb{E}_X \int_0^{\Delta_X(\mathcal{F})/2} \sqrt{\log(2\mathcal{N}(\mathcal{F}, d_{X,2}, u))} du, \end{aligned}$$

where the metric $d_{X,2}$ is given as

$$d_{X,2}(f, g) = \left(\frac{1}{N} \sum_{j=1}^N (f(X_j) - g(X_j))^2 \right)^{1/2} = \left(\mathbb{E}_\varepsilon \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N \varepsilon_j (f(X_j) - g(X_j)) \right)^2 \right)^{1/2}$$

and $\mathcal{N}(\mathcal{F}, d_{X,2}, u)$ denote the covering numbers of \mathcal{F} , i.e., the minimal number of balls of radius u in the metric $d_{X,2}$ required to cover \mathcal{F} and

$$\Delta_X(\mathcal{F}) := \left(\sup_{f \in \mathcal{F}} \frac{1}{N} \sum_{j=1}^N f(X_j)^2 \right)^{1/2} = \underbrace{\left(\sup_{f \in \mathcal{F}} \frac{1}{N} \sum_{j=1}^N f(X_j)^2 \right)^{1/2}}_{=: Y},$$

where we have used that f takes only values in $\{0, 1\}$ in the equality. It follows from Haussler's theorem [21] and the fact that the functions in \mathcal{F} are $\{0, 1\}$ -valued (so that $\|f - g\|_{L_2(\mu)}^2 = \|f - g\|_{L_1(\mu)}$ for any probability measure μ) that the covering numbers can be estimated via the VC-dimension d as

$$\mathcal{N}(\mathcal{F}, d_{X,2}, t) \leq e(d+1)(2e)^d t^{-2d}.$$

Plugging this into our estimate of $\mathbb{E}Z$ above and noting that $(2e(d+1))^{1/2d}$ takes the maximum for $d = 1$, so that $(2e(d+1))^{1/2d} \leq 2\sqrt{e}$ for all $d \geq 1$, gives

$$\mathbb{E}Z \leq 8\sqrt{2N} \mathbb{E} \int_0^{\sqrt{Y}/2} \sqrt{\log(2e(d+1)(2e/u)^{2d})} du \leq 16\sqrt{Nd} \mathbb{E} \int_0^{\sqrt{Y}/2} \sqrt{\log(2e\sqrt{2}/u)} du$$

We use the Cauchy-Schwarz inequality to estimate the integral

$$\int_0^\alpha \sqrt{\log(\gamma/u)} du \leq \sqrt{\int_0^\alpha 1 du} \sqrt{\int_0^\alpha \log(\gamma/u) du} = \sqrt{\alpha} \sqrt{\gamma \int_{\gamma/\alpha}^\infty \log(t) t^{-2} dt} = \alpha \sqrt{\log(e\gamma/\alpha)}.$$

Setting $\alpha = \sqrt{Y}/2$ and $\gamma = 2e\sqrt{2}$, noting that $t \mapsto \sqrt{t \log(2e\sqrt{2}/t)}$ is concave and applying Jensen's inequality gives

$$\mathbb{E}Z \leq 16\sqrt{Nd}\mathbb{E}\sqrt{\frac{Y}{4} \log\left(\frac{2e^2\sqrt{2}}{Y/4}\right)} \leq 16\sqrt{Nd}\sqrt{\frac{\mathbb{E}Y}{4} \log\left(\frac{2e^2\sqrt{2}}{\mathbb{E}Y/4}\right)}.$$

Now observe that by the triangle inequality and since each f takes values in $\{0, 1\}$,

$$Y \leq Z/N + \sup_{f \in \mathcal{F}} \mathbb{E}f(X) = Z/N + \sup_{f \in \mathcal{F}} \mathbb{E}f^2(X) \leq Z/N + \sigma^2.$$

Since $t \mapsto \sqrt{t \log(2e^2\sqrt{2}/t)}$ is increasing, this yields

$$\mathbb{E}Z \leq 8\sqrt{Nd}\sqrt{\frac{\mathbb{E}Z + N\sigma^2}{N} \log\left(\frac{8e^2\sqrt{2}}{\mathbb{E}Z/N + \sigma^2}\right)} \leq 8\sqrt{d}\sqrt{(\mathbb{E}Z + N\sigma^2) \log\left(\frac{8e^2\sqrt{2}}{\sigma^2}\right)}.$$

Setting $Q := 8\sqrt{d \log\left(\frac{8e^2\sqrt{2}}{\sigma^2}\right)}$ and squaring leads to the inequality $(\mathbb{E}Z)^2 \leq Q^2(\mathbb{E}Z + N\sigma^2)$ so that

$$\begin{aligned} \mathbb{E}Z &\leq Q^2/2 + \sqrt{Q^2N\sigma^2 + Q^4/4} \leq Q^2 + Q\sqrt{N}\sigma \\ &= 64d \log\left(\frac{8e^2\sqrt{2}}{\sigma^2}\right) + 8\sigma\sqrt{Nd \log\left(\frac{8e^2\sqrt{2}}{\sigma^2}\right)} = NR. \end{aligned}$$

It follows from (A.2) that

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} \left| \sum_{j=1}^N (f(X_j) - \mathbb{E}f(X_j)) \right| \geq NR + t\right) \leq \exp\left(-\frac{t^2/2}{N\sigma^2 + 2NR + t/3}\right),$$

which is equivalent to the statement of Theorem (2.5).

B Sparse recovery

We begin this section with an outline of the proof of how the ℓ_1 -quotient property leads to (1.11). The null space property of A of order s with constant $\rho < 1$ requiring that

$$\sum_{j \in S} |v_j| \leq \rho \sum_{j \in S^c} |v_j| \quad \text{for all } v \in \ker A \setminus \{0\} \text{ and all } S \subset \{1, \dots, N\} \text{ with } \#S = s, \quad (\text{B.1})$$

implies by [16, Theorem 4.12] that the solution x^\sharp of equality constrained ℓ_1 -minimization (1.7) with $y = Ax$ satisfies

$$\|x - x^\sharp\|_1 \leq \frac{2(1+\rho)}{1-\rho} \sigma_s(x)_1. \quad (\text{B.2})$$

If $y = Ax + w$, then the ℓ_1 -quotient property yields the existence of $v \in \mathbb{R}^n$ satisfying (1.10), so that we can write $y = A(x + v)$. The error bound (B.2) then leads to

$$\begin{aligned} \|x^\sharp - x\|_1 &\leq \frac{2(1+\rho)}{1-\rho} \inf_{z: \|z\|_0 \leq s} \|x + v - z\|_1 \leq \frac{2(1+\rho)}{1-\rho} \left(\inf_{z: \|z\|_0 \leq s} \|x - z\|_1 + \|v\|_1 \right) \\ &\leq \frac{2(1+\rho)}{1-\rho} \sigma_s(x)_1 + \frac{2(1+\rho)}{c_2(1-\rho)} \|w\|_p, \end{aligned}$$

which is (1.11).

Next, let us turn to the example of noise-blind recovery when the measurement matrix has i.i.d. columns, selected according to the random vector X , which has i.i.d. coordinates, distributed according to the (L_2 -normalized) Student- t entries with $d = 2 \log N$ degrees of freedoms. In particular, the first moments of each coordinate are equivalent to that of a Gaussian random variable: for any $q \leq \log N$, $c_1 \|g\|_{L_q} \leq \|\xi\|_{L_q} \leq c_2 \|g\|_{L_q}$. This example is particularly interesting because it was recently shown (see, e.g., [31] and [13, Example 9]) that the corresponding random matrix satisfies the null space property (B.1) of order s with high probability as long as $s \sim n/\log(eN/n)$. In addition, numerical tests in [13] show that this random matrix behaves precisely like a Gaussian random matrix in practical sparse recovery problems. However, the ℓ_1 -quotient property of a Student- t matrix was previously open.

It is straightforward to verify that for any $q \leq \log N$ and every $w \in \mathbb{R}^n$, $\|\langle X, w \rangle\|_{L_q} \sim \|\langle G, w \rangle\|_{L_q}$. Moreover, setting $p = \alpha \log(eN/n)$, the results in Section 3.2 imply that

$$K_p(X) \sim B(L_p(X)) \sim B(L_p(G)) \sim \sqrt{\log(eN/n)} B_2^n;$$

therefore,

$$\|w\|_p \sim_\alpha \|w\|_2 \sqrt{\log(eN/n)}.$$

The general error estimate (1.11) and Theorem 1.6 together with $s \sim n/\log(eN/n)$ lead to

$$\|x^\# - x\|_1 \lesssim \sigma_s(x)_1 + \sqrt{\log(eN/n)} \|w\|_2 \sim \sigma_s(x)_1 + \|w\|_2 \sqrt{\frac{s}{n}}. \quad (\text{B.3})$$

Note that (B.3) yields the same error estimate as (1.9) (up to absolute constants), but while (1.9) requires an a priori threshold for the noise level, (B.3) does not, and the error depends on the true noise level $\|w\|_2$ rather than a potentially pessimistic upper bound. We refer to [22] for more results in this direction and corresponding numerical experiments.

Acknowledgements

HR would like to thank the Isaac Newton Institute for Mathematical Science for support and hospitality during the program Approximation, Sampling and Compression in Data Science when work on this paper was undertaken. This work was supported by EPSRC Grant Number EP/R014604/1. OG thanks the funding of the Fondation Simone et Cino Del Duca for the project "Phénomènes en grande dimension".

FK was supported by the German Science Foundations in the context of an Emmy Noether Junior Research Group (KR 4512/1-1)

AMS 2010 Classification: primary: 52A22, 46B06, 60B20, 65K10 secondary: 52A23, 46B09, 15B52.

Keywords: Random polytopes, random matrices, heavy tails, small ball probability, compressed sensing, ℓ_1 -quotient property.

References

- [1] R. Adamczak, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. AMS*, 23(2):535–561, 2010.
- [2] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. *Constr. Approx.*, 34:61–88, 2011.
- [3] I. Bárány. Random polytopes, convex bodies, and approximation. In *Stochastic geometry*, volume 1892 of *Lecture Notes in Math.*, pages 77–118. Springer, Berlin, 2007.

- [4] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *J. ACM*, 36:929–965, 1989.
- [5] C. Borell. Convex measures on locally convex spaces. *Ark. Mat.*, 12(1):239–252, 1974.
- [6] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math.*, 334(6):495–500, 2002.
- [7] S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou. *Geometry of isotropic convex bodies*. American Mathematical Society, Providence, Rhode Island, 2014.
- [8] S. Brugiapaglia and B. Adcock. Robustness to unknown error in sparse regularization. *IEEE Trans. Inform. Theory*, 64(10):6638–6661, 2018.
- [9] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interactions between compressed sensing random matrices and high dimensional geometry*, volume 37 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2012.
- [10] N. Dafnis, A. Giannopoulos, and A. Tsolomitis. Asymptotic shape of a random polytope in a convex body. *J. Funct. Anal.*, 257(9):2820 – 2839, 2009.
- [11] V. De la Peña and E. Giné. *Decoupling. From dependence to independence. Randomly stopped processes, U-statistics and processes, martingales and beyond*. Springer Science & Business Media, 2012.
- [12] R. DeVore, G. Petrova, and P. Wojtaszczyk. Instance-optimality in probability with an ℓ_1 -minimization decoder. *Appl. Comput. Harmon. Anal.*, 27(3):275 – 288, 2009.
- [13] S. Dirksen, G. Lecué, and H. Rauhut. On the gap between restricted isometry properties and sparse recovery conditions. *IEEE Trans. Inform. Theory*, 64(8):5478–5487, 2018.
- [14] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [15] S. Foucart. Stability and robustness of ℓ_1 -minimizations with Weibull matrices and redundant dictionaries. *Linear Algebra Appl.*, 441:4 – 21, 2014.
- [16] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. Springer New York, 2013.
- [17] A. Giannopoulos and M. Hartzoulaki. Random spaces generated by vertices of the cube. *Discrete Comput. Geom.*, 28(2):255–273, 2002.
- [18] E. D. Gluskin. Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces. *Sb. Math.*, 64(1):85, 1989.
- [19] O. Guédon, A. Litvak, and K. Tatarko. Random polytopes obtained by matrices with heavy tailed entries. *To appear in Commun. Contemp. Math.*, 2019.
- [20] O. Guédon, P. Nayar, and T. Tkocz. Concentration inequalities and geometry of convex bodies. In *Analytical and probabilistic methods in the geometry of convex bodies*, volume 2 of *IMPAN Lect. Notes*, pages 9–86. Polish Acad. Sci. Inst. Math., Warsaw, 2014.
- [21] D. Haussler. Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. *J. Combinat. Theory, Ser. A*, 69:217–232, 1995.
- [22] F. Krahmer, C. Kümmeler, and H. Rauhut. A quotient property for matrices with heavy-tailed entries and its application to noise-blind compressed sensing. *Preprint arXiv:1806.04261*, 2018.
- [23] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.*, 67(11):1877–1904, 2014.

- [24] S. Kwapien and W. A. Woyczyński. *Random series and stochastic integrals: single and multiple*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1992.
- [25] M. Ledoux and M. Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.
- [26] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.*, 195(2):491 – 523, 2005.
- [27] E. Lutwak and G. Zhang. Blaschke-Santaló inequalities. *J. Diff. Geom.*, 45:1–16, 1997.
- [28] S. Mendelson. A few notes on statistical learning theory. In S. Mendelson and A. Smola, editors, *Advanced Lectures on Machine Learning*, volume 2600, pages 1–40. Springer, 2003.
- [29] S. Mendelson. Learning without concentration. *J. ACM*, 62(3):21:1–21:25, June 2015.
- [30] S. Mendelson. On the geometry of random polytopes. *arXiv preprint arXiv:1902.01664*, 2019.
- [31] S. Mendelson and G. Lecué. Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc.*, 19(3):881–904, 2017.
- [32] M. Meyer and S. Reisner. Characterizations of affinely-rotation-invariant log-concave measures by section-centroid location. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 145–152. Springer, Berlin, 1991.
- [33] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of Machine Learning*. MIT Press, 2012.
- [34] S. J. Montgomery-Smith. The distribution of Rademacher sums. *Proc. Amer. Math. Soc.*, 109(2):517–522, 1990.
- [35] G. Paouris. Concentration of mass on convex bodies. *Geom. Funct. Anal.*, 16(5):1021–1049, 2006.
- [36] C. Schütt and E. Werner. The convex floating body. *Math. Scand.*, 66(2):275–290, 1990.
- [37] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1):28–76, 1994.
- [38] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.
- [39] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probab. Appl.*, 16:264–280, 1971.
- [40] P. Wojtaszczyk. Stability and instance optimality for Gaussian measurements in compressed sensing. *Found. Comput. Math.*, 10(1):1–13, 2010.