On the Banach–Mazur distance to cross-polytope

Konstantin Tikhomirov*

May 23, 2018

Abstract

Let $n \geq 3$, and let B_1^n be the standard n–dimensional cross-polytope (i.e. the convex hull of standard coordinate vectors and their negatives). We show that there exists a symmetric convex body $\mathcal{G}_m$ in $\mathbb{R}^n$ such that the Banach–Mazur distance $\mathrm{d_{BM}}(B_1^n, \mathcal{G}_m)$ satisfies $\mathrm{d_{BM}}(B_1^n, \mathcal{G}_m) \geq n^{5/9} \log^{-C} n$, where $C > 0$ is a universal constant. The body $\mathcal{G}_m$ is obtained as a typical realization of a random polytope in $\mathbb{R}^n$ with $2m := 2n^C$ vertices (for a large constant C). The result improves upon an earlier estimate of S. Szarek which gives $\mathrm{d_{BM}}(B_1^n, \mathcal{G}_m) \geq cn^{1/2} \log n$ (with a different choice of m). This shows in a strong sense that the cross-polytope (or the cube $[-1, 1]^n$) cannot be an "approximate" center of the Minkowski compactum.

1 Introduction

The Minkowski (or the Banach–Mazur) compactum $\mathcal{M}_n$ is defined as the collection of all origin-symmetric n–dimensional convex bodies equipped with the distance function

$$\mathrm{d_{BM}}(K, L) := \inf \big\{ d \geq 1 : \; \exists T \in \mathrm{GL_n}(\mathbb{R}) \text{ such that } K \subset T(L) \subset dK \big\}$$

(in this note, we do not consider non-symmetric bodies). The classical theorem of F. John [6] asserts that $\mathrm{d_{BM}}(K, B_2^n) \leq \sqrt{n}$ for all $K \in \mathcal{M}_n$, where B_2^n is the standard Euclidean ball. The question of estimating $\sup\limits_{K \in \mathcal{M}_n} \mathrm{d_{BM}}(K, B_\infty^n)$ (or, equivalently, $\sup\limits_{K \in \mathcal{M}_n} \mathrm{d_{BM}}(K, B_1^n)$) has attracted considerable attention of researchers. Here and in what follows, by B_p^n ($1 \leq p \leq \infty$) we denote the unit ball of space ℓ_p^n; in particular, $B_\infty^n = [-1, 1]^n$ and $B_1^n = \mathrm{conv}\, \{\pm e_1, \ldots, \pm e_n\}$, where $e_1, e_2, \ldots, e_n$ is the standard vector basis in $\mathbb{R}^n$.

Currently best upper bound for the quantity $\sup\limits_{K \in \mathcal{M}_n} \mathrm{d_{BM}}(K, B_1^n)$ is $Cn^{5/6}$ due to A. Giannopoulos [4] (we refer to earlier results of J. Bourgain–S. Szarek [1] and S. Szarek–M. Talagrand [13] giving upper estimates $n \exp(-c\sqrt{\log n})$ and $Cn^{7/8}$, respectively, as well as to a result of P. Youssef [14] for estimate with improved constant $2n^{5/6}$). The connection of the problem with the property of *restricted invertibility* of matrices and the proportional Dvoretzky–Rogers factorization has been intensively explored in literature; we refer, among others, to papers [1, 13, 4, 14, 3, 10, 11].

*Princeton University, NJ; email: kt12@math.princeton.edu. The research is partially supported by the Simons Foundation.

For the lower bound, the only available result up to this writing, due to S. Szarek, asserts that $\sup_{K\in\mathcal{M}_n} \mathrm{d_{BM}}(K,B_1^n)\geq c\sqrt{n}\log n$ for a universal constant $c>0$ [12]. The proof of the lower estimate in [12] involves two crucial ingredients. First is construction of a family of random polytopes which, together with an ε–net argument and some probabilistic relations, reduces the problem to studying relative positions of independent Gaussian vectors in $\mathbb{R}^n$. In a different form, such construction was first used in context of geometric functional analysis by E. Gluskin [5] and allowed him to solve a crucial problem of estimating the diameter of the Minkowski compactum (we refer to survey [8] for more information; see also [7] for more recent applications of Gluskin's construction). The second element of Szarek's proof is an estimate of singular values of the standard Gaussian matrix which can be viewed as a non-asymptotic analog of the Marchenko–Pastur law [9] for the spectrum of sample covariance matrices, with very strong probability bounds.

The result of S. Szarek shows that the cross-polytope (or the cube) cannot be a "center" of the Minkowski compactum similar to the Euclidean ball. However, the result leaves open a possibility that the distance of any convex body to the cube is bounded above by the square root of dimension times a polylogarithmic multiple. A conjecture of A. Naor (personal communication) that the bound obtained in [12] is suboptimal, is confirmed in the main result of this note (see also [13, remark 2] for related discussion).

Theorem A. *Let $n\geq 3$ and let $m:=n^3$. Further, let $\mathcal{G}_m$ be the random symmetric polytope constructed as the convex hull of $2m$ vectors $\pm G_1,\pm G_2,\dots,\pm G_m$ where $G_1,G_2,\dots,G_m$ are independent standard Gaussian ($N(0,\mathrm{Id_n})$) vectors in $\mathbb{R}^n$. Then with positive probability $\mathrm{d_{BM}}(\mathcal{G}_m,B_1^n)\geq n^{5/9}\log^{-C}n$. Here, $C>0$ is a universal constant.*

The starting point of our proof is the same as in the S. Szarek's work [12]: we construct Gluskin's random polytope $\mathcal{G}_m$ to estimate its Banach–Mazur distance to an n–dimensional cross-polytope. However, instead of working with singular values we develop a combination of geometric and probabilistic arguments to "directly" estimate the Gaussian measure of cross-polytopes inscribed into a given realization of Gluskin's polytope.

The structure of the paper is the following: in Section 2 we recall those elements of Szarek's construction [12] that are also used in the present paper. In Section 3 we give a high-level overview of our strategy, which comprises two essential parts: constructing a special event of probability close to one which catches the geometric properties useful for us and estimating the Gaussian measure of cross-polytopes from a special class. The event is constructed in Section 4, while the Gaussian measure of cross-polytopes is computed in Section 5. Finally, in Section 6 we choose parameters and complete the proof.

2 Preliminaries

Let us start with basic notation. Given a finite set I, we denote its cardinality by $|I|$. The convex hull of a set of points S in a linear space is denoted by $\mathrm{conv}\,(S)$. Given a matrix A, its columns are denoted by $\mathrm{col}_1(A),\mathrm{col}_2(A),\dots$ Given a vector v, let $\mathrm{supp}\,v$ be its support. For two convex bodies K and L, $K+L$ denotes their Minkowski sum. Further, for any set K in $\mathbb{R}^n$ and an $m\times n$ matrix A, by $A(K)$ we denote the linear image of K in $\mathbb{R}^m$. Let $(\Omega, \mathbb{P})$ be a probability space. For any random variable/vector/matrix ξ on Ω, by $\xi(\omega)$ we denote the realization of ξ at a point $\omega \in \Omega$.

We introduce several global parameters:

$$\text{A large integer } n;\text{ a number } m \geq n^2;\ \varepsilon \in (0, 1/2] \text{ and } \rho \geq 1.$$

In what follows, m will be responsible for the geometry of our random polytope (we will take $2m$ to be the number of generating vectors in Gluskin's polytope in $\mathbb{R}^n$); ε will be used to define a net (discretization) on a set of linear operators; ρ will serve as a lower bound for the Banach–Mazur distance of the Gluskin polytope to a cross-polytope.

We will impose additional assumptions on the parameters in various statements below; the parameters will be explicitly chosen at the end of the note at the optimization stage.

Let G_i, $i \leq m$ be jointly independent standard Gaussian vectors in $\mathbb{R}^n$, and Γ be the $n \times m$ standard Gaussian matrix with columns G_i. We define the random symmetric convex body $\mathcal{G}_m$ as the convex hull of $\pm G_i$, $i \leq m$. These random convex polytopes, introduced to high-dimensional convex geometry by E. Gluskin [5] (the original definition was slightly different), turned out extremely useful.

2.1 Discretization

All observations in this subsection, up to minor modifications, repeat those from [12].

To show that there is a symmetric convex set with the Banach–Mazur distance to B_1^n at least ρ, it is sufficient to show that

$$\mathbb{P}\big\{\text{There is a symmetric cross-polytope } P \text{ with } \rho\,\mathcal{G}_m \supset \rho P \supset \mathcal{G}_m\big\} < 1. \tag{1}$$

By Caratheodory's theorem, for a given point $\omega \in \Omega$ of the probability space, any cross-polytope P inscribed into the realization $\mathcal{G}_m(\omega)$, can be represented as $P = \Gamma(\omega)A(B_1^n)$ for an $m \times n$ matrix A (determined by ω) such that the cardinality of support of every column of A is at most n and the ℓ_1^n–norm of each column at most one. Define the class $\mathcal{A}_{m,n}$ of matrices A in $\mathbb{R}^{m\times n}$ satisfying the conditions

$$|\operatorname{supp}\operatorname{col}_i(A)| \leq n; \qquad \|\operatorname{col}_i(A)\|_1 \leq 1; \qquad i \leq n.$$

With this definition, (1) would follow as long as we show that the event

$$\mathcal{E}_* := \big\{\text{There is } A \in \mathcal{A}_{m,n} \text{ such that } \rho\,\Gamma A(B_1^n) \supset \mathcal{G}_m\big\}$$

has probability strictly less than one.

Further, define a discretization of $\mathcal{A}_{m,n}$ as follows. Let $\mathcal{N} \subset \mathcal{A}_{m,n}$ be the set of all $m \times n$ matrices $A = (a_{ij})$ in $\mathcal{A}_{m,n}$ such that $\varepsilon^{-1}a_{ij} \in \mathbb{Z}$ for all indices i, j. It is easy to see that for any $A = (a_{ij}) \in \mathcal{A}_{m,n}$ there is $A' = (a'_{ij}) \in \mathcal{N}$ such that $\max_{i,j} |a_{ij} - a'_{ij}| < \varepsilon$.

We have the following elementary lemma:

Lemma 2.1. *Let parameters m, n, ρ, ε satisfy the additional assumptions $m \leq n^{10}$ and $\varepsilon\rho n^2 \leq 1$. Denote by $\mathcal{E}_{2.1}$ the event*

$$\mathcal{E}_{2.1} := \big\{\textit{There is } A \in \mathcal{N} \textit{ such that } 2\rho\,\Gamma A(B_1^n) \supset \mathcal{G}_m\big\}.$$

Then $\mathbb{P}(\mathcal{E}_*) \leq \mathbb{P}(\mathcal{E}_{2.1}) + 2^{-n}$.

Proof. Assume that the difference $\mathcal{E}_* \setminus \mathcal{E}_{2.1}$ is non-empty, and fix any $\omega \in \mathcal{E}_* \setminus \mathcal{E}_{2.1}$. Let $r \geq 0$ be the largest real number such that $rB_2^n \subset \mathcal{G}_m(\omega)$, and let $A = A(\omega) = (a_{ij})$ be a matrix in $\mathcal{A}_{m,n}$ such that $\rho\,\Gamma(\omega)A(\omega)(B_1^n) \supset \mathcal{G}_m(\omega)$. Take $A' = (a'_{ij}) \in \mathcal{N}$ with $\max\limits_{i,j}|a_{ij} - a'_{ij}| < \varepsilon$ and $\operatorname{supp}\operatorname{col}_i(A') \subset \operatorname{supp}\operatorname{col}_i(A)$, $i \leq n$. It is not difficult to see that $\Gamma(\omega)A'(B_1^n) + \varepsilon n \max\limits_{j\leq m}\|G_j(\omega)\|_2\, B_2^n \supset \Gamma(\omega)A(B_1^n)$. Indeed, take any $x \in \Gamma(\omega)A(B_1^n)$; then $x = \sum_{i=1}^n \alpha_i \Gamma(\omega)\operatorname{col}_i(A)$ for some $(\alpha_1,\dots,\alpha_n) \in B_1^n$. Set $y := \sum_{i=1}^n \alpha_i\Gamma(\omega)\operatorname{col}_i(A')$ and observe that $\|\Gamma(\omega)\operatorname{col}_i(A'-A)\|_2 \leq \varepsilon n \max\limits_{j\leq m}\|G_j(\omega)\|_2$, $i \leq n$. Hence, $\|x-y\|_2 \leq \varepsilon n \max\limits_{j\leq m}\|G_j(\omega)\|_2$, so $x \in \Gamma(\omega)A'(B_1^n) + \varepsilon n \max\limits_{j\leq m}\|G_j(\omega)\|_2\, B_2^n$.

The above inclusion, together with the definition of r, gives

$$\Gamma(\omega)A'(B_1^n) + \varepsilon n \max_{j\leq m}\|G_j(\omega)\|_2\,\frac{\rho}{r}\,\Gamma(\omega)A(B_1^n) \supset \Gamma(\omega)A(B_1^n).$$

This implies that $\varepsilon n \max\limits_{j\leq m}\|G_j(\omega)\|_2\,\frac{\rho}{r} > \frac{1}{2}$, since otherwise we would have $\Gamma(\omega)A'(B_1^n) \supset \frac{1}{2}\Gamma(\omega)A(B_1^n)$, which is impossible as $\omega \notin \mathcal{E}_{2.1}$. Next, observe that the definition of r implies that there is $u \in S^{n-1}$ such that $|\langle u, G_i(\omega)\rangle| \leq r$ for all $i \leq m$ whence $s_{\min}(\Gamma(\omega)^T) \leq r\sqrt{m}$.

Summarizing, we showed that the difference $\mathcal{E}_* \setminus \mathcal{E}_{2.1}$ is contained in the event

$$\Big\{\varepsilon n \max_{j\leq m}\|G_i\|_2\,\frac{\rho\sqrt{m}}{s_{\min}(\Gamma^T)} > \frac{1}{2}\Big\}.$$

Standard, by now, concentration properties for $\|G_i\|_2$ and $s_{\min}(\Gamma^T)$ (see, for example, [2, Theorem II.6 and Theorem II.13]), together with our assumptions on parameters, imply the result. □

2.2 Distances to random linear spans

The next lemma is an elementary application of standard concentration results for Gaussian variables. We provide the proof for Reader's convenience. Recall that Γ is the "global object" of the proof defined as the $n \times m$ standard Gaussian matrix with columns G_i, $i \leq m$.

Lemma 2.2 (Distances to linear spans)**.** *There are universal constants $C_{2.2}, c_{2.2} > 0$ with the following property. Assume that $n \geq C_{2.2}$, $n/2 \leq u \leq n$, $1 \leq k \leq u/2$, $\tau \geq C_{2.2}$, $\delta \in (1/k, 1]$; and fix any $m \times u$ non-random matrix B of full rank u such that each column has Euclidean norm at most one. Denote $H_i := \Gamma(\operatorname{col}_i(B))$, $i \leq u$. Further, for any permutation σ of $[u]$ let $\mathcal{E}_\sigma$ be the event that*

$$\Big|\Big\{i:\ u-k+1 \leq i \leq u,\ \operatorname{dist}(H_{\sigma(i)}, \operatorname{span}\{H_{\sigma(j)},\ j \leq i-1\}) \leq \tau\sqrt{n-u+k}\Big\}\Big| \geq (1-\delta)k.$$

Then $\mathbb{P}(\mathcal{E}_\sigma) \geq 1 - e^{-c_{2.2}\tau^2\delta(n-u+k)k}$ and, moreover, $\mathbb{P}(\bigcap\limits_{\sigma\in\Pi_u}\mathcal{E}_\sigma) \geq 1 - u^k\, e^{-c_{2.2}\tau^2\delta(n-u+k)k}$, where Π_u is the set of all permutations on $[u]$.

Proof. Since the linear span of $\{H_{\sigma(j)},\ j \leq u-k\}$ is completely determined by Γ and the set $\{\sigma(j),\ u \geq j \geq u-k+1\}$, it is enough to show the first assertion of the lemma: that for any fixed permutation σ of $[u]$ the probability of the event

$$\Big|\Big\{i:\ u-k+1 \leq i \leq u,\ \operatorname{dist}(H_{\sigma(i)}, \operatorname{span}\{H_{\sigma(j)},\ j \leq i-1\}) \leq \tau\sqrt{n-u+k}\Big\}\Big| \geq (1-\delta)k$$

is bounded from below by $1-e^{-c\tau^2\delta(n-u+k)k}$, for an appropriate universal constant $c>0$.

Fix a permutation σ and any subset $I\subset\{u-k+1,\dots,u\}$ of cardinality $\lceil\delta k\rceil$. Basic properties of the Gaussian distribution (specifically, rotational invariance), together with the definition of H_i's, immediately imply that there exists a decomposition

$$H_{\sigma(i)}=\widetilde{H}_{\sigma(i)}+\hat{H}_{\sigma(i)},\qquad i\le u,$$

such that for every $i>1$, $\widetilde{H}_{\sigma(i)}$ is a linear combination of $H_{\sigma(j)}$'s, $j<i$ (i.e. belongs to their linear span), while $\hat{H}_{\sigma(i)}$ is a non-zero multiple of the standard Gaussian vector in $\mathbb{R}^n$ independent from $\{H_{\sigma(j)},\ j<i\}$, with $\mathbb{E}\|\hat{H}_{\sigma(i)}\|_2^2\le n$. In particular, conditioned on any $(i-1)$–dimensional realization of $\mathrm{span}\{H_{\sigma(j)},\ j<i\}$, the distance

$$\mathrm{dist}(\hat{H}_{\sigma(i)},\mathrm{span}\{H_{\sigma(j)},\ j<i\})=\mathrm{dist}(H_{\sigma(i)},\mathrm{span}\{H_{\sigma(j)},\ j<i\})$$

is equidistributed with the Euclidean norm of a multiple of the standard Gaussian vector in $\mathbb{R}^{n-i+1}$ (with the multiplication coefficient not greater than one). Thus, applying a standard concentration inequality for Lipschitz functions in the Gauss space (see, for example, [2, Theorem II.6]), we get

$$\mathbb{P}\Big\{\mathrm{dist}(H_{\sigma(i)},\mathrm{span}\{H_{\sigma(j)},\ j<i\})>\tau\sqrt{n-u+k}\text{ for all } i\in I\Big\}\le e^{-c'\tau^2(n-u+k)|I|}$$

for a suitable universal constant $c'>0$. Taking the union bound over all possible choices of I, we get that the event

$$\Big|\Big\{i:\ u-k+1\le i\le u,\ \mathrm{dist}(H_{\sigma(i)},\mathrm{span}\{H_{\sigma(j)},\ j<i\})\le\tau\sqrt{n-u+k}\Big\}\Big|\ge(1-\delta)k$$

has probability at least $1-\delta^{-\delta k}e^{-c'\tau^2\delta(n-u+k)k}$. The result follows. □

3 Decomposition and structuring

In this section we develop a way to estimate from above probability of the event $\mathcal{E}_{2.1}$ from the previous section. The approach is significantly different from the one in paper [12]. One of main challenges of the net-based approach is to control probabilities in a way that admits some sort of a union bound. If the cardinality of the net $\mathcal{N}$ introduced in the previous section were very small, we would be able to bound the probability of $\mathcal{E}_{2.1}$ simply by summing up probability estimates of inclusion $2\rho\,\Gamma A(B_1^n)\supset\mathcal{G}_m$ for each $A\in\mathcal{N}$. However, the size of $\mathcal{N}$ is greater than 2^{n^2} and this approach would require to take rather small value for ρ to make sure the summation produces a number less than one.

To deal with this issue, we will partition every matrix A from $\mathcal{N}$ into a matrix with "very sparse" columns and a matrix with columns of small Euclidean norms. We introduce another global parameter $\alpha\in(0,1/2]$ whose value will be determined at the optimization stage. Let x be a vector in $\mathbb{R}^m$. We say that x is of *type* $(\alpha+)$ if $\|x\|_1\le 1$, and $|\mathrm{supp}\,x|\le 1/\alpha$. Further, x is of *type* $(\alpha-)$ if $|\mathrm{supp}\,x|\le n$, $\|x\|_1\le 1$ and $\|x\|_2<\sqrt{\alpha}$. Roughly speaking, type $(\alpha+)$ corresponds to very sparse vectors while $(\alpha-)$ consists of moderately sparse vectors of small Euclidean norm. It is easy to see that any vector $y\in\mathbb{R}^m$ with

$\|y\|_1 \leq 1$ and with cardinality of support at most n can be decomposed into the sum $y_1 + y_2$, where y_1 is of type $(\alpha+)$, y_2 is of type $(\alpha-)$, and y_1, y_2 have disjoint supports.

Define two mappings $\mathcal{F}_1, \mathcal{F}_2 : \mathcal{N} \to \mathcal{N}$ as follows: given $A = (a_{ij}) \in \mathcal{N}$, let $\mathcal{F}_1(A)$ be the $m \times n$ matrix with entries $a_{ij}\mathbf{1}_{|a_{ij}|\geq\alpha}$, where $\mathbf{1}_{|a_{ij}|\geq\alpha}$ is the indicator of the boolean expression "$|a_{ij}| \geq \alpha$". Further, we set $\mathcal{F}_2(A)$ to be the $m \times n$ matrix with entries $a_{ij}\mathbf{1}_{|a_{ij}|<\alpha}$. Finally, set $\mathcal{F}(A)$ to be the $m \times 2n$ matrix obtained by concatenating $\mathcal{F}_1(A)$ and $\mathcal{F}_2(A)$. Obviously, $A(B_1^n) \subset \mathcal{F}(A)(2B_1^{2n})$ for any $A \in \mathcal{N}$. This elementary relation turns out extremely useful in our context. It shows that every point of the random cross-polytope $\Gamma A(B_1^n)$ is a convex combination of random vectors of two types: vectors which are α^{-1}–sparse linear combinations of G_i's and vectors which have relatively small expected Euclidean norms. The number of vectors of the first type is relatively small (because of the α^{-1}–sparsity of corresponding linear combinations) allowing an efficient net-argument. At the same time, vectors of the second type are "short", which enables us to control their influence even though the number of corresponding linear combinations is relatiely large.

In the following lemma we formulate sufficient conditions which allow us to bound the Banach–Mazur distance between the Gluskin polytope $\mathcal{G}_m$ and an n-dimensional cross-polytope.

Lemma 3.1 (Decomposition)**.** *Let $\mathcal{E}$ be an event of non-zero probability having the following structure: $\mathcal{E} = \bigcap\limits_{A\in\mathcal{N}} \mathcal{E}_A$, where for every $A \in \mathcal{N}$ the event $\mathcal{E}_A$ is measurable with respect to σ-algebra generated by vectors G_j, with $j \in \bigcup\limits_{i\leq n} \operatorname{supp} \operatorname{col}_i(A)$. Let $\widetilde{G}$ be the standard Gaussian vector in $\mathbb{R}^n$ independent from Γ. Then*

$$\mathbb{P}(\mathcal{E}_{2.1} \cap \mathcal{E}) \leq |\mathcal{N}| \max_{A\in\mathcal{N}} \sup_{\omega\in\mathcal{E}_A} \mathbb{P}\Big(\big\{\omega' \in \Omega : \textit{ For some } I \subset [2n] \textit{ with } |I| = n \textit{ we have } \widetilde{G}(\omega') \in 4\rho \left(\operatorname{conv}\left\{ \pm \operatorname{col}_i(\Gamma(\omega)\mathcal{F}(A)),\ \ i \in I\right\}\right) \big\}\Big)^{m-n^2}.$$

Proof. Pick any point $\omega \in \mathcal{E}_{2.1} \cap \mathcal{E}$ (we assume that the intersection is non-empty as otherwise there is nothing to prove). In view of the definition of $\mathcal{E}_{2.1}$, there is $A = A(\omega) \in \mathcal{N}$ such that for any $j \leq m$ we have $G_j(\omega) \in 2\rho\,\Gamma(\omega)A(B_1^n) \subset 4\rho\,\Gamma(\omega)\mathcal{F}(A)(B_1^{2n})$. By Caratheodory's theorem, this implies that there is a subset $I = I(\omega, j) \subset [2n]$ of cardinality n such that

$$G_j(\omega) \in 4\rho \left(\operatorname{conv}\left\{ \pm \operatorname{col}_i(\Gamma(\omega)\mathcal{F}(A)),\ \ i \in I\right\}\right).$$

In what follows, we are only interested in indices $j \in [m] \setminus \bigcup\limits_{i\leq n} \operatorname{supp} \operatorname{col}_i(A)$, which will enable us to use independence. In view of the above, we get

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{2.1}\cap\mathcal{E}) \le \mathbb{P}\big(\big\{\exists A\in\mathcal{N} \quad &\text{such that for any } j\notin \bigcup_{i\le n}\text{supp col}_i(A)\\ &\text{and some } I=I(j)\subset[2n] \text{ with } |I|=n \text{ we have}\\ &G_j\in 4\rho\left(\text{conv}\left\{\pm\text{col}_i(\Gamma\mathcal{F}(A)),\ \ i\in I\right\}\right)\big\}\cap\mathcal{E}\big)\\ \le |\mathcal{N}|\max_{A\in\mathcal{N}}\mathbb{P}\big(\big\{&\text{For any } j\notin\bigcup_{i\le n}\text{supp col}_i(A) \text{ and some}\\ &I=I(j)\subset[2n] \text{ with } |I|=n \text{ we have}\\ &G_j\in 4\rho\left(\text{conv}\left\{\pm\text{col}_i(\Gamma\mathcal{F}(A)),\ \ i\in I\right\}\right)\big\}\cap\mathcal{E}_A\big).\end{aligned}$$

Observe that, by the conditions on $\mathcal{E}_A$, vectors G_j, $j\notin\bigcup_{i\le n}\text{supp col}_i(A)$, are independent from $\mathcal{E}_A$ and $\Gamma\mathcal{F}(A)$. Thus, the last expression can be estimated from above by

$$\begin{aligned}|\mathcal{N}|\max_{A\in\mathcal{N}}\sup_{\omega\in\mathcal{E}_A}\mathbb{P}\big(\big\{&\omega'\in\Omega: \text{For some } I\subset[2n] \text{ with } |I|=n \text{ we have}\\ &\widetilde{G}(\omega')\in 4\rho\left(\text{conv}\left\{\pm\text{col}_i(\Gamma(\omega)\mathcal{F}(A)),\ \ i\in I\right\}\right)\big\}\big)^{m-n^2}.\end{aligned}$$

□

The last lemma provides a very useful mechanism of bounding probability $\mathbb{P}(\mathcal{E}_{2.1})$ from above. Indeed, if the events $\mathcal{E}_A$ are such that for any $\omega\in\mathcal{E}_A$ we have

$$\begin{aligned}\mathbb{P}\big(\big\{\omega'\in\Omega: &\text{For some } I\subset[2n] \text{ with } |I|=n \text{ we have}\\ &\widetilde{G}(\omega')\in 4\rho\left(\text{conv}\left\{\pm\text{col}_i(\Gamma(\omega)\mathcal{F}(A)),\ \ i\in I\right\}\right)\big\}\big)\le 1/2,\end{aligned}\tag{2}$$

then, by choosing m sufficiently large (so that the power $m-n^2$ is big enough) we will beat the cardinality of $\mathcal{N}$ and obtain an upper estimate for $\mathbb{P}(\mathcal{E}_{2.1}\cap\mathcal{E})$. If, at the same time, the probability of $\mathcal{E}$ is close to one, this will imply upper bounds for $\mathbb{P}(\mathcal{E}_{2.1})$. Thus, the rest of the argument consists of two major steps:

- Find an appropriate event $\mathcal{E}_\cap=\bigcap_{A\in\mathcal{N}}\mathcal{E}_A$ satisfying conditions of Lemma 3.1;
- Show that for any $A\in\mathcal{N}$ and $\omega\in\mathcal{E}_A$ we have (2).

4 Constructing event $\mathcal{E}_\cap$

We define the following parametric family of cross-polytopes in $\mathbb{R}^n$. Given $1\le k\le n$ and $h>0$, define

$\mathcal{K}(k,h)$ — collection of all origin-symmetric cross-polytopes $T\subset\mathbb{R}^n$ having the following structure: $T=\text{conv}\{\pm x_1,\pm x_2,\dots,\pm x_n\}$, where for each permutation σ of $[n]$ we have:
$\big|\big\{i:\ n-k+1\le i\le n,\ \text{dist}(x_{\sigma(i)},\text{span}\{x_{\sigma(j)},\ j<i\})\le h\big\}\big|\ge k/4.$

Further, let $s, \widetilde{s}, \tau, \delta$ be parameters satisfying $n \geq s \geq \widetilde{s} \geq 1$ and $\tau \geq C_{2.2}$. For every matrix $A \in \mathcal{N}$ define two events $\mathcal{E}_A^1(s, \widetilde{s})$ and $\mathcal{E}_A^2(\widetilde{s}, \delta, \tau)$ as follows:

$$\begin{aligned}\mathcal{E}_A^1(s,\widetilde{s}) := \big\{&\text{For every } I_1 \subset [n],\ I_2 \subset \{n+1,\dots,2n\} \text{ with } |I_1 \cup I_2| = n,\ |I_1| \geq n-\widetilde{s},\\ &\text{such that vectors } \mathrm{col}_i(\Gamma\mathcal{F}(A)),\ i \in I_1\cup I_2, \text{ are linearly independent, we}\\ &\text{have conv}\left\{\pm\mathrm{col}_i(\Gamma\mathcal{F}(A)),\ i\in I_1\cup I_2\right\} \in \mathcal{K}(s, C_{2.2}\sqrt{2s})\big\}\end{aligned}$$

and

$$\begin{aligned}\mathcal{E}_A^2(\widetilde{s},\delta,\tau) := \big\{&\text{For every } I_1 \subset [n],\ I_2 \subset \{n+1,\dots,2n\} \text{ with } |I_1 \cup I_2| = n,\ |I_2| > \widetilde{s}, \text{ such}\\ &\text{that vectors } \mathrm{col}_i(\Gamma\mathcal{F}(A)),\ i \in I_1\cup I_2, \text{ are linearly independent, we have}\\ &\big|\big\{i \in I_2:\ \mathrm{dist}(\mathrm{col}_i(\Gamma\mathcal{F}(A)), \mathrm{span}\{\mathrm{col}_j(\Gamma\mathcal{F}(A)),\ j \in (I_1\cup I_2)\cap[i-1]\})\\ &\leq \tau\sqrt{\alpha|I_2|}\big\}\big| \geq (1-\delta)|I_2|\big\}.\end{aligned}$$

Set $\mathcal{E}_A := \mathcal{E}_A^1 \cap \mathcal{E}_A^2$ and $\mathcal{E}_\cap := \bigcap\limits_{A\in\mathcal{N}} \mathcal{E}_A$ (for brevity, we sometimes suppress the list of parameters). It is not difficult to see that by the definition of $\mathcal{E}_A$ (regardless of the values of the parameters), the event is measurable with respect to the σ-algebra generated by G_j, with $j \in \bigcup\limits_{i\leq n} \mathrm{supp}\ \mathrm{col}_i(A)$. Thus, the intersection $\mathcal{E}_\cap$ satisfies conditions of Lemma 3.1. In the next two lemmas, which conclude this section, we will show that, with an appropriate choice of parameters, the event $\mathcal{E}_\cap$ has probability close to one.

Lemma 4.1. *There is a universal constant $C_{4.1} > 0$ with the following property. Assume that $m/\varepsilon \leq n^{10}$, $1 > \delta \geq (\log n)^{-1}$, $n \geq \widetilde{s} \geq \log^2 n$ and, additionally, assume that $\tau \geq C_{2.2}$ is such that*

$$\min\left(\frac{\tau^2\delta\widetilde{s}^2\alpha}{n}, \frac{\tau^2\delta\widetilde{s}}{n}\right) \geq C_{4.1}\log n.$$

Then

$$\mathbb{P}\Big(\bigcap_{A\in\mathcal{N}} \mathcal{E}_A^2(\widetilde{s},\delta,\tau)\Big) \geq 1 - \frac{1}{n}.$$

Proof. Fix any $p \in \{0,1,\dots,n\}$ and consider the collection T_p of all $m\times n$ matrices B satisfying the following condition: there is $A \in \mathcal{N}$ such that B is a submatrix of $\mathcal{F}(A)$, where the first p columns of B are from $\mathcal{F}_1(A)$ and the last $n-p$ columns are from $\mathcal{F}_2(A)$. Obviously, the set T_p is contained within the collection T_p' of all $m\times n$ matrices where the first p columns are of type $(\alpha+)$, the last $n-p$ columns are of type $(\alpha-)$, and, additionally, all entries of the matrix are from $\varepsilon\mathbb{Z}$. It is not difficult to see that the cardinality of T_p' (hence, T_p) can be bounded from above as

$$|T_p'| \leq \left(\binom{m}{\lfloor\alpha^{-1}\rfloor}(\varepsilon/3)^{-1/\alpha}\right)^p \left(\binom{m}{n}(\varepsilon/3)^{-n}\right)^{n-p}.$$

We need to show that probability of the event $\bigcap\limits_{A\in\mathcal{N}} \mathcal{E}_A^2$, that is, the event

$$\begin{aligned}\big\{&\text{For any } p \in \{0,1,\dots,n-\widetilde{s}-1\} \text{ and any } B \in T_p \text{ of full rank, we have}\\ &\big|\big\{p+1 \leq i \leq n:\ \mathrm{dist}(\mathrm{col}_i(\Gamma B), \mathrm{span}\{\mathrm{col}_j(\Gamma B),\ j<i\}) \leq \tau\sqrt{\alpha(n-p)}\big\}\big|\\ &\geq (1-\delta)(n-p)\big\},\end{aligned}$$

is close to one. Take any $p \in \{0,1,\dots,n-\widetilde{s}-1\}$. By the definition of T_p, we have $\|\mathrm{col}_i(B)\|_2 \le \sqrt{\alpha}$ for all $B \in T_p$ and $i > p$. Hence, in view of Lemma 2.2, for any $B \in T_p$ of full rank we have

$$\mathbb{P}\Big\{\big|\{p+1 \le i \le n:\ \mathrm{dist}(\mathrm{col}_i(\Gamma B), \mathrm{span}\{\mathrm{col}_j(\Gamma B),\ j<i\}) \\ \le \tau\sqrt{\alpha(n-p)}\}\big| \ge (1-\delta)(n-p)\Big\} \ge 1 - e^{-c_{2.2}\tau^2\delta(n-p)^2}.$$

This immediately implies

$$\mathbb{P}\big(\bigcap_{A\in\mathcal{N}} \mathcal{E}_A^2\big) \ge 1 - \sum_{p=0}^{n-\widetilde{s}-1} \bigg(\binom{m}{\lfloor \alpha^{-1}\rfloor}(\varepsilon/3)^{-1/\alpha}\bigg)^p \bigg(\binom{m}{n}(\varepsilon/3)^{-n}\bigg)^{n-p} e^{-c_{2.2}\tau^2\delta(n-p)^2} \\ =: 1 - \sum_{p=0}^{n-\widetilde{s}-1} J_p.$$

A direct computation shows that for every $p \in \{0,1,\dots,n-\widetilde{s}-1\}$, we have

$$J_p \le \Big(\frac{3e\alpha m}{\varepsilon}\Big)^{\alpha^{-1}p}\Big(\frac{3em}{\varepsilon n}\Big)^{n(n-p)} e^{-c_{2.2}\tau^2\delta(n-p)^2}.$$

Note that, by the assumptions on the parameters assuming the constant $C_{4.1}$ is sufficiently large), we have

$$c_{2.2}\tau^2\delta(n-p)^2 \ge 2\Big(\alpha^{-1}p\log\frac{3e\alpha m}{\varepsilon} + n(n-p)\log\frac{3em}{\varepsilon n}\Big),$$

whence $J_p \le e^{-c_{2.2}\tau^2\delta(n-p)^2/2}$. Summing over all admissible p, we get the result. □

Lemma 4.2. *There is a universal constant $C_{4.2}$ with the following property. Assume that $m/\varepsilon \le n^{10}$, $n \ge s \ge 4\widetilde{s} \ge 4\log^2 n$, and that $\frac{\widetilde{s}^2\alpha}{n} \ge C_{4.2}\log n$. Then*

$$\mathbb{P}\big(\bigcap_{A\in\mathcal{N}} \mathcal{E}_A^1(s,\widetilde{s})\big) \ge 1 - \frac{1}{n}.$$

Proof. Let T_p be defined the same way as in the proof of Lemma 4.1. For any $p = n-\widetilde{s}, n-\widetilde{s}+1,\dots,n$, denote by $\mathrm{Proj}_p : \mathbb{R}^n \to \mathbb{R}^n$ the orthogonal projection onto first p standard coordinate vectors and observe that the set of matrices

$$Q_p := \big\{B\mathrm{Proj}_p :\ B \in T_p\big\}$$

has cardinality

$$|Q_p| \le \bigg(\binom{m}{\lfloor\alpha^{-1}\rfloor}(\varepsilon/3)^{-1/\alpha}\bigg)^p \le \Big(\frac{3e\alpha m}{\varepsilon}\Big)^{\alpha^{-1}p}.$$

Further, observe that the intersection $\bigcap_{A\in\mathcal{N}} \mathcal{E}_A^1(s,\widetilde{s})$ coincides with the event $\bigcap_{p=n-\widetilde{s}}^{n} \mathcal{E}_p$, where

$$\mathcal{E}_p := \big\{\text{For every } A \in \mathcal{N},\ I_1 \subset [n],\ I_2 \subset \{n+1,\dots,2n\} \text{ with } |I_1\cup I_2| = n,\ |I_1| = p, \\ \text{such that vectors } \mathrm{col}_i(\Gamma\mathcal{F}(A)),\ i \in I_1\cup I_2, \text{ are linearly independent, we} \\ \text{have } \mathrm{conv}\{\pm\mathrm{col}_i(\Gamma\mathcal{F}(A)),\ i\in I_1\cup I_2\} \in \mathcal{K}(s, C_{2.2}\sqrt{2s})\big\}.$$

In turn, probability of each event $\mathcal{E}_p$ can be bounded from below by probability of the event

$$\mathcal{E}'_p := \Big\{\text{For any } B \in Q_p \text{ of rank } p \text{ and any permutation } \sigma \text{ of } [p] \text{ we have} \\ \big|\{i:\ p-s+1 \le i \le p,\ \mathrm{dist}(\mathrm{col}_{\sigma(i)}(\Gamma B), \\ \mathrm{span}\{\mathrm{col}_{\sigma(j)}(\Gamma B),\ j \le i-1\}) \le C_{2.2}\sqrt{2s}\}\big| \ge s/2\Big\}.$$

Let us prove the last assertion. Take any $\omega \in \mathcal{E}'_p \setminus \Omega_0$, where Ω_0 is the event (of probability zero) that for some $A \in \mathcal{N}$, $\Gamma\mathcal{F}(A)$ contains an $n\times n$ submatrix of deficient rank. Further, let σ be any permutation of $[n]$; let $A \in \mathcal{N}$ and let $I_1 \subset [n]$, $I_2 \subset \{n+1,\dots,2n\}$ with $|I_1 \cup I_2| = n$, $|I_1| = p$. Denote by B the $m \times n$ matrix with first p columns coincident with $\mathrm{col}_i(\mathcal{F}(A))$, $i \in I_1$ (with ordering of columns preserved), and last $n-p$ zero columns. Clearly, $B \in Q_p$. Define permutation σ' of $[p]$ is such a way that $\sigma^{-1}(\sigma'(\ell))$ is increasing with ℓ on $[p]$. By the definition of event $\mathcal{E}'_p$, we have that there are at least $s/2$ indices $i \in \{p-s+1,\dots,p\}$ such that

$$\mathrm{dist}(\mathrm{col}_{\sigma'(i)}(\Gamma(\omega)B), \mathrm{span}\{\mathrm{col}_{\sigma'(j)}(\Gamma(\omega)B),\ j \le i-1\}) \le C_{2.2}\sqrt{2s}.$$

Let $R : I_1 \cup I_2 \to [n]$ be the order-preserving bijection and set $x_i := \mathrm{col}_{R^{-1}(i)}(\Gamma(\omega)\mathcal{F}(A))$, $i \le n$. Then the last condition can be rewritten as

$$\mathrm{dist}(x_{\sigma'(i')}, \mathrm{span}\{x_{\sigma'(j)},\ j \le i'-1\}) \le C_{2.2}\sqrt{2s} \text{ for at least } s/2 \text{ ind. } i' \ge p-s+1. \quad (3)$$

The conditions on p and $\widetilde{s}$ imply that the set $S := \{\sigma(i):\ n-s+1 \le i \le n\} \cap [p]$ has cardinality at least $s - \widetilde{s} \ge 3s/4$. Further, for any i with $\sigma(i) \in S$ we have $\sigma(i) = \sigma'(i')$, where, by the definition of σ', necessarily $i' \ge p-(n-i) \ge p-s+1$, and, moreover, $\mathrm{span}\{x_{\sigma'(j)},\ j \le i'-1\} \subset \mathrm{span}\{x_{\sigma(j)},\ j \le i-1\}$. In particular, we have

$$\mathrm{dist}(x_{\sigma(i)}, \mathrm{span}\{x_{\sigma(j)},\ j \le i-1\}) \le \mathrm{dist}(x_{\sigma'(i')}, \mathrm{span}\{x_{\sigma'(j)},\ j \le i'-1\}).$$

This, together with (3), implies that

$$\mathrm{dist}(x_{\sigma(i)}, \mathrm{span}\{x_{\sigma(j)},\ j \le i-1\}) \le C_{2.2}\sqrt{2s} \text{ for at least } s/4 \text{ indices } i \ge n-s+1,$$

whence $\omega \in \mathcal{E}_p$. Thus, indeed $\mathbb{P}(\mathcal{E}_p) \ge \mathbb{P}(\mathcal{E}'_p)$.

Applying Lemma 2.2 (with $\delta := 1/2$, $u := p$, $k := s$, $\tau := C_{2.2}$) and taking the union over all $B \in Q_p$, we get

$$\mathbb{P}(\mathcal{E}'_p) \ge 1 - n^s\, e^{-c_{2.2}C_{2.2}^2 s^2/2}\left(\frac{3e\alpha m}{\varepsilon}\right)^{\alpha^{-1}n} \ge 1 - e^{-c_{2.2}C_{2.2}^2 s^2/4},$$

where the last relation follows by the assumptions on parameters. Taking the union bound over admissible p, we get the result. □

5 Gaussian measure of tilted cross-polytopes

Let us recall our proof strategy as outlined in Section 3. We have constructed the event $\mathcal{E}_\cap = \bigcap_{A\in\mathcal{N}} \mathcal{E}_A$ and essentially showed in Lemmas 4.1 and 4.2 that under an appropriate choice of parameters $\mathcal{E}_\cap$ has probability close to one. As the second step of the proof, we will show that (again, with appropriately chosen parameters) each $\omega \in \mathcal{E}_A = \mathcal{E}_A^1 \cap \mathcal{E}_A^2$ satisfies (2). Clearly, (2) can be interpreted as a statement about the Gaussian measure of a union of cross-polytopes of the form $4\rho\,\mathrm{conv}\big\{\pm\,\mathrm{col}_i(\Gamma(\omega)\mathcal{F}(A)),\ i\in I\big\}$, where I is a subset of $[2n]$ of cardinality n. We will estimate the measure of each such cross-polytope in one of the two ways depending on the cardinality of the set $I\cap[n]$. When $I\cap[n]$ is "large" (that is, vast majority generating vectors of the cross-polytope are realized as α^{-1}–sparse combinations of columns of $\Gamma(\omega)$, assuming an appropriate rescaling), we will use the condition that $\omega\in\mathcal{E}_A^1$, whence the cross-polytope is from the class $\mathcal{K}$. The Gaussian measure of such polytopes is computed below in Lemma 5.4, the central statement of this section. In the other case, when $I\cap[n]$ is "not very large", we will use the assumption $\omega\in\mathcal{E}_A^2$, which allows a relatively simple upper bound for the measure (essentially owing to the fact that the expected norm of the columns of $\Gamma\mathcal{F}_2(A)$ is rather small). The above description does not include the process of optimizing all the involved parameters, which we leave for the last section.

The next lemma is elementary; its proof is given for Reader's convenience.

Lemma 5.1. *Let $P = \mathrm{conv}\{\pm x_1, \pm x_2, \dots, \pm x_n\}$ be a symmetric non-degenerate cross-polytope in $\mathbb{R}^n$, and let $d_i := \mathrm{dist}(x_i, \mathrm{span}\{x_j, j<i\})$, $i\le n$. Further, let $1\le r\le n$, and let $P' = \mathrm{conv}\{\pm y_1, \pm y_2, \dots, \pm y_n\}$ be a symmetric cross-polytope in $\mathbb{R}^n$ such that $y_i = x_i$ for $i\le r$; $\|y_i\|_2 = d_i$ for $i = r+1,\dots,n$, and vectors y_i, $i\ge r+1$, are mutually orthogonal and orthogonal to* $\mathrm{span}\{x_j, j\le r\}$. *Then $\gamma_n(P')\ge\gamma_n(P)$.*

Proof. The lemma can be proven inductively on r, starting with $r=n$. Namely, we will construct a sequence of cross-polytopes $P_n, P_{n-1}, \dots, P_r$, with P_r being an orthogonal transformation of P' and P_n an orthogonal transformation of P, such that $\gamma_n(P_n)\le\gamma_n(P_{n-1})\le\dots\le\gamma_n(P_r)$.

At w-th step ($r+1\le w\le n$, in this argument we step backwards), let P_w be a cross-polytope generated by $\{y_1^w, y_2^w, \dots, y_n^w\}$ where y_i^w, $i = w+1,\dots,n$ are mutually orthogonal and orthogonal to $\mathrm{span}\{y_j^w,\ j\le w\}$. Denote by z_w the orthogonal projection of y_w^w onto the linear span of $\mathrm{span}\{y_j^w,\ j<w\}$, and let P_w^s be the symmetric cross-polytope generated by vectors $\{y_j^w, j\ne w\}\cup\{z_w + (z_w - y_w^w)\}$. Observe that P_w^s is an orthogonal transformation of P_w (reflection with respect to hyperplane orthogonal to $z_w - y_w^w$). Now, let P_{w-1} be generated by vectors $\{y_j^w, j\ne w\}\cup\{z_w - y_w^w\}$. It is not difficult to see that P_{w-1} can be represented as

$$P_{w-1} := \Big\{\frac{1}{2}(v_1+v_2):\ v_1\in P_w,\ v_2\in P_w^s,\ \langle v_1, z_w - y_w^w\rangle = \langle v_2, z_w - y_w^w\rangle\Big\}.$$

This, together with log-concavity of the Gaussian distribution, implies that $\gamma_n(P_{w-1})\ge\gamma_n(P_w)$, and the result follows. □

As a corollary of the last lemma, we obtain

Lemma 5.2. *Let $P = \operatorname{conv}\{\pm x_1, \pm x_2, \dots, \pm x_n\}$ be a symmetric cross-polytope in $\mathbb{R}^n$ such that for some $1 \le r < n$ and some $0 < h$ we have*

$$\operatorname{dist}(x_i, \operatorname{span}\{x_j, j < i\}) \le h, \qquad i \ge r+1.$$

Then $\gamma_n(P) \le \left(\frac{eh}{n-r}\right)^{n-r}$.

Proof. In view of the previous lemma, we can assume without loss of generality that $x_i = he_i$, $i \ge r+1$, and that the linear span of $\{x_j,\ j \le r\}$ coincides with $\operatorname{span}\{e_1, \dots, e_r\}$. Let $\widetilde{G} = (\widetilde{g}_1, \dots, \widetilde{g}_n)$ be the standard Gaussian vector in $\mathbb{R}^n$. Clearly, the event $\widetilde{G} \in P$ implies that

$$\sum_{i=r+1}^{n} |\widetilde{g}_i| \le h.$$

Thus,

$$\mathbb{P}\{\widetilde{G} \in P\} \le (2\pi)^{-(n-r)/2} \frac{(2h)^{n-r}}{(n-r)!} \le \frac{(2\pi)^{-(n-r)/2-1/2}(2h)^{n-r}}{(n-r)^{n-r+1/2} e^{r-n}} \le \left(\frac{eh}{n-r}\right)^{n-r}.$$

The result follows. □

As an immediate consequence of the last lemma, we get

Lemma 5.3. *Let $P = \operatorname{conv}\{\pm x_1, \pm x_2, \dots, \pm x_n\}$ be a symmetric cross-polytope such that for some $h > 0$, $\delta \in (0, 1/2]$ and $1 \le k \le n$ we have*

$$\left|\left\{i :\ n-k+1 \le i \le n,\ \operatorname{dist}(x_i, \operatorname{span}\{x_j,\ j < i\}) \le h\right\}\right| \ge (1-\delta)k.$$

Then $\gamma_n(P) \le \left(\frac{2eh}{k}\right)^{(1-\delta)k}$.

Estimates analogous to Lemma 5.3 are easily available for cross-polytopes from $\mathcal{K}(k, h)$. Indeed, appropriately rearranging the generating vectors, we obtain that for any cross-polytope P from $\mathcal{K}(k, h)$ we have $\gamma_n(P) \le e^{-c'k}$ as long as $h \le c'k$, for some universal constant $c' > 0$. However, these estimates (with such a condition on h) are absolutely useless in our context: they are too weak because they do not take into consideration invariance of $\mathcal{K}(k, h)$ under permutations of the generating vectors. Applying a more elaborate argument, we can derive a much stronger statement, which is the core of our approach.

Lemma 5.4 (Gaussian measure of tilted cross-polytopes from $\mathcal{K}(k, h)$)**.** *Let*

$$P = \operatorname{conv}\{\pm x_1, \dots, \pm x_n\} \in \mathcal{K}(k, h),$$

with $1 \le k \le n$ and $h \le c_{5.4} n$. Then $\gamma_n(P) \le 2e^{-c_{5.4}k}$, for a universal constant $c_{5.4} > 0$.

Proof. Let $\widetilde{G}$ be the standard Gaussian vector in $\mathbb{R}^n$, and let $\widetilde{\mathcal{E}}$ be the event that $\widetilde{G} \in P$. Denote $q := \gamma_n(P) = \mathbb{P}(\widetilde{\mathcal{E}})$. Conditioned on $\widetilde{\mathcal{E}}$, there is a (random) vector $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ with $\|v\|_1 \le 1$ such that

$$\widetilde{G} = \sum_{i=1}^{n} v_i x_i.$$

As $\mathbb{E}\big(\sum_{i=1}^{n}|v_i| \mid \widetilde{\mathcal{E}}\big) \leq 1$, there is a non-random subset $I \subset [n]$ of cardinality k such that $\mathbb{E}\big(\sum_{i\in I}|v_i| \mid \widetilde{\mathcal{E}}\big) \leq k/n$, whence, by Markov's inequality, there is an event $\mathcal{E}' \subset \widetilde{\mathcal{E}}$ of probability at least $q/2$ such that

$$\sum_{i\in I}|v_i| \leq \frac{2k}{n}$$

everywhere on $\mathcal{E}'$. Without loss of generality (since the definition of the class $\mathcal{K}$ is invariant under permutations of generating vectors), we can assume that $I = \{n-k+1, \dots, n\}$. Now, for any $\omega \in \mathcal{E}'$ we have

$$\widetilde{G}(\omega) = \sum_{i=1}^{n} v_i(\omega)x_i = 4\sum_{i=1}^{n-k}\frac{v_i(\omega)}{4}x_i + 4\sum_{i=n-k+1}^{n}\frac{nv_i(\omega)}{4k}\Big(\frac{k}{n}x_i\Big),$$

where, setting $v_i' := \frac{v_i}{4}$ for $i \leq n-k$ and $v_i' := \frac{nv_i}{4k}$ for $i > n-k$, we get $\|(v_1', \dots, v_n')\|_1 \leq 1$. Thus, if we define non-random cross-polytope P' as

$$P' := 4\,\mathrm{conv}\,\big\{\pm x_1, \dots, \pm x_{n-k}, \pm kx_{n-k+1}/n, \dots, \pm kx_n/n\big\},$$

then

$$\mathbb{P}\{\widetilde{G} \in P'\} \geq q/2.$$

On the other hand, appropriately rearranging the generating vectors and using the definition of the class $\mathcal{K}(k,h)$, we get that there is a presentation

$$P' := \mathrm{conv}\,\{\pm y_1, \dots, \pm y_n\}$$

where for the last $\lceil k/4\rceil$ vectors y_i we have

$$\mathrm{dist}(y_i, \mathrm{span}\{y_j, j<i\}) \leq \frac{4kh}{n}, \qquad i = n-\lceil k/4\rceil+1, \dots, n.$$

Finally, applying Lemma 5.2 to P', we obtain

$$q/2 \leq e^{-c'k},$$

for a universal constant $c' > 0$. □

6 Completion of the proof

We assume that n is a large positive integer, and set $\varepsilon := n^{-3}$; $m := n^3$. Although parameter ρ will be determined at the very last stage, we assume that $\rho \leq n$, whence $\varepsilon\rho n^2 \leq 1$. Thus, applying Lemma 2.1, we get that, in order to prove (1), it is sufficient to show that the event $\mathcal{E}_{2.1}$ from Lemma 2.1 has probability strictly less than $1-2^{-n}$.

Let $\delta := \log^{-1} n$ and let parameters $s, \widetilde{s}, \tau, \alpha$ satisfy

$$n \geq s \geq 4\widetilde{s} \geq 4\log^2 n; \qquad \frac{\widetilde{s}^2\alpha}{n} \geq C_{4.2}\log n; \qquad \min\left(\frac{\tau^2\widetilde{s}^2\alpha}{n}, \frac{\tau^2\widetilde{s}}{n}\right) \geq C_{4.1}\log^2 n. \quad (4)$$

Let $\mathcal{E}_A = \mathcal{E}_A^1(s,\widetilde{s}) \cap \mathcal{E}_A^2(\widetilde{s},\delta,\tau)$ $(A \in \mathcal{N})$ and $\mathcal{E}_\cap := \bigcap\limits_{A\in\mathcal{N}} \mathcal{E}_A$ be defined as in Section 4.

Lemmas 4.1 and 4.2, together with assumptions (4), imply that $\mathbb{P}(\mathcal{E}_\cap) \geq 1 - \frac{2}{n}$. Hence, applying Lemma 3.1, we get

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{2.1}) \leq |\mathcal{N}| \max_{A\in\mathcal{N}} \sup_{\omega\in\mathcal{E}_A} \mathbb{P}\big(\big\{\omega' \in \Omega :\ &\text{For some } I \subset [2n] \text{ with } |I| = n \text{ we have} \\ &\widetilde{G}(\omega') \in 4\rho\left(\operatorname{conv}\left\{\pm \operatorname{col}_i(\Gamma(\omega)\mathcal{F}(A)),\ \ i \in I\right\}\right)\big\}\big)^{m-n^2} + \frac{2}{n}.\end{aligned}$$

It is not difficult to see that, with our choice of m, ε, the cardinality of the set $\mathcal{N}$ of matrices can be estimated as

$$|\mathcal{N}| \leq \left(\binom{m}{n}(\varepsilon/3)^{-n}\right)^n \leq e^{Cn^2\log n}$$

for an appropriate universal constant $C > 0$. Thus, if we show that (for a specific choice of parameters) for every matrix $A \in \mathcal{N}$ and $\omega \in \mathcal{E}_A$ we have

$$\begin{aligned}\mathbb{P}\big(\big\{\omega' \in \Omega :\ &\text{For some } I \subset [2n] \text{ with } |I| = n \text{ we have} \\ &\widetilde{G}(\omega') \in 4\rho\left(\operatorname{conv}\left\{\pm \operatorname{col}_i(\Gamma(\omega)\mathcal{F}(A)),\ \ i \in I\right\}\right)\big\}\big) \leq 1/2,\end{aligned} \tag{5}$$

this would imply that $\mathbb{P}(\mathcal{E}_{2.1})$ is close to zero, and we will obtain (1). For the rest of the proof, we are concerned with finding values for parameters so that condition (5) is satisfied.

The definition of the event $\mathcal{E}_A$ implies that for every $\omega \in \mathcal{E}_A$ and $I \subset [2n]$, the cross-polytope $T := \operatorname{conv}\left\{\pm \operatorname{col}_i(\Gamma(\omega)\mathcal{F}(A)),\ \ i \in I\right\}$ satisfies one of the three conditions:

- T is degenerate, whence $\mathbb{P}\{\omega' \in \Omega :\ \widetilde{G}(\omega') \in 4\rho T\} = 0$;
- $T \in \mathcal{K}(s, C_{2.2}\sqrt{2s})$ (if $|I \cap [n]| \geq n - \widetilde{s}$);
- $|I \cap [n]| < n - \widetilde{s}$ and

$$\begin{aligned}&\big|\big\{i \in I \setminus [n] :\ \operatorname{dist}(\operatorname{col}_i(\Gamma(\omega)\mathcal{F}(A)), \operatorname{span}\{\operatorname{col}_j(\Gamma(\omega)\mathcal{F}(A)),\ j \in I \cap [i-1]\}) \\ &\leq \tau\sqrt{\alpha|I \setminus [n]|}\big\}\big| \geq (1-\delta)|I \setminus [n]|.\end{aligned}$$

Observe that for each $p = 0, 1, \ldots, n$, there are $\binom{n}{p}^2$ ways to choose a subset $I \subset [2n]$ of cardinality n and with $|I \cap [n]| = p$. Thus, in order to satisfy (5), it is sufficient to have

$$\begin{aligned}&\sum_{p=0}^{n-\widetilde{s}-1} \binom{n}{p}^2 \sup\left\{\mathbb{P}\{\omega' :\ \widetilde{G}(\omega') \in 4\rho P\}\ ,\ P \in \mathcal{U}\right\} \\ +&\sum_{p=n-\widetilde{s}}^{n} \binom{n}{p}^2 \sup\left\{\mathbb{P}\{\omega' :\ \widetilde{G}(\omega') \in 4\rho P\}\ ,\ P \in \mathcal{K}(s, C_{2.2}\sqrt{2s})\right\} \leq \frac{1}{2}.\end{aligned}$$

Here, $\mathcal{U}$ denotes the collection of all cross-polytopes $P = \mathrm{conv}\{\pm x_1, \pm x_2, \dots, \pm x_n\}$ such that $\big|\{i:\ p+1 \le i \le n,\ \mathrm{dist}(x_i, \mathrm{span}\{x_j,\ j<i\}) \le \tau\sqrt{\alpha(n-p)}\}\big| \ge (1-\delta)(n-p)$. For any $p \le n - \widetilde{s} - 1$ and $P \in \mathcal{U}$, we get, by Lemma 5.3, that

$$\mathbb{P}\big\{\widetilde{G} \in 4\rho\, P\big\} \le \left(\frac{8e\rho\tau\sqrt{\alpha(n-p)}}{n-p}\right)^{(1-\delta)(n-p)}.$$

Further, for $p \ge n - \widetilde{s}$ and $P \in \mathcal{K}(s, C_{2.2}\sqrt{2s})$ we obtain, by Lemma 5.4,

$$\mathbb{P}\big\{\widetilde{G} \in 4\rho\, P\big\} \le 2e^{-c_{5.4}s},$$

provided that $4C_{2.2}\rho\sqrt{2s} \le c_{5.4}n$.

Summarizing, condition (5) is satisfied whenever we have a set of parameters $s, \widetilde{s}, \tau, \alpha, \rho$ satisfying conditons (4), together with condition $4C_{2.2}\rho\sqrt{2s} \le c_{5.4}n$ and condition

$$\sum_{p=0}^{n-\widetilde{s}-1} \binom{n}{p}^2 \left(\frac{8e\rho\tau\sqrt{\alpha(n-p)}}{n-p}\right)^{(1-\delta)(n-p)} + 2\sum_{p=n-\widetilde{s}}^{n} \binom{n}{p}^2 e^{-c_{5.4}s} \le \frac{1}{2}.$$

It is not difficult to see that

$$\sum_{p=n-\widetilde{s}}^{n} \binom{n}{p}^2 e^{-c_{5.4}s} \le (\widetilde{s}+1)\left(\frac{en}{\widetilde{s}}\right)^{2\widetilde{s}} e^{-c_{5.4}s} \le \frac{1}{n},$$

as long as $s \ge C\widetilde{s}\log n$ for a sufficiently large universal constant $C > 0$. Further,

$$\sum_{p=0}^{n-\widetilde{s}-1} \binom{n}{p}^2 \left(\frac{8e\rho\tau\sqrt{\alpha(n-p)}}{n-p}\right)^{(1-\delta)(n-p)} \le \frac{1}{n}$$

as long as $\frac{n^2\rho\tau\sqrt{\alpha}}{\widetilde{s}^{5/2}} \le c$ for a sufficiently small universal constant $c > 0$. Thus, condition (5) is satisfied whenever our set of parameters satisfies the relations

$$n \ge s \ge C\widetilde{s}\log n \ge 4C\log^3 n; \qquad 4C_{2.2}\rho\sqrt{2s} \le c_{5.4}n; \qquad \frac{\widetilde{s}^2\alpha}{n} \ge C_{4.2}\log n;$$
$$\min\left(\frac{\tau^2\widetilde{s}^2\alpha}{n}, \frac{\tau^2\widetilde{s}}{n}\right) \ge C_{4.1}\log^2 n; \qquad \frac{n^2\rho\tau\sqrt{\alpha}}{\widetilde{s}^{5/2}} \le c.$$

Our goal is to find the largest possible ρ so that the above conditions can be satisfied. Setting

$$\tau := \sqrt{C_{4.1}}\log n \max\big(\sqrt{n/\widetilde{s}}, \sqrt{n/(\widetilde{s}^2\,\alpha)}\big),$$

we get that we can take

$$\rho := c' \min\left(n/\sqrt{s}, \frac{1}{\log n}\frac{\widetilde{s}^3}{n^{5/2}\sqrt{\alpha}}, \frac{1}{\log n}\frac{\widetilde{s}^{7/2}}{n^{5/2}}\right)$$

for a small enough universal constant $c' > 0$. Since it is better for us to take α as small as possible, we can take $\alpha := \frac{C' n \log n}{\widetilde{s}^2}$ for a large enough universal constant $C' > 0$. Plugging in and optimizing over $\widetilde{s}$, s (up to logarithmic multiples, we should take s, $\widetilde{s}$ of order $n^{8/9}$), we get the result.

Acknowledgement. I would like to thank Assaf Naor and Pierre Youssef for introducing me to recent results on restricted invertibility of matrices, in connection with the Banach–Mazur distance to the cube, Assaf Naor for suggesting to work on the problem, Alexander Litvak and Nicole Tomczak-Jaegermann for valuable remarks.

References

[1] J. Bourgain and S. J. Szarek, The Banach-Mazur distance to the cube and the Dvoretzky-Rogers factorization, Israel J. Math. **62** (1988), no. 2, 169–180. MR0947820

[2] K. R. Davidson and S. J. Szarek, Local operator theory, random matrices and Banach spaces, in *Handbook of the geometry of Banach spaces, Vol. I*, 317–366, North-Holland, Amsterdam. MR1863696

[3] O. Friedland, P. Youssef, Approximating matrices and convex bodies, International Mathematics Research Notices, Vol. 2017, No. 00, pp. 1–19 doi: 10.1093/imrn/rnx206, arXiv:1605.03861

[4] A. A. Giannopoulos, A note on the Banach-Mazur distance to the cube, in *Geometric aspects of functional analysis (Israel, 1992–1994)*, 67–73, Oper. Theory Adv. Appl., 77, Birkhäuser, Basel. MR1353450

[5] E. D. Gluskin, The diameter of the Minkowski compactum is roughly equal to n, Funktsional. Anal. i Prilozhen. **15** (1981), no. 1, 72–73. MR0609798

[6] F. John, Extremum problems with inequalities as subsidiary conditions, in *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, 187–204, Interscience Publishers, Inc., New York, NY. MR0030135

[7] A. E. Litvak, M. Rudelson and N. Tomczak-Jaegermann, On approximation by projections of polytopes with few facets, Israel J. Math. **203** (2014), no. 1, 141–160. MR3273435

[8] P. Mankiewicz and N. Tomczak-Jaegermann, Quotients of finite-dimensional Banach spaces; random phenomena, in *Handbook of the geometry of Banach spaces, Vol. 2*, 1201–1246, North-Holland, Amsterdam. MR1999195

[9] V. A. Marchenko and L. A. Pastur, Distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (N.S.) **72 (114)** (1967), 507–536. MR0208649

[10] A. Naor and P. Youssef, Restricted invertibility revisited, in *A journey through discrete mathematics*, 657–691, Springer, Cham. MR3726618

[11] A. Naor and P. Youssef, Refined restricted invertibility, preprint

[12] S. J. Szarek, Spaces with large distance to ℓ_∞^n and random matrices, Amer. J. Math. **112** (1990), no. 6, 899–942. MR1081810

[13] S. J. Szarek and M. Talagrand, An "isomorphic" version of the Sauer-Shelah lemma and the Banach-Mazur distance to the cube, in *Geometric aspects of functional analysis (1987–88)*, 105–112, Lecture Notes in Math., 1376, Springer, Berlin. MR1008718

[14] P. Youssef, Restricted invertibility and the Banach-Mazur distance to the cube, Mathematika **60** (2014), no. 1, 201–218. MR3164527