LOW-DIMENSIONAL FACES OF RANDOM 0/1-POLYTOPES

VOLKER KAIBEL

Abstract. Let P be a random 0/1-polytope in $\mathbb{R}^d$ with $n(d)$ vertices, and denote by $\nu_r(P)$ the quotient of the number of faces of P with exactly r vertices and $\binom{n(d)}{r}$ (the *r-density* of P). For each $r \geq 3$, we establish the existence of a sharp threshold for the r-density and determine the values of the threshold numbers τ_r such that, for all $\varepsilon > 0$,

$$\mathbb{E}\left[\nu_r(P)\right] \;=\; \begin{cases} 1-\mathrm{o}(1) & \text{if } n(d) \leq 2^{(\tau_r-\varepsilon)d} \text{ for all } d \\ \mathrm{o}(1) & \text{if } n(d) \geq 2^{(\tau_r+\varepsilon)d} \text{ for all } d \end{cases}$$

holds for the expected value of $\nu_r(P)$. The threshold for $r = 2$ has already been determined in [8].

In particular, these results indicate that the high densities often encountered in polyhedral combinatorics (e.g., the cut-polytope has both 2- and 3-density equal to one) is due to the geometry of 0/1-polytopes rather than to the special combinatorics of the underlying problems.

1. Introduction and Results

Over the last decades, investigations of various special classes of 0/1-polytopes (convex hulls of sets of 0/1-points) have not only lead to beautiful structural results on combinatorial optimization problems, but also to powerful algorithms. Consequently, there has been some effort to learn more about the general class of 0/1-polytopes (see [11]).

In the 1980's, e.g., several results on the graphs of 0/1-polytopes have been obtained, most notably Naddef's proof [9] that they satisfy the Hirsch-conjecture. A quite spectacular achievement in 2000 was Bárány and Pór's theorem [2] stating that random 0/1-polytopes (within a certain range of vertex numbers) have super-exponentially (in the dimension) many facets. Their proof is based on the methods developed in the early 1990's by Dyer, Füredi, and McDiarmid [4], in order to show that the expected volume of a random d-dimensional 0/1-polytope with n vertices drops from (almost) zero to (almost) one very quickly with n passing the threshold $2^{(2/\sqrt{e})d}$.

While Bárány and Pór's result sheds some light on the highest-dimensional faces of random 0/1-polytopes, we investigate their lower dimensional faces in this paper. We define the *r-density* $\nu_r(P)$ of a polytope P with n vertices to be the number of faces of P with exactly r vertices devided by $\binom{n}{r}$. Thus, $\nu_r(P)$=1 if and only if P is r-neighborly in the usual sense (see, e.g., [10]).

For $r = 2$, $\tau_2(P)$ is the density of the graph of P. In this case, a threshold result for random 0/1-polytopes has recently been obtained in [8]. However, for specific classes of 0/1-polytopes, high r-density has been observed also for larger values of r: For example, cut-polytopes have both 2- and 3-density equal to one, i.e., every triple of vertices makes a triangle-face (see [1, 3]). Note that cut-polytopes (of complete graphs) have $2^{\Theta(\sqrt{d})}$ vertices.

Here, we obtain that there is a sharp threshold for the r-density of random 0/1-polytopes for all (fixed) r. The threshold values nicely extend the results for $r = 2$, while the proof becomes more involved and needs a heavier machinery (the one developed in the above mentioned paper by Dyer, Füredi, and McDiarmid). As a pay-back, the proof, however, reveals several interesting insights into the geometry of (random) 0/1-polytopes.

1.1. Results.

Throughout the paper, $\log(\cdot)$ and $\ln(\cdot)$ will denote the binary and the natural logarithm, respectively. For $0 < \xi < 1$, define

$$h(\xi) \;:=\; \xi \log \frac{1}{\xi} + (1-\xi) \log \frac{1}{1-\xi}$$

This work was done while the author was a member of the *Mathematical Sciences Research Institute* at Berkeley, CA, during Oct/Nov 2003. He furthermore was supported by the DFG Research Center *Mathematics for key*

(i.e., $h(\cdot)$ is the entropy function). For any $a \in \mathbb{R}$, let $[a] := \{1, 2, \ldots, \lfloor a \rfloor\}$ be the set of all positive integers not greater than a.

For every $r \in \{2, 3, \ldots\}$ and for each $i \in [\frac{r}{2}]$, we denote by $\mathrm{b}(r, i)$ the number of subsets of $[r]$ of size i or $r - i$, i.e.,

$$\mathrm{b}(r,i) \;=\; \begin{cases} 2\binom{r}{i} & \text{if } i < \frac{r}{2} \\ \binom{r}{\frac{r}{2}} & \text{if } i = \frac{r}{2} \end{cases} .$$

In particular, we have $\sum_{i \in [\frac{r}{2}]} \mathrm{b}(r,i) = 2^r - 2$. Let us define

$$H_r \;:=\; \frac{1}{2^r - 2} \sum_{i \in [\frac{r}{2}]} \mathrm{b}(r,i) h\big(\frac{i}{r}\big) .$$

Note that, for $r \geq 3$ we have $0 < H_r < 1$, and $H_2 = 1$.

We define

$$\mathrm{V}_d \;:=\; \{0,1\}^d \qquad \text{and} \qquad \mathrm{Q}_d \;:=\; [0,1]^d \;=\; \mathrm{conv}\ \mathrm{V}_d \ .$$

Let $r \in \{2, 3, \ldots\}$ be fixed. For every d and $n \in [2^d]$, choose the points $S_1, \ldots, S_r, X_1, \ldots, X_n \in \mathrm{V}_d$ independently uniformly at random. Let

$$S := \{S_1, \ldots, S_r\} , \quad X := \{X_1, \ldots, X_n\} , \quad P := \mathrm{conv}\,(X \cup S) ,$$

and denote by

$$f_r(d,n) \;:=\; \mathbb{P}\left[\mathrm{conv}\,(S) \text{ is a face of } P\right]$$

the probability that S is the vertex set of a face of P. Note that, since r is constant, for large d, S will be affinely independent with (very) high probability (see [7]). Therefore, if $\mathrm{conv}\,(S)$ is a face of P, then it will be an $(r-1)$-dimensional simplex-face with high probability.

Now we can formulate our main result.

Theorem 1. *For every $r \in \{3, 4, \ldots\}$ and for each $\varepsilon > 0$ the following holds: If $n : \mathbb{N} \to \mathbb{N}$ is any function, then we have*

$$f_r(d, n(d)) \;=\; \begin{cases} 1 - \mathrm{o}(1) & \text{if } n(d) \leq 2^{(\tau_r - \varepsilon)d} \text{ for all } d \\ \mathrm{o}(1) & \text{if } n(d) \geq 2^{(\tau_r + \varepsilon)d} \text{ for all } d \end{cases} ,$$

where

$$\tau_r \;=\; 1 - (1 - 2^{1-r}) H_r \ .$$

The evolution result on the density of the graphs of random 0/1-polytope obtained in [8] implies that the statement of Theorem 1 is also true for $r = 2$ (yielding $\tau_2 = \frac{1}{2}$).

Figure 1 illustrates the values τ_r for $r \in \{2, \ldots, 50\}$.

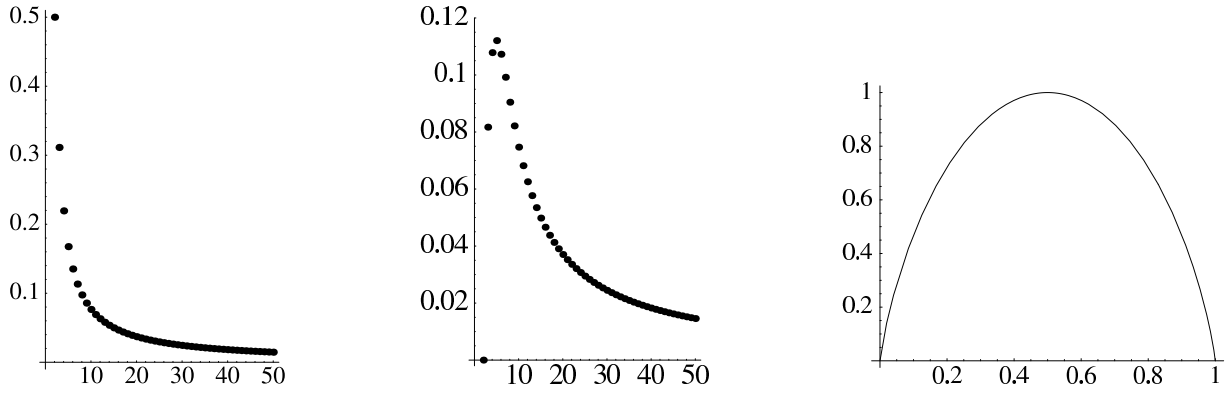

FIGURE 1. The values τ_r, $1 - H_r$ (see Proposition 1), as well as the function $h(\cdot)$.

A slightly different random model is the following: First, choose X' among the n-element subsets of V_d uniformly at random, and then choose S' among the r-element subsets of X' uniformly at random. Define $P' := \mathrm{conv}\ X'$ and

However, if for some $\varepsilon > 0$, $n(d) \leq 2^{(\frac{1}{2}-\varepsilon)d}$ holds for all n, then we have

$$\mathbb{P}\left[|\{X_1, \ldots, X_{n(d)}\}| = n(d)\right] \;=\; 1 - \mathrm{o}(1)\ .$$

Therefore, the statement of Theorem 1 is also true with $f_r(d, n(d))$ replaced by $g_r(d, n(d))$.

Since $\mathbb{P}[\text{conv } S' \text{ is a face of } P'] = \nu_r(P')$, this proves the result refered to in the abstract:

Theorem 2. *Let $r \in \{2, 3, \ldots\}$, $\varepsilon > 0$, and $n : \mathbb{N} \to \mathbb{N}$ be any function. For each $d \in \mathbb{N}$, choose an $n(d)$-element subset X of $\{0,1\}^d$ uniformly at random, and denote $P := \text{conv } X$. Then*

$$\mathbb{E}\left[\nu_r(P)\right] \;=\; \begin{cases} 1 - \mathrm{o}(1) & \text{if } n(d) \leq 2^{(\tau_r - \varepsilon)d} \text{ for all } d \\ \mathrm{o}(1) & \text{if } n(d) \geq 2^{(\tau_r + \varepsilon)d} \text{ for all } d \end{cases}$$

holds for the expected r-density of P.

1.2. **Overview of the proof.** The following basic polyhedral facts (consult [10] in case of doubts – or for background information) are important for our proof :

Remark 1. *Let P be a polytope with vertex set V, and $S \subset V$.*

1. $\text{conv}\,(S)$ *is a face of* P only if *the barycenter* $\frac{1}{|S|}\sum_{s\in S} s$ *of* S *is not contained in* $\text{conv}\,(V \setminus S)$.
2. $\text{conv}\,(S)$ *is a face of* P if *(and only if) there is an affine halfspace containing S in its boundary and $V \setminus S$ in its interior.*
3. *If $P \subset \mathbb{R}^d$ is a 0/1-polytope, and $F(S)$ is the smallest face of Q_d containing S, then* conv S *is a face of P* if and only if conv S *is a face of $P \cap F(S)$.*

The structure of the proof is as follows: First, we will (in Section 2) reduce the problem of determining a threshold for the probability that conv S is a face to the problem of determining a threshold for the corresponding probability conditioned on the event that S is not contained in a proper face of the cube (S is *spanning*). The latter problem finally is resolved in Section 5. There we need the results of Section 3 (for showing that *behind* the threshold conv S almost surely *is not* a face) and Section 4 (for showing that *below* the threshold conv S almost surely is a face).

Some of the calculations omitted in this extended abstract are given in the Appendix.

Acknowledgements. Im grateful to the Mathematical Sciences Research Institute at Berkeley for the generous financial support and the excellent working conditions I enjoyed during my visit in October/November 2003, when this work has been done. I thank Günter M. Ziegler for comments on an earlier version of the paper.

2. Reduction to the spanning case

For the rest of the paper, let $r \in \{3, 4, \ldots\}$ be fixed.

Again, choose the points $S_1, \ldots, S_r, X_1, \ldots, X_n \in \mathrm{V}_d$ independently uniformly at random, and let $S := \{S_1, \ldots, S_r\}$, $X := \{X_1, \ldots, X_n\}$, and $P := \text{conv}\,(X \cup S)$. Denote by $F(S)$ the smallest face of the cube Q_d that contains S. Let $d(S)$ be the dimension of $F(S)$ (i.e., $d(S)$ is the number of coordinates where not all elements of S agree). If $F(S) = \mathrm{Q}_d$ (i.e., $d(S) = d$), then we call S *spanning*. Let us define

$$\tilde{f}_r(d, n) \;:=\; \mathbb{P}\left[\text{conv}\,(S) \text{ is a face of } P \,|\, S \text{ is spanning}\right]\ .$$

In Section 5, we will prove the following result.

Proposition 1. *Let $r \in \{3, 4, \ldots, \}$. For each $\varepsilon > 0$, the following holds: If $n : \mathbb{N} \to \mathbb{N}$ is any function, then we have*

$$\tilde{f}_r(d, n(d)) \;=\; \begin{cases} 1 - \mathrm{o}(1) & \text{if } n(d) \leq 2^{(1 - H_r - \varepsilon)d} \text{ for all } d \\ \mathrm{o}(1) & \text{if } n(d) \geq 2^{(1 - H_r + \varepsilon)d} \text{ for all } d \end{cases}\ .$$

Figure 1 illustrates the threshold values $1 - H_r$. Note that they are not monotone in r. Actually, this is not surprising: In [8] it was shown that the threshold for $\tilde{f}_2(d, n(d))$ is at $(2 \pm \varepsilon)d$ (thus, Proposition 1 holds for $r = 2$ as well, due to $1 - H_2 = 0$). On the other hand, if r is large, than

will approximately be equal to that for the general case, for which it is quite plausible that it is monotonically decreasing with increasing r.

If for all $j \in [d]$ we have $\sum_{i\in[r]} S_{ij} \leq \frac{r}{2}$ (where S_{ij} is the j-th component of S_i), then we call S *reduced*. By appropriate "coordinate flips," we have

$$\begin{aligned}(1)\quad &\mathbb{P}\left[\text{conv}\,(S) \text{ is a face of } P \,|\, S \text{ is spanning}\right] \\ &\qquad = \mathbb{P}\left[\text{conv}\,(S) \text{ is a face of } P \,|\, S \text{ is spanning and reduced}\right] .\end{aligned}$$

This is not relevant in this section, but we will exploit it for the proof of Proposition 1 in Sections 3, 4, and 5.

The aim of the current section is to show that Proposition 1 implies Theorem 1.

2.1. **Preliminaries.** Let A be the $(r \times d)$-matrix whose rows are $S_1, \ldots, S_r$. Clearly, $d(S)$ equals the number of columns of A which are neither $\mathbf{0}$ (the all-zero vector) nor $\mathbf{1}$ (the all-one vector).

The random matrix A is distributed in the same way as an $r \times d$ matrix is distributed whose columns are chosen independently uniformly at random from $\{0,1\}^r$. For $t \in \{0,1\}^r$ chosen uniformly at random, we have $\mathbb{P}\left[t \notin \{\mathbf{0},\mathbf{1}\}\right] = 1 - 2^{1-r}$.

The de Moivre-Laplace Theorem (see, e.g., [5, Chap. 7]) yields that, for every $\delta > 0$, there is a $B_\delta > 0$ such that

$$\mathbb{P}\left[\left|d(S) - (1-2^{1-r})d\right| \leq B_\delta\sqrt{d}\right] \;\geq\; 1-\delta \tag{2}$$

holds for all large enough d.

For each $\delta > 0$, define

$$K_\delta(d) \;:=\; \left\{k \in [d] \,:\, |k - (1-2^{1-r})d| \leq B_\delta\sqrt{d}\right\} .$$

Thus, by (2) we have

$$\mathbb{P}\left[d(S) \in K_\delta\right] \;\geq\; 1-\delta \tag{3}$$

for all large enough d.

Throughout the section, we denote by FACE the event that conv S is a face of $P = \text{conv}\,(S \cup X)$. Furthermore, we denote

$$n(S) \;:=\; \left|\{i \in [n] \,:\, X_i \in F(S)\}\right| .$$

2.2. **The case $\mathbf{n(d) \leq 2^{(\tau_r-\varepsilon)d}}$.** Let $\delta > 0$ be fixed and let $k_{\min} \in K_\delta$ such that

$$\mathbb{P}\left[\text{FACE} \,|\, d(S) = k_{\min}\right] \;=\; \min\left\{\mathbb{P}\left[\text{FACE} \,|\, d(S) = k\right] \,:\, k \in K_\delta(d)\right\} .$$

Then we have

$$\left|d - \frac{k_{\min}}{1-2^{1-r}}\right| \;=\; o(k_{\min}) .$$

We therefore obtain

$$\begin{aligned}\mathbb{E}\left[n(S) \,\middle|\, d(S) = k_{\min}\right] &= 2^{k_{\min}-d} n(d) \\ &\leq 2^{k_{\min}-d+(\tau_r-\varepsilon)d} \\ &\leq 2^{\frac{1-2^{1-r}+\tau_r-1-\varepsilon}{1-2^{1-r}}k_{\min}+o(k_{\min})} . \qquad (4)\end{aligned}$$

The fraction in the exponent equals (see Appendix) $1-H_r-\varepsilon'$ where $\varepsilon' := \frac{\varepsilon}{1-2^{1-r}} > 0$. By Markov's inequality, we obtain

$$\mathbb{P}\left[n(S) \leq 2^{(1-H_r-\varepsilon'/2)k_{\min}} \,|\, d(S) = k_{\min}\right] \;=\; 1-o(1) . \tag{5}$$

Proposition 1 implies

$$\mathbb{P}\left[\text{FACE} \,|\, d(S) = k_{\min}, n(S) \leq 2^{(1-H_r-\varepsilon'/2)k_{\min}}\right] \;=\; 1-o(1) .$$

2.3. **The case $\mathbf{n(d) \geq 2^{(\tau_r+\varepsilon)d}}$.** The calculations that are necessary to prove the following lemma can be found in the Appendix.

Lemma 1. *Let $\alpha, \beta, \gamma > 0$ with $\alpha + \beta > 1 + \beta\gamma$, $n(d) := \lfloor 2^{\alpha d} \rfloor$, $k(d) = \beta d + o(d)$, and let F be any $k(d)$-face of Q_d. If $X_1, \ldots, X_{n(d)}$ are chosen independently uniformly at random from V_d, then we have*

$$\mathbb{P}\left[\left|\{i \in [n(d)] : X_i \in F\}\right| \geq 2^{\gamma k(d)}\right] = 1 - o(1) .$$

Now we can prove the second part of Theorem 1 (using Proposition 1). Let $\delta > 0$ be fixed and let $k_{\max} \in K_\delta$ such that

$$\mathbb{P}[\text{FACE} \mid d(S) = k_{\max}] = \max\left\{\mathbb{P}[\text{FACE} \mid d(S) = k] : k \in K_\delta(d)\right\} .$$

With $\alpha := \tau_r + \varepsilon$, $\beta := 1 - 2^{1-r}$, and $\gamma := 1 - H_r + \varepsilon$, one easily verifies (see Appendix) $\alpha + \beta > 1 + \beta\gamma$. Since $k_{\max} = (1 - 2^{1-r})d + o(d)$ we thus obtain from Lemma 1

$$\mathbb{P}[n(S) \geq 2^{(1-H_r+\varepsilon)k_{\max}} \mid d(S) = k_{\max}] = 1 - o(1) . \tag{6}$$

The second part of Proposition 1 implies

$$\mathbb{P}[\text{FACE} \mid d(S) = k_{\max}, n(S) \geq 2^{(1-H_r+\varepsilon)k_{\max}}] = o(1) .$$

Together with (6), the definition of $k_{\max}$, and (3), this proves the second part of Theorem 1.

3. Membership probabilities

In this section, we derive (from Dyer, Füredi, and McDiarmid's paper [4]) suitable lower bounds on $n(d)$ that, for specified points of Q_d, guarantee their membership in our random 0/1-polytopes with high probablity.

For any $z \in Q_d$, let us define

$$p(z) := \frac{1}{2^d} \min\left\{|T^{\leq} \cap V_d| : T^{\leq} \subset \mathbb{R}^d \text{ (closed affine) halfspace}, z \in T^{\leq}\right\} .$$

For each $\alpha > 0$, denote

$$Q_d^\alpha := \{z \in Q_d : p(z) \geq 2^{-\alpha d}\} .$$

For $z = (\zeta_1, \ldots, \zeta_d) \in \operatorname{int} Q_d$ (the interior of Q_d), define

$$H(z) := \frac{1}{d} \sum_{j \in [d]} h(\zeta_j) .$$

From Lemmas 2.1 and 4.1 of [4] one can deduce the following fact (for details see the Appendix). Let us mention that in particular the proof of Lemma 4.1 (needed for part (2) of Lemma 2) is quite hard. It is the core of Dyer, Füredi, and McDiarmid's beautiful paper.

Lemma 2. *For every $\alpha, \varepsilon > 0$ the following holds:*

(1) *If $n(d) \geq 2^{(\alpha+\varepsilon)d}$ holds for all d, and $X_1, \ldots, X_{n(d)} \in V_d$ are chosen independently uniformly at random, then we have*

$$\mathbb{P}[Q_d^\alpha \subseteq \operatorname{conv}\{X_1, \ldots, X_{n(d)}\}] = 1 - o(1) .$$

(2) *For large enough d,*

$$\{z \in \operatorname{int} Q_d : H(z) \geq 1 - \alpha + \varepsilon\} \subseteq Q_d^\alpha$$

holds.

The following straight consequence (choose $\alpha := 1 - \beta + \varepsilon/2$) of Lemma 2 is the key to the proof of the second part of Proposition 1.

Corollary 1. *Let $\beta > 0$. If $n(d) \geq 2^{(1-\beta+\varepsilon)d}$ holds for all d, and $X_1, \ldots, X_{n(d)} \in V_d$ are chosen independently uniformly at random, then we have*

4. Shallow cuts of the cube

This section is the heart of the proof of (the first part of) Proposition 1.

For $m \in \{1,2,\dots\}$, let $A(m)$ be an $r \times M$ matrix with $M := (2^r-2)m$ that has as its columns $2m$ copies of each vector $v \in \{0,1\}^r$ with $1 \le \mathbf{1}^T v < \frac{r}{2}$ and m copies of each $v \in \{0,1\}^r$ with $\mathbf{1}^T v = \frac{r}{2}$ (if r is even). This choice is motivated by the following fact (which is, however, irrelevant in this section): If $S_1,\dots,S_r$ are chosen independently uniformly at random from V_M, then the "multiplicity" of each vector $v \in \{0,1\}^r$ among the columns of $A(m)$ equals the expected number of appearances of v as a column of the matrix with rows $S_1,\dots,S_r$ — conditioned on the event that S is spanning and reduced.

Let $s_1,\dots,s_r \in \{0,1\}^M$ be the rows of $A(m)$, and let $L(i)$ be the set of indices of columns that have precisely i ones. We have $|L(i)| = \mathrm{b}(r,i)m$. Denote by $\sigma(i)$ the number of ones that any of the rows has in columns indexed by $L(i)$ (these numbers are equal for all rows). Obviously, we have $\sigma(i) = \frac{i}{r}\,\mathrm{b}(r,i)m$.

Let $b := (\beta_1,\dots,\beta_M)$ be the barycenter of the rows $s_1,\dots,s_r$. For each $l \in [M]$ we thus have $\beta_l = \frac{i(l)}{r}$, if $l \in L(i(l))$. Consequently (with the definition of $H(\cdot)$ from Section 3),

$$H(b) \;=\; \frac{1}{M}\sum_{i\in[\frac{r}{2}]} \mathrm{b}(r,i)\, m\, h\big(\frac{i}{r}\big) \;=\; H_r\ .$$

From Section 3 (see Lemma 2) we know that no hyperplane in $\mathbb{R}^M$ that contains b can therefore cut off significantly *less* than $2^{H_r M}$ points from V_M, and that there are indeed hyperplanes containing b that do also not cut off significantly *more* than $2^{H_r M}$ cube vertices. However, for our purposes, it will be necessary to know that there is a hyperplane containing not only b, but even the entire set $\{s_1,\dots,s_r\}$, and nevertheless cutting off not significantly more than $2^{H_r M}$ cube vertices.

The next result guarantees the existance of such a hyperplane, i.e., a certain shallow cut of the cube. Its proof will also reveal the basic reason for the appearance of the function $h(\cdot)$: It is due to the well-known fact that, for any constant $\alpha > 0$,

$$\sum_{p\in[\alpha q]} \binom{q}{p} \;=\; 2^{h(\alpha)q+\mathrm{o}(q)} \qquad (\text{for } q\to\infty) \tag{7}$$

(see, e.g., [6, Chap. 9,Ex. 42]).

Proposition 2. *There are coefficients $a_1,\dots,a_{\lfloor r/2\rfloor} \in \mathbb{R}$, such that the inequality*

$$\sum_{i\in[\frac{r}{2}]}\sum_{l\in L(i)} a_i\xi_l \;\le\; \sum_{i\in[\frac{r}{2}]} a_i\sigma(i) \tag{8}$$

has at most $2^{H_r M+\mathrm{o}(M)}$ 0/1-solutions. (By construction, the 0/1-points $s_1,\dots,s_r$ satisfy the inequality with equality.)

Proof. For all $a_1,\dots,a_{\lfloor r/2\rfloor} \in \mathbb{R}$ and for each $(k_1,\dots,k_{\lfloor r/2\rfloor}) \in \mathbb{N}^{\lfloor r/2\rfloor}$, let $\omega(k_1,\dots,k_{\lfloor r/2\rfloor})$ be the number of 0/1-solutions to (8) with precisely k_i ones in components indexed by $L(i)$ for each i. Let us further define

$$K(a_1,\dots,a_{\lfloor r/2\rfloor}) \;:=\; \Big\{(k_1,\dots,k_{\lfloor r/2\rfloor}) \in \mathbb{N}^{\lfloor r/2\rfloor} \;:\; \sum_{i\in[\frac{r}{2}]} a_i k_i \;\le\; \sum_{i\in[\frac{r}{2}]} a_i\sigma(i)\Big\}\ .$$

Then we have

$$\omega(k_1,\dots,k_{\lfloor r/2\rfloor}) \;=\; \begin{cases} \prod_{i\in[\frac{r}{2}]} \binom{\mathrm{b}(r,i)m}{k_i} & \text{if } (k_1,\dots,k_{\lfloor r/2\rfloor}) \in K(a_1,\dots,a_{\lfloor r/2\rfloor}) \\ 0 & \text{otherwise} \end{cases}\ .$$

Consequently, the number of 0/1-points satisfying (8) is precisely

$$\sum_{(k_1,\dots,k_{\lfloor r/2\rfloor})\in K(a_1,\dots,a_{\lfloor r/2\rfloor})} \omega(k_1,\dots,k_{\lfloor r/2\rfloor})\ . \tag{9}$$

If, for some i, we have $k_i > \mathrm{b}(r,i)m$, then clearly $\omega(k_1,\ldots,k_{\lfloor r/2\rfloor}) = 0$. Thus, the number of nonzero summands in (9) is $\mathrm{O}(m^r)$. Below, we will exhibit (constant) coefficients $a_1,\ldots,a_{\lfloor r/2\rfloor} \in \mathbb{R}$ with

$$\omega(k_1,\ldots,k_{\lfloor r/2\rfloor}) \;\le\; \omega(\sigma(1),\ldots,\sigma(\lfloor r/2\rfloor))\, 2^{\mathrm{o}(M)} \tag{10}$$

for all $(k_1,\ldots,k_{\lfloor r/2\rfloor}) \in K(a_1,\ldots,a_{\lfloor r/2\rfloor})$. This will eventually prove the proposition, since we have

$$\begin{aligned}
\omega(\sigma(1),\ldots,\sigma(\lfloor r/2\rfloor)) &= \textstyle\prod_{i\in[r/2]} \binom{\mathrm{b}(r,i)m}{\sigma(i)} \\
&= \textstyle\prod_{i\in[r/2]} \binom{\mathrm{b}(r,i)m}{(i/r)\,\mathrm{b}(r,i)m} \\
&= \textstyle\prod_{i\in[r/2]} 2^{h(i/r)\,\mathrm{b}(r,i)m+\mathrm{o}(m)} \\
&= 2^{\sum_{i\in[r/2]} h(i/r)\,\mathrm{b}(r,i)m+\mathrm{o}(m)} \\
&= 2^{H_r M+\mathrm{o}(m)}
\end{aligned}$$

(where the third equation is due to (7), and the last one comes from $M = (2^r-2)m$).

For simplicity, let us define $M_i := \mathrm{b}(r,i)m$. Furthermore, we denote by

$$B \;:=\; (0,M_1/2] \times (0,M_2/2] \times \cdots \times (0,M_{\lfloor r/2\rfloor}/2]$$

the box of all points in the open positive orthant of $\mathbb{R}^{\lfloor r/2\rfloor}$ for which no i-th coordinate exceeds $M_i/2$. (Note that $(\sigma(1),\ldots,\sigma(\lfloor r/2\rfloor) \in B$.)

For $a_1,\ldots,a_{\lfloor r/2\rfloor} \in \mathbb{R}$ define the halfspace

$$U(a_1,\ldots,a_{\lfloor r/2\rfloor}) \;:=\; \Big\{(\zeta_1,\ldots,\zeta_{\lfloor r/2\rfloor}) \in \mathbb{R}^{\lfloor r/2\rfloor} \;:\; \sum_{i\in[\frac{r}{2}]} a_i\zeta_i \;\le\; \sum_{i\in[\frac{r}{2}]} a_i\sigma(i)\Big\}\;.$$

Thus,

$$K(a_1,\ldots,a_{\lfloor r/2\rfloor}) = U(a_1,\ldots,a_{\lfloor r/2\rfloor}) \cap \mathbb{N}^{\lfloor r/2\rfloor}$$

holds. If all a_i are *nonnegative*, then we have

$$\begin{aligned}
&\max\{\omega(k_1,\ldots,k_{\lfloor r/2\rfloor}) \;:\; (k_1,\ldots,k_{\lfloor r/2\rfloor}) \in K(a_1,\ldots,a_{\lfloor r/2\rfloor})\} \\
&\qquad = \;\max\{\omega(k_1,\ldots,k_{\lfloor r/2\rfloor}) \;:\; (k_1,\ldots,k_{\lfloor r/2\rfloor}) \in B \cap U(a_1,\ldots,a_{\lfloor r/2\rfloor}) \cap \mathbb{N}^{\lfloor r/2\rfloor}\}\;.
\end{aligned}$$

Let us now approximate the function ω by Sterlings formula (see, e.g., [6, Eq. (9.40)])

$$N! \;=\; \Theta\Big(\sqrt{N}\frac{N^N}{e^N}\Big)\;.$$

We obtain (for $(k_1,\ldots,k_{\lfloor r/2\rfloor}) \in K(a_1,\ldots,a_{\lfloor r/2\rfloor})$)

$$\Omega(M^{-r}) \prod_{i\in[r/2]} \frac{M_i^{M_i}}{k_i^{k_i}(M_i-k_i)^{M_i-k_i}} \;\le\; \omega(k_1,\ldots,k_{\lfloor r/2\rfloor}) \;\le\; \mathrm{O}(M^r) \prod_{i\in[r/2]} \frac{M_i^{M_i}}{k_i^{k_i}(M_i-k_i)^{M_i-k_i}}\;.$$

Therefore, if we define the function $\eta : B \to \mathbb{R}$ via

$$\eta(\zeta_1,\ldots,\zeta_{\lfloor r/2\rfloor}) \;:=\; \prod_{i\in[r/2]} \frac{M_i^{M_i}}{\zeta_i^{\zeta_i}(M_i-\zeta_i)^{M_i-\zeta_i}}\;,$$

then it suffices to find nonnegative coefficients $a_1,\ldots,a_{\lfloor r/2\rfloor} \in \mathbb{R}$ such that $z^\star := (\sigma(1),\ldots,\sigma(\lfloor r/2\rfloor)$ maximizes η over $B \cap U(a_1,\ldots,a_{\lfloor r/2\rfloor})$ (since then (10) holds for $(k_1,\ldots,k_{\lfloor r/2\rfloor}) \in K(a_1,\ldots,a_{\lfloor r/2\rfloor})$).

In fact, since $\ln(\cdot)$ is monotonically increasing, we may equivalently investigate the function

$$\ln\big(\eta(\zeta_1,\ldots,\zeta_{\lfloor r/2\rfloor})\big) = \sum_{i\in[r/2]} M_i \ln M_i - \sum_{i\in[r/2]} \big(\zeta_i \ln\zeta_i + (M_i-\zeta_i)\ln(M_i-\zeta_i)\big)\;,$$

and thus find nonnegative coefficients $a_1,\ldots,a_{\lfloor r/2\rfloor} \in \mathbb{R}$ such that $z^\star$ *minimizes*

$$\tilde\eta(\zeta_1,\ldots,\zeta_{\lfloor r/2\rfloor}) \;:=\; \sum_{i\in[r/2]} \big(\zeta_i \ln\zeta_i + (M_i-\zeta_i)\ln(M_i-\zeta_i)\big)$$

Now we choose a_i as (-1) times the i-th partial derivative of $\tilde{\eta}$ at $z^\star$, i.e., $-(a_1,\ldots,a_{\lfloor r/2\rfloor})$ is the gradient of $\tilde{\eta}$ at $z^\star$. One easily calculates (see Appendix)

$$a_i = -\ln\frac{\sigma(i)}{M_i-\sigma(i)}$$

(which is nonnegative due to $\sigma(i)\leq\frac{M_i}{2}$).

In order to prove that this choice indeed makes $z^\star$ a minimizer of $\tilde{\eta}$ over $B\cap U(a_1,\ldots,a_{\lfloor r/2\rfloor})$, let $z\in B\cap U(a_1,\ldots,a_{\lfloor r/2\rfloor})$ be arbitrary ($z\neq z^\star$). Define $v:=z-z^\star$, and consider the function $\tilde{\eta}_{z^\star,z}:[0,1]\to\mathbb{R}$ defined via $\tilde{\eta}_{z^\star,z}(t):=\tilde{\eta}(z^\star+tv)$. The derivative of this function on $(0,1)$ is (see Appendix)

$$\tilde{\eta}'_{z^\star,z}(t) \;=\; \sum_{i\in[r/2]} v_i\ln\frac{\sigma(i)+tv_i}{M_i-\sigma(i)-tv_i}\;. \tag{11}$$

Consider any $i\in[r/2]$, and define $\varrho(t):=\frac{\sigma(i)+tv_i}{M_i-\sigma(i)-tv_i}$ (for all $t\in(0,1)$). If $v_i\geq 0$, then $\varrho(t)\geq\varrho(0)$, therefore, $v_i\ln\varrho(t)\geq v_i\ln\varrho(0)=-a_iv_i$. If $v_i<0$, then $\varrho(t)<\varrho(0)$, and thus, $v_i\ln\varrho(t)>v_i\ln\varrho(0)=-a_iv_i$. Hence, in any case the i-th summand in (11) is at least as large as $-a_iv_i$. Therefore, we obtain

$$\tilde{\eta}'_{z^\star,z}(t) \;\geq\; -\sum_{i\in[r/2]} a_iv_i\;.$$

Since $z\in U(a_1,\ldots,a_{\lfloor r/2\rfloor})$, we have $\sum_{i\in[r/2]}a_iv_i\leq 0$. Thus, $\tilde{\eta}'_{z^\star,z}(t)\geq 0$ for all $t\in(0,1)$. Since $\tilde{\eta}_{z^\star,z}$ is continuous on $[0,1]$, we hence conclude $\tilde{\eta}(z^\star)\leq\tilde{\eta}(z)$. □

5. The spanning case

Using the material collected in Sections 3 and 4, we will now prove Proposition 1 (and thus, as shown in Section 2) Theorem 1.

Towards this end, let $S_1,\ldots,S_r,X_1,\ldots,X_n\in\mathrm{V}_d$ be chosen according to the probability distribution induced by our usual distribution (choosing all points independently uniformly at random) on the event that $S:=\{S_1,\ldots,S_r\}$ is spanning and reduced (see (1) in Section 2). Let A be the $(r\times d)$-matrix with rows $S_1,\ldots,S_r$. Then A is a random matrix that has the same distribution as the $(r\times d)$-random matrix A' which arises from choosing each column independently uniformly at random from $\{0,1\}^r\setminus\{\mathbf{0},\mathbf{1}\}$, and then "flipping" all columns with more than $r/2$ ones. Therefore, if we denote the columns of A by $t_1,\ldots,t_d\in\{0,1\}^r$, then the t_j are (independently) distributed according to the distribution

$$\mathbb{P}\left[t_j=t\right] \;=\; \frac{\kappa(t)}{2^r-2} \;=:\; \pi(t)$$

whith

$$\kappa(t) \;=\; \begin{cases} 2 & \text{if } \mathbf{1}^Tt<r/2\\ 1 & \text{if } \mathbf{1}^Tt=r/2\\ 0 & \text{if } \mathbf{1}^Tt>r/2\end{cases}$$

for each $t\in\{0,1\}^r\setminus\{\mathbf{0},\mathbf{1}\}$.

Define

$$T_r \;:=\; \left\{t\in\{0,1\}^r\,:\,\mathbf{1}^Tt\leq\frac{r}{2}\right\}\;,$$

and denote, for every $t\in T_r$,

$$J(t) \;:=\; \{j\in[d]\,:\,t_j=t\}\;.$$

Let $m\in\mathbb{N}$ be the largest number such that $\kappa(t)m\leq|J(t)|$ holds for all $t\in T_r$. For each t, choose an arbitrary subset $\tilde{J}(t)\subseteq J(t)$ with $|\tilde{J}(t)|=\kappa(t)m$.

Denote by

$$\Delta_{\max} \;:=\; \max\left\{\left||J(t)|-\pi(t)d\right|\,:\,t\in T_r\right\}$$

From the de Moivre-Laplace Theorem (see, e.g., [5, Chap. 7]) one deduces the following for each $t \in T_r$: For every $\gamma' > 0$ there is a $C'_{\gamma'} > 0$ such that

$$\mathbb{P}\left[\big||J(t)| - \pi(t)d\big| \le C'_{\gamma'}\sqrt{d}\right] \;\ge\; 1-\gamma'$$

holds for all large enough d. Since $|T_r|$ is a constant, one can even derive the following stronger result from this: For every $\gamma > 0$ there is a constant $C_\gamma > 0$ such that

$$\mathbb{P}\left[\Delta_{\max} \le C_\gamma\sqrt{d}\right] \;\ge\; 1-\gamma \tag{12}$$

holds for all large enough d.

Let us define

$$\tilde{D} \;:=\; \bigcup_{t\in T_r} \tilde{J}(t)$$

and $\tilde{d} := |\tilde{D}| = m(2^r-2)$. In case of $\Delta_{\max} \le C_\gamma\sqrt{d}$, we can deduce

$$\tilde{d} \;\ge\; d - \mathrm{o}(d)\;. \tag{13}$$

5.1. **The case $\mathbf{n(d) \le 2^{(1-H_r-\varepsilon)d}}$.** Let $\tilde{S}_1,\dots,\tilde{S}_r$ be the canonical projections of $S_1,\dots,S_r$, respectively, to the coordinates in $\tilde{D}$. Then $\tilde{S}_1,\dots,\tilde{S}_r$ form a matrix $A(m)$ as defined in Section 4. Denote, for each $i \in [r/2]$,

$$\tilde{L}(i) \;:=\; \bigcup_{t\in T_r\,:\,\mathbf{1}^T t = i} \tilde{J}(t)\;.$$

Due to Proposition 2, there are coefficients $\tilde{a}_1,\dots,\tilde{a}_{\lfloor r/2\rfloor} \in \mathbb{R}$ such that the inequality

$$\sum_{i\in[r/2]} \tilde{a}_i \sum_{j\in\tilde{L}(i)} \tilde{a}_i\xi_j \;\le\; \sum_{i\in[r/2]} \tilde{a}_i \frac{i}{r}\,\mathrm{b}(r,i)m \;=:\; a_0 \tag{14}$$

has at most $2^{H_r\tilde{d}+\mathrm{o}(\tilde{d})}$ many 0/1-solutions (and $\tilde{S}_1,\dots,\tilde{S}_r$ satisfy the inequality with equality).

For each $j \in [d]$ let

$$a_j \;:=\; \begin{cases} \tilde{a}_i & \text{if } j \in \tilde{L}(i) \\ 0 & \text{if } j \in [d]\setminus\tilde{D} \end{cases}\;,$$

i.e., $a_1,\dots,a_d$ are the coefficients of (14) considered as an inequality for $\mathbb{R}^d$.

The inequality

$$\sum_{j\in[d]} a_j\xi_j \;\le\; a_0 \tag{15}$$

is satisfied with equality by $S_1,\dots,S_r$.

Let us, for the moment, restrict our attention to the event $\Delta_{\max} \le C_\gamma\sqrt{d}$. Then (15) has at most

$$2^{H_r\tilde{d}+\mathrm{o}(\tilde{d})}2^{d-\tilde{d}} \;=\; 2^{H_r d+\mathrm{o}(d)}$$

solutions (due to (13).

Define the hyperplane

$$T^{\le} \;:=\; \big\{(\xi_1,\dots,\xi_d)\in\mathbb{R}^d \,:\, \sum_{j\in[d]} a_j\xi_j \le a_0\big\}\;,$$

and let $T^{=}$ be its bounding hyperplane. Thus, we have

$$S_1,\dots,S_r \;\in\; T^{=} \qquad\text{and}\qquad \big|T^{\le}\cap \mathrm{V}_d\big| \;\le\; 2^{H_r d+\mathrm{o}(d)}\;. \tag{16}$$

Since $n(d) \le 2^{(1-H_r-\varepsilon)d}$, the expected number of points from X lying in $T^{\le}$ is at most

$$\frac{2^{H_r d+\mathrm{o}(d)}}{2^d} n(d) \;\le\; 2^{-\varepsilon d+\mathrm{o}(d)}\;.$$

Therefore, by Markov's inequality,

$$\mathbb{P}\left[X\cap T^{\le} = \varnothing \,\middle|\, \Delta_{\max} \le C_\gamma\sqrt{d}\right] \;= \mathrm{o}(1) \tag{17}$$

5.2. **The case $\mathbf{n(d) \geq 2^{(1-H_r+\varepsilon)d}}$.** From the remarks in the Introduction, we know

$$\mathbb{P}\left[|S| = r\right] \;=\; 1 - \mathrm{o}(1)\;. \tag{18}$$

Let $\gamma > 0$ be fixed, and assume $|S| = r$, i.e., the points $S_1, \ldots, S_r$ are pairwise disjoint. Denote by $b(S) = (\beta_1, \ldots, \beta_d)$ the barycenter of S. For each $t \in T_r$ and $j \in \tilde{J}(t)$, we have

$$\beta_j \;=\; \frac{\mathbf{1}^T t}{r}\;.$$

If $\Delta_{\max} \leq C_\gamma \sqrt{d}$ holds, we thus have (due to (13))

$$\begin{aligned} H(b(S)) = &\quad \tfrac{1}{d}\big(\textstyle\sum_{t\in T_r} m\kappa(t) h\big(\frac{\mathbf{1}^T t}{r}\big) + \mathrm{o}(d)\big) \\ = &\quad \tfrac{1}{d}\big(\textstyle\sum_{i\in[r/2]} m\,\mathrm{b}(r,i) h(i/r) + \mathrm{o}(d)\big) \\ = &\quad \tfrac{m(2^r-2)}{d} H_r + \mathrm{o}(1) \\ = &\quad (1-\mathrm{o}(1)) H_r + \mathrm{o}(1) \end{aligned}$$

Hence, in this case

$$H(b(S)) \;\geq\; H_r - \varepsilon$$

holds for large enough d, and, due to $n(d) \geq 2^{(1-H_r+\varepsilon)d}$, Corollary 1 implies

$$\mathbb{P}\,\left[b(S) \in \mathrm{conv}\,(X \setminus S)\,\middle|\,|S| = r, \Delta_{\max} \leq C_\gamma\sqrt{d}\right] \;\geq\; 1 - \mathrm{o}(1)\;.$$

Together with (18) and (12), this proves the second part of Proposition 1 (recall Remark 1).

References

[1] Francisco Barahona and Ali Ridha Mahjoub, *On the cut polytope*, Math. Programming **36** (1986), no. 2, 157–173.

[2] Imre Bárány and Attila Pór, *On* 0-1 *polytopes with many facets*, Adv. Math. **161** (2001), no. 2, 209–228.

[3] Michel Marie Déza and Monique Laurent, *Geometry of cuts and metrics*, Algorithms and Combinatorics, vol. 15, Springer-Verlag, Berlin, 1997.

[4] Martin E. Dyer, Zoltan Füredi, and Colin McDiarmid, *Volumes spanned by random points in the hypercube*, Random Structures Algorithms **3** (1992), no. 1, 91–106.

[5] William Feller, *An introduction to probability theory and its applications. Vol. I*, Third edition, John Wiley & Sons Inc., New York, 1968.

[6] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics*, second ed., Addison-Wesley Publishing Company, Reading, MA, 1994.

[7] Jeff Kahn, János Komlós, and Endre Szemerédi, *On the probability that a random* ± 1*-matrix is singular*, J. Amer. Math. Soc. **8** (1995), no. 1, 223–240.

[8] Volker Kaibel and Anja Remshagen, *On the graph-density of random 0/1-polytopes*, Approximation, Randomization, and Combinatorial Optimization (Proc. RANDOM03) (S. Arora, K. Jansen, J.D.P. Rolim, and A. Sahai, eds.), Lecture Notes in Computer Science, vol. 2764, Springer, 2003, pp. 318–328.

[9] Denis Naddef, *The Hirsch conjecture is true for* $(0,1)$*-polytopes*, Math. Programming **45** (1989), no. 1, (Ser. B), 109–110.

[10] Günter M. Ziegler, *Lectures on polytopes*, Graduate Texts in Mathematics, vol. 152, Springer-Verlag, New York, 1995, 2nd edition: 1998.

[11] ———, *Lectures on* 0/1*-polytopes*, Polytopes—Combinatorics and Computation (Oberwolfach, 1997), DMV Sem., vol. 29, Birkhäuser, Basel, 2000, pp. 1–41.

Appendix

Evaluation of the fraction in (4).

$$\begin{aligned}\frac{1-2^{1-r}+\tau_r-1-\varepsilon}{1-2^{1-r}} &= 1+\tfrac{1-(1-2^{1-r})H_r-1-\varepsilon}{1-2^{1-r}}\\ &= 1+\tfrac{-(1-2^{1-r})H_r-\varepsilon}{1-2^{1-r}}\\ &= 1-H_r-\tfrac{\varepsilon}{1-2^{1-r}}\end{aligned}$$

Proof of Lemma 1. For the sake of readibility, let $n := n(d)$ and $k := k(d)$. For each $m \in [2^{\gamma k}]$ we have

$$\begin{aligned}\mathbb{P}\left[|\{i\in[d] : X_i\in F\}| = m\right] &= \tfrac{\binom{n}{m}(2^k)^m(2^d-2^k)^{n-m}}{(2^d)^n}\\ &\le n^m 2^{km}\left(\tfrac{2^d-2^k}{2^d}\right)^n\\ &= 2^{(\log n+k)m}\left(\left(1-\tfrac{1}{2^{d-k}}\right)^{2^{d-k}}\right)^{\frac{n}{2^{d-k}}}.\end{aligned}$$

For $d\to\infty$ we have $d-k\to\infty$ (due to $k=\beta d+\mathrm{o}(d)$), and therefore, the expression in the biggest brackets converges to $1/e<1/2$ (where $e=2.71\ldots$ is Euler's constant). Thus, for large enough d, we obtain

$$\begin{aligned}\mathbb{P}\left[|\{i\in[d] : X_i\in F\}| = m\right] &\le 2^{(\log n+k)m}(1/2)^{\frac{n}{2^{d-k}}}\\ &\le 2^{2^{\gamma\beta d+\mathrm{o}(d)}-2^{\alpha d-d+\beta d-\mathrm{o}(d)}}\\ &= 2^{2^{\gamma\beta d+\mathrm{o}(d)}-2^{(\alpha+\beta-1)d-\mathrm{o}(d)}}\end{aligned}$$

Since we have $\alpha+\beta-1>\gamma\beta$, we thus obtain

$$\mathbb{P}\left[|\{i\in[d] : X_i\in F\}| = m\right] \;=\; \mathrm{o}(2^{-2^{\delta d}})$$

for $\delta := (\alpha+\beta-1)/2$ (which is positive due to $\alpha+\beta-1>\gamma\beta>0$).

Therefore, we have

$$\sum_{m\in[2^{\gamma k}]}\mathbb{P}\left[|\{i\in[d] : X_i\in F\}| = m\right] \;\le\; 2^{\gamma k}\,\mathrm{o}(2^{-2^{\delta d}}) \;=\; \mathrm{o}(1)\,,$$

which proves the lemma.

Verification of $\alpha+\beta>1+\beta\gamma$ for (6). We have

$$\begin{aligned}\alpha+\beta &= \tau_r+\varepsilon+1-2^{1-r}\\ &= 1-(1-2^{1-r})H_r+\varepsilon+1-2^{1-r}\\ &= 1+(1-2^{1-r})(1-H_r)+\varepsilon\end{aligned}$$

and

$$\begin{aligned}1+\beta\gamma &= 1+(1-2^{1-r})(1-H_r+\varepsilon)\\ &= 1+(1-2^{1-r})(1-H_r)+(1-2^{1-r})\varepsilon\,,\end{aligned}$$

which proves the claim due to $1-2^{1-r}<1$ and $\varepsilon>0$.

The proof of Lemma 2. For any $x\in[-1,1]^d$, let us define

$$q(x) \;:=\; \frac{1}{2^d}\min\left\{|T^{\le}\cap\{-1,1\}^d| : T^{\le}\subset\mathbb{R}^d \text{ (closed affine) halfspace}, x\in T^{\le}\right\}.$$

For each $\gamma>0$, denote

$$C_d^{\gamma} \;:=\; \{x\in[-1,1]^d : q(x)\ge e^{-\gamma d}\}$$

(where e is Euler's constant). For $\xi\in(-1,1)$ define

$$\frac{1}{\quad}\qquad\qquad\frac{1}{\quad}$$

and for $x = (\xi_1, \ldots, \xi_d) \in (-1,1)^d$, define

$$F(x) \;:=\; \frac{1}{d} \sum_{j \in [d]} f(\xi_j) \;.$$

For every $\mu > 0$, let

$$F_d^\mu \;:=\; \{x \in (-1,1)^d \,:\, F(x) \leq \mu\} \;.$$

Lemma 2.1 of [4] implies (more precisely: the proof of part (b) of that lemma shows) the following:

DFM 1. *Let $\gamma, \delta > 0$. If $n(d) \geq e^{(\gamma+\delta)d}$ for all d and $Y_1, \ldots, Y_{n(d)}$ are chosen independently uniformly at random from $\{-1,1\}^d$, then*

$$\mathbb{P}\left[C_d^\gamma \subseteq \operatorname{conv}\{Y_1, \ldots, Y_{n(d)}\}\right] \;=\; 1 - \mathrm{o}(1)$$

holds.

Lemma 4.1 of [4] states the following:

DFM 2. *Let $\gamma, \delta > 0$. Then, for all large enough d,*

$$F_d^{\gamma-\delta} \;\subseteq\; C_d^\gamma$$

holds.

Let us now show, how the results DFM 1 and DFM 2 imply Lemma 2. The linear transformation $\Psi : \mathbb{R}^d \to \mathbb{R}^d$ defined via

$$\Psi(x) \;:=\; \frac{1}{2}(x + \mathbf{1})$$

maps $[-1,1]^d$ to Q_d by translating and shrinking it. Clearly, we have

$$q(x) \;=\; p(\Psi(x)) \tag{19}$$

for all $x \in [-1,1]^d$.

For each $\alpha > 0$, we have, for every $x \in (-1,1)^d$,

$$\begin{aligned} x \in C_d^{\alpha \ln 2} &\Leftrightarrow\; q(x) \geq e^{-(\alpha \ln 2)d} \\ &\Leftrightarrow\; q(x) \geq 2^{-\alpha d} \\ &\Leftrightarrow\; p(\Psi(x)) \geq 2^{-\alpha d} \\ &\Leftrightarrow\; \Psi(x) \in \mathrm{Q}_d^\alpha \;. \end{aligned}$$

This shows

$$\Psi\big(C_d^{\alpha \ln 2}\big) \;=\; \mathrm{Q}_d^\alpha \;. \tag{20}$$

Hence DFM 1 (applied with $\gamma = \alpha \ln 2$ and $\delta = \varepsilon \ln 2$) implies the first part of Lemma 2.

In order to show the second part, let us first calculate (for each $\xi \in (-1,1)$)

$$f(\xi) \;=\; \frac{1 - h(\frac{1}{2}(\xi+1))}{\log e} \;.$$

Indeed, this follows from the following computation (for every $\zeta \in (0,1)$):

$$\begin{aligned} f(2\zeta - 1) &= \tfrac{1}{2}\big(2\zeta \ln 2\zeta + (2-2\zeta)\ln(2-2\zeta)\big) \\ &= \zeta \ln 2\zeta + (1-\zeta)\ln(2-2\zeta) \\ &= \zeta(\ln 2 + \ln\zeta) + (1-\zeta)(\ln 2 + \ln(1-\zeta)) \\ &= \ln 2 + \zeta \ln\zeta + (1-\zeta)\ln(1-\zeta) \\ &= \ln 2 - \big(\zeta \ln \tfrac{1}{\zeta} + (1-\zeta)\ln \tfrac{1}{(1-\zeta)}\big) \\ &= \tfrac{1}{\log e}\big(\log 2 - \big(\zeta \log \tfrac{1}{\zeta} + (1-\zeta)\log \tfrac{1}{(1-\zeta)}\big)\big) \end{aligned}$$

In particular, we have (for all $x = (\xi_1, \ldots, \xi_d) \in (-1,1)^d$):

$$\begin{aligned} F(x) &= \tfrac{1}{d} \textstyle\sum_{j\in[d]} f(\xi_j) \\ &= \tfrac{1}{d} \textstyle\sum_{j\in[d]} \frac{1-h(\frac{1}{2}(\xi_j+1))}{\log e} \\ &= \tfrac{1-H(\Psi(x))}{\log e} \ . \end{aligned}$$

Thus, for each $\mu > 0$, and for all $x \in (-1,1)^d$

$$\begin{aligned} x \in F_d^{\mu \ln 2} &\Leftrightarrow F(x) \leq \mu \ln 2 \\ &\Leftrightarrow \tfrac{1-H(\Psi(x))}{\log e} \leq \mu \ln 2 \\ &\Leftrightarrow H(\Psi(x)) \geq 1-\mu \end{aligned}$$

holds. This yields

$$\Psi(F_d^{\mu \ln 2}) \ = \ \{z \in \text{int } \mathrm{Q}_d \ : \ H(z) \geq 1-\mu\} \ ,$$

which proves the second part of Lemma 2 by applying DFM 2 with $\gamma = \alpha \ln 2$ and $\delta = \varepsilon \ln 2$.

The partial derivatives of $\tilde{\eta}$. For any $(\zeta_1, \ldots, \zeta_{[r/2]}) \in B$, the i-th partial derivative of $\tilde{\eta}$ at $(\zeta_1, \ldots, \zeta_{[r/2]})$ is

$$\begin{aligned} \frac{\partial \tilde{\eta}}{\partial \zeta_i}(\zeta_1, \ldots, \zeta_{[r/2]}) = & \ \ln \zeta_i + \zeta_i \tfrac{1}{\zeta_i} + (-1) \ln(M_i - \zeta_i) + (M_i - \zeta_i) \tfrac{1}{M_i - \zeta_i}(-1) \\ = & \ \ln \zeta_i - \ln(M_i - \zeta_i) \\ = & \ \ln \tfrac{\zeta_i}{M_i - \zeta_i} \ . \end{aligned}$$

The derivative of $\tilde{\eta}_{z^\star,z}$. If $\mathrm{grad}_{\tilde{z}}\, \tilde{\eta}$ denotes the gradient of $\tilde{\eta}$ at the point $\tilde{z} \in B$, then we have (for $t \in (0,1)$)

$$\tilde{\eta}_{z^\star,z}(t) \ = \ \langle v, \mathrm{grad}_{z^\star + tv}\, \tilde{\eta} \rangle \ ,$$

which by the result of the previous paragraph equals

$$\sum_{i \in [r/2]} v_i \ln \frac{\sigma(i) + t v_i}{M_i - (\sigma(i) + t v_i)} \ ,$$

proving the claim.

Volker Kaibel, DFG Research Center *Mathematics for key technologies*, TU Berlin, MA 6–2, Strasse des 17. Juni 136, 10623 Berlin, Germany, kaibel@math.tu-berlin.de