COVERING THE HYPERCUBE, THE UNCERTAINTY PRINCIPLE, AND AN INTERPOLATION FORMULA

PAATA IVANISVILI, OHAD KLEIN, AND ROMAN VERSHYNIN

Abstract. We show that the minimal number of skewed hyperplanes that cover the hypercube $\{0,1\}^n$ is at least $\frac{n}{2}+1$, and there are infinitely many n's when the hypercube can be covered with $n-\log_2(n)+1$ skewed hyperplanes. The minimal covering problems are closely related to the uncertainty principle on the hypercube, where we also obtain an interpolation formula for multilinear polynomials on $\mathbb{R}^n$ of degree less than $\lfloor n/m \rfloor$ by showing that its coefficients corresponding to the largest monomials can be represented as a linear combination of values of the polynomial over the points $\{0,1\}^n$ whose Hamming weights are divisible by m.

2020 Mathematics Subject Classification. Primary: 06E30; Secondary: 42C10, 68Q32.
Key words. Discrete hypercube, Uncertainty principle, minimal covering, Littlewood–Offord

1. Introduction

1.1. **Covering the hypercube.** How many affine hyperplanes are needed to cover the hypercube $\{-1,1\}^n$? Notice that two affine hyperplanes $x_1=-1$ and $x_1=1$ cover the hypercube, and clearly this is the minimal number. However, if one requires that the affine hyperplanes are skewed, i.e., $a_1x_1+\ldots+a_nx_n+b=0$ with all $a_1,\ldots,a_n\neq 0$, then the problem becomes challenging[1].

It follows from Littlewood–Offord inequalities that any skewed hyperplane covers at most $n^{-1/2}$ fraction of the points in $\{-1,1\}^n$ (up to a universal constant factor), therefore, one needs at least $\Omega(n^{1/2})$ skewed hyperplanes to cover the hypercube. In [7], this lower bound was improved to $\Omega(n^{0.51})$, and recently in [4] to $\Omega(n^{2/3}\log(n)^{-4/3})$ by the second named author of the present paper.

The family of $n+1$ hyperplanes, $x_1+\ldots+x_n=2k-n$ for all $k=0,\ldots,n$, covers the hypercube. In fact, if n is even, one can cover with n skewed hyperplanes just by replacing the two hyperplanes corresponding to $k=0$ and $k=n$ in the previous example with one hyperplane $x_1+\ldots+x_{n/2}-x_{n/2-1}-\ldots-x_n=0$. Moreover, it follows from [1] that for even n, the upper bound n on the minimal cover is also a lower bound if one restricts the covering to the family of "regular" hyperplanes, i.e., the ones $\varepsilon_1x_1+\ldots+\varepsilon_nx_n+b=0$ with $\varepsilon_j=\pm1$ for all $j=1,\ldots,n$.

Looking at the results for the case of "regular" hyperplane cover in [1], one may suspect that in analogy to Littlewood–Offord problem the sharp lower bound on the minimal skew hyperplane cover should be n. Surprisingly, one can cover the hypercube $\{-1,1\}^5$ with the following 4 skewed hyperplanes

$$\begin{aligned}
x_1+x_2+x_3+x_4+2x_5&=0,\\
x_1+x_2+x_3-x_4+2x_5&=0,\\
x_1+x_2+x_3+x_4-2x_5&=0,\\
x_1+x_2+x_3-x_4-2x_5&=0.
\end{aligned}$$

Also the hypercube $\{-1,1\}^6$ can be covered with 5 skewed hyperplanes (see Section 2.3). In fact, one can cover the hypercube $\{-1,1\}^n$ with $n-\log_2(n)+1$ skewed hyperplanes for infinitely many n's.

[1]In what follows we will omit the word affine and we will be referring to such hyperplanes as skewed hyperplanes.

Proposition 1. *For any integer $m \geq 1$ the hypercube $\{-1,1\}^{2^m+m-1}$ can be covered with 2^m skewed hyperplanes.*

This proposition shows that the skew hyperplane covering problem is genuinely different from the original "regular" problem solved in [1].

Question 2. *What is the minimal number of skewed hyperplane cover of the hypercube $\{-1,1\}^n$?*

We prove the following lower bound.

Theorem 3. *The minimal number of skewed hyperplane cover of $\{-1,1\}^n$ is at least $\frac{n}{2}+1$.*

There is a close relation between the minimal hyperplane covering problem and the uncertainty principle on the hypercube. Let $p(x)$ be a polynomial on $\mathbb{R}^n$, and let $\mathrm{supp}(p) = \{x \in \mathbb{R}^n : p(x) \neq 0\}$. Under what conditions on $\mathrm{supp}(p) \cap \{-1,1\}^n$ and $\deg(p)$ does it follow that $p(x) \equiv 0$ on $\{-1,1\}^n$?

It turns out that the support of a nonzero low degree polynomial cannot be contained in a skewed hyperplane:

Theorem 4 (Linial–Radhakrishnan [5])**.** *If $\deg(p) < \frac{n}{2}$ and $\mathrm{supp}(p) \cap \{-1,1\}^n$ belongs to a skewed hyperplane, then $p(x) \equiv 0$ on $\{-1,1\}^n$.*

Observe that Theorem 3 follows from Theorem 4. Indeed, let $H_1,\dots,H_{k+1}$ be a minimal skew hyperplane cover of $\{-1,1\}^n$. If $H_j's$ are given via equations $\ell_j(x) = a_{1j}x_1 + \ldots + a_{nj}x_n + b_j = 0$, for all $j = 1,\dots,k+1$, then it follows that $p(x)\ell_{k+1}(x) \equiv 0$ on $\{-1,1\}^n$, where $p(x) = \prod_{j=1}^{k} \ell_j(x)$ is a not identically zero polynomial on $\{-1,1\}^n$ of degree at most k. Hence, $\mathrm{supp}(p) \cap \{-1,1\}^n$ belongs to H_{k+1} and Theorem 4 implies that $k \geq n/2$.

After the current paper was completed, independently and concurrently the paper [6] appeared on arXiv where Theorem 3 was derived from Theorem 4 proved in [5] (see Lemma 2 in [5]). The proof of Theorem 4 in [5] in turn is based on either Combinatorial Nullstellensatz or the spectral properties of the Johnson graph (the authors [5] attribute the nonsingularity of the Johnson graph to [3]). Our proof of Theorem 4, given in Section 2.1, is simple and self-contained.

1.2. **An interpolation formula.** In [1] sharp lower bound n on the minimal number of "regular" hyperplane cover of the n dimensional hypercube (for even n) was based on the following technical observation: if a multilinear polynomial $p(x)$ vanishes on all those points of $\{-1,1\}^n$ which have even number of 1's in its coordinates, and $\deg(p) < n/2$, then p is identically zero (see Lemma 2.1 in [1]). This observation suggests that perhaps the coefficients of the multilinear polynomials of small degree can be reconstructed by its values at sparse points of $\{-1,1\}^n$. The goal of this section is to obtain such an interpolation formula.

Recall that any function $f : \{-1,1\}^n \mapsto X$, where X is a normed space, has Fourier–Walsh representation

$$f(x) = \sum_{S \subset \{1,\dots,n\}} \widehat{f}(S) x^S, \tag{1}$$

for some $\widehat{f}(S) \in X$, where $x = (x_1,\dots,x_n)$, $x^S = \prod_{j \in S} x_j$ and $x^{\emptyset} = 1$. We say that f has degree $\deg(f)$ if $\widehat{f}(S) = 0$ for all $S \subset \{1,\dots,n\}$ with $|S| > \deg(f)$, and there exists a subset S of cardinality $\deg(f)$ such that $\widehat{f}(S) \neq 0$.

Definition 5. *For any integer $m > 1$ the symbol $W(m)$ denotes the subset of $\{-1,1\}^n$ consisting of all points $x = (x_1,\dots,x_n) \in \{-1,1\}^n$ such that $\#\{j : x_j = -1\}$ is divisible by m.*

In this section we obtain the following interpolation formula

Theorem 6. *Let $f\colon \{-1,1\}^n \mapsto X$ and let $m \geq 2$ be an integer divisible by 2 such that*

$$\deg(f) \leq \frac{n}{m} - \frac{1}{2}. \tag{2}$$

Then for any $S \subset [n]$ with $|S| = \deg(f)$, there exists a probability measure $d\mu(x)$ supported on $W(m)$ and a sign function $h\colon W(m) \mapsto \{-1,1\}$ such that

$$\widehat{f}(S) = \int_{W(m)} h(x) f(x) d\mu(x) \tag{3}$$

Both $d\mu$ and h depend only on $S, m, \deg(f), n$.

The next corollary follows from the theorem

Corollary 7. *If $f\colon \{-1,1\}^n \mapsto X$ vanishes on a set $W(m)$ for some even integer m satisfying (2), then $f \equiv 0$.*

Remark 8. *When $m = 2$, Corollary 7 is the classical result* [1, Lemma 2.1].

Remark 9. *In the proof of Theorem 6 both the measure $d\mu$ and $h(x)$ are constructed explicitly.*

Notice that since $\widehat{f}(S) = \mathbb{E} f(x) x^S$ then $\|\widehat{f}(S)\| \leq \max_{x \in \{-1,1\}^n} \|f(x)\|$. However, if f has low degree, then $\max_{x \in \{-1,1\}^n} \|f(x)\|$ can be replaced by a maximum over sparse family of points of $\{-1,1\}^n$ provided that $|S| = \deg(f)$. Indeed, Theorem 6 gives the following

Corollary 10. *Let $f: \{-1,1\}^n \mapsto X$, where X is a normed space, be a function whose degree satisfies (2), then*

$$\|\widehat{f}(S)\| \leq \max_{x \in W(m)} \|f(x)\|$$

for all $S \subset \{1, \ldots, n\}$ with $|S| = \deg(f)$.

2. Proofs

2.1. The proof of Theorem 4.

Denote $[n] := \{1, \ldots, n\}$. Every polynomial $p(x)$ of degree d, when restricted to $\{-1,1\}^n$, can be written as $f(x) = \sum_{|S| \leq d} c_S x^S$ for some $c_S \in \mathbb{R}$. The assumption in Theorem 4 that the support of f is contained within a skewed hyperplane means that

$$\forall x \in \{-1,1\}^n\colon\ (a_1 x_1 + \ldots + a_n x_n + b) \sum_{|S| \leq d} c_S x^S = 0, \tag{4}$$

where

$$\forall i\colon\ a_i \neq 0. \tag{5}$$

When expanding (4), the right hand side means that all coefficients of the monomials x^T must vanish, in particular for degree $d+1$ monomials. This means that

$$\sum_{j \in T} a_j c_{T \setminus j} = 0 \tag{6}$$

for all $T \subseteq [n]$ with $|T| = d + 1$. We can view (6) as a system of linear equations $Ac = 0$ in $c = (c_S : S \subset [n], |S| = d)$. It suffices to prove

Lemma 11. *We have* $\mathrm{Ker}(A) = 0$ *as long as* $n \geq 2d + 1$.

The theorem follows from the lemma as follows: The lemma implies that in (4) we have $c_S = 0$ for all $S \subset [n]$ with $|S| = d$. This means that $p(x)$ is in fact a polynomial of degree $d - 1$. Since d was merely defined as the degree of p (and assumed to satisfy $2d+1 \leq n$), we can repeat and deduce p is of degree $d-2$ and similarly of degree 0. Once $n \geq 2$ a hyperplane can not cover the entire hypercube, so p must be the zero polynomial, concluding the proof of Theorem 4.

Proof of Lemma 11. Note that it is sufficient to prove the lemma only for $n = 2d+1$, and it would follow for any $n \geq 2d+1$. To see this, let $n > 2d+1$ and let $S \subset [n]$ with $|S| = d$; we must show $c_S = 0$. Fix a set $N \subset [n]$ with $S \subset N$ and $|N| = 2d+1$, and focus on equations (6) for $T \subset N$. From the $n = 2d+1$ case we conclude that all involved variables $c_{S'}$ with $S' \subset N$ are 0, and in particular $c_S = 0$.

Next, we prove the lemma by induction on d for the $n = 2d+1$ case.

Base case: When $d = 1$ and $n = 3$, Equation (6) applied on the sets $T = \{2,3\}, \{1,3\}, \{1,2\}$ yields:

$$\begin{aligned} a_3c_2 + a_2c_3 &= 0 \\ a_1c_3 + a_3c_1 &= 0 \\ a_1c_2 + a_2c_1 &= 0, \end{aligned}$$

Or in matrix form:

$$\begin{bmatrix} 0 & a_3 & a_2 \\ a_3 & 0 & a_1 \\ a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = 0.$$

The determinant of this 3×3 matrix equals $2a_1a_2a_3$, which is nonzero by (5). We learn that $c_i = 0$ and the lemma follows in this case.

Inductive step: Assume that the lemma holds for $d-1$, we prove it for d and $n = 2d+1$. Let c be a solution to (6). In order to complete the induction step, we must show that $c = 0$.

Note that the induction hypothesis applied for $d-1$ and $n-2$ (that has $n-2 = 2(d-1)+1$) implies that $\bar{c} = 0$ is the unique solution to the system of equations

$$a_{i_1}\bar{c}_{i_2,i_3,\ldots,i_d} + a_{i_2}\bar{c}_{i_1,i_3,\ldots,i_d} + \cdots + a_{i_d}\bar{c}_{i_1,i_2,\ldots,i_{d-1}} = 0 \quad \forall \{i_1,\ldots,i_d\} \subset [n-2], \tag{7}$$

where $\{i_1,\ldots,i_d\}$ ranges over all subsets of size d of $[n-2]$.

Fix $j := i_{d+1} \in \{n-1, n\}$, and consider all those linear equations (6) in c that arise for $T = \{i_1,\ldots,i_d,j\}$ where $\{i_1,\ldots,i_d\} \subset [n-2]$:

$$a_{i_1}c_{i_2,i_3,\ldots,i_d,j} + a_{i_2}c_{i_1,i_3,\ldots,i_d,j} + \cdots + a_{i_d}c_{i_1,i_2,\ldots,i_{d-1},j} + a_jc_{i_1,i_2,\ldots,i_d} = 0 \quad \forall \{i_1,\ldots,i_d\} \subset [n-2].$$

Moving the last term into the right hand side, we get:

$$a_{i_1}c_{i_2,i_3,\ldots,i_d,j} + a_{i_2}c_{i_1,i_3,\ldots,i_d,j} + \cdots + a_{i_d}c_{i_1,i_2,\ldots,i_{d-1},j} = -a_jc_{i_1,i_2,\ldots,i_d} \quad \forall \{i_1,\ldots,i_d\} \subset [n-2]. \tag{8}$$

Compare this system with the one in (7). This is the same system, up to a relabeling of the variables, and different right hand side. Let A_{d-1} be the matrix defining the $d-1$ case, then (8) can be written as

$$A_{d-1}\bar{c} = -a_ju,$$

where u is the vector of variables $c_{i_1,i_2,\ldots,i_d}$ and $\bar{c}$ is the vector of variables $c_{i_1,i_2,\ldots,i_d,j}$ for $\{i_1,i_2,\ldots,i_d\} \subset [n-2]$. Note that since $\binom{2d-1}{d-1} = \binom{2d-1}{d}$, the matrix A_{d-1} is square, and is hence invertible by the induction hypothesis. In particular,

$$\bar{c} = -a_jA_{d-1}^{-1}u.$$

Note that on the right hand side, the only thing that depends on j is a_j. That is, both A_{d-1} and u do not depend on whether $j = n-1$ or $j = n$. By comparing the two options $j = n-1$, and $j' = n$, we conclude that for all $i_1, i_2, \ldots, i_d \subset [n-2]$,

$$c_{i_1,i_2,\ldots,i_{d-1},j}/a_j = c_{i_1,i_2,\ldots,i_{d-1},j'}/a_{j'}. \tag{9}$$

Note that in (9), we can choose the indices $i_1,\ldots,i_d,j',j$ arbitrarily so long as they are distinct.

Claim 12. *Equation* (9) *implies the existence of a single constant* $K \in \mathbb{R}$ *such that*

$$c_S = K\prod_{i \in S} a_i, \tag{10}$$

for all $S \subset [n]$ *with* $|S| = d$.

Plugging the formula for c_S from Claim 12 into the system of equations (6), we get that for any $T \subset [n]$ with $|T| = d+1$,

$$0 = \sum_j a_j c_{T\setminus\{j\}} = \sum_{j\in T} a_j \left(K \prod_{i\in T\setminus\{j\}} a_i \right) = \sum_{j\in T} K \prod_{i\in T} a_i = (d+1)K \prod_{i\in T} a_i$$

In (5), we assumed $a_i \neq 0$ for all i, so it follows that $K = 0$. and in particular $c_S = 0$ for all $S \subset [n]$. The induction step and the lemma follows.

Proof of Claim 12. Let $I \subset [n]$ and $J, K \subset [n] \setminus I$ be sets with $|J| = |K| = d - |I|$. We prove by induction on the size of J and K that

$$c_{I\cup J} / \prod_{j\in J} a_j = c_{I\cup K} / \prod_{k\in K} a_k. \tag{11}$$

When J, K are of size d, and $I = \emptyset$, then we derive (10) and the proof is complete.
Base case: When J, K are of size 0, (11) is obvious as $J = K$.
Inductive step: Let $J, K \subset [n]$ be of size $\ell \geq 1$ and let $I \subset [n] \setminus (J \cup K)$ be of size $d - \ell$. We prove (11). Take $j \in J$ and $k \in K$ and denote $I' = I \cup \{j\}$, $J' = J \setminus \{j\}$ and $K' = K \setminus \{k\}$. Then (11) follows:

$$\begin{aligned} c_{I\cup J} / \prod_{i\in J} a_i &\underbrace{=}_{I\cup J = I'\cup J'} \frac{1}{a_j} c_{I'\cup J'} / \prod_{i\in J'} a_i \\ &\underbrace{=}_{\text{induction}} \frac{1}{a_j} c_{I'\cup K'} / \prod_{i\in K'} a_i \\ &\underbrace{=}_{(9)} \frac{1}{a_k} c_{I\cup K} / \prod_{i\in K'} a_i \\ &= c_{I\cup K} / \prod_{i\in K} a_i. \end{aligned}$$

□

□

2.2. **The proof of Theorem 6.** Let $d = \deg(f)$ and assume that $S = \{m, 2m, \ldots, dm\}$ (any S with $|S| = d$ is obtained by relabeling of variables). Recall Equation (1) defining the Fourier representation: for any $u = (u_1, \ldots, u_n) \in \{-1,1\}^n$ we have

$$f(u) = \sum_{|S|\leq d} \widehat{f}(S) u^S \quad \text{where} \quad u^S = \prod_{j\in S} u_j.$$

For $(y_1, \ldots, y_d) \in \{-1,1\}^d$, and $(x_1, \ldots, x_n) \in \{-1,1\}^n$, we define $y \circ x \in \{-1,1\}^n$ by splitting the vector x into disjoint sets of indices $I_1, I_2, \ldots, I_d, I_{\text{extra}}, I_{\text{rest}}$, where

$$\begin{aligned} I_1 &= (1, \ldots, m), \\ I_2 &= (m+1, \ldots, 2m), \\ &\ldots \\ I_d &= ((d-1)m+1, \ldots, dm), \\ I_{\text{extra}} &= (md+1, \ldots, md+m/2), \\ I_{\text{rest}} &= (md+m/2+1, \ldots, n). \end{aligned}$$

Then, we define

$$y_j x_{I_j} := (y_j x_{(j-1)m+1}, y_j x_{(j-1)m+2}, \ldots, y_j x_{jm}),$$

and finally,

$$\begin{aligned}
y \circ x &:= (y_1 x_{I_1}, \ldots, y_d x_{I_d}, x_{I_{\text{extra}}}, x_{I_{\text{rest}}}) \\
&= (y_1 x_1,\ y_1 x_2, \ldots,\ y_1 x_m, \\
&\quad\ y_2 x_{m+1},\ y_2 x_{m+2}, \ldots,\ y_2 x_{2m}, \\
&\qquad\qquad \vdots \\
&\quad\ y_d x_{(d-1)m+1},\ y_d x_{(d-1)m+2}, \ldots,\ y_d x_{md}, \\
&\quad\ x_{md+1},\ x_{md+2}, \ldots,\ x_{md+m-\frac{m}{2}}, \ldots,\ x_n)
\end{aligned}$$

Note that the variables in I_{extra} and I_{rest} are unchanged. The coordinates I_{rest} do not play a role in the proof (and may be empty, e.g. if we have equality in (2)), but the coordinates I_{extra} have the important role of "parity" in the proof.

Claim 13. *There exists a distribution $\mathcal{D}$ for x (on $\{-1,1\}^n$) such that*

$$\widehat{f}(S) = \mathop{\mathbb{E}}_{x\sim\mathcal{D}} \ \mathop{\mathbb{E}}_{y\sim \text{unif}(\{-1,1\}^d)} [f(y \circ x)\cdot y_1 \cdots y_d], \tag{12}$$

and moreover,

$$y \circ x \in W(m) \tag{13}$$

for all $y \in \{-1,1\}^d$ and all $x \in \text{supp}(\mathcal{D})$. $\mathcal{D}$ depends only on d, m, n but not on f.

Claim 13 concludes the proof of (3) by using the sign function $h = y_1 \cdots y_d$ and the measure $d\mu$ depicting the distribution[2] of $y \circ x$ where $x \sim D$ and $y \sim \text{unif}(\{-1,1\}^d)$.

Proof of Claim 13. Observe the formula

$$\mathop{\mathbb{E}}_{y\sim \text{unif}(\{-1,1\}^d)} [f(y\circ x) y_1 \ldots y_d] = \sum_{T:\ \forall j:\ |T\cap I_j|=1} \widehat{f}(T) x^T. \tag{14}$$

To verify (14), we expand $f(y\circ x)$ on the left hand side as $\sum_T \widehat{f}(T)(y\circ x)^T$. The number of times y_j appears in that expression is exactly $|T \cap I_j|$. If $T \cap I_j$ is empty for some j, then $(y\circ x)^T$ does not depend on y_j, and the multiplication by $y_1 \ldots y_d$ in (14) zeroes out the term $\widehat{f}(T)(y\circ x)^T$. Hence relevant terms are only those with $|T\cap I_j| \geq 1$ for all $j = 1, \ldots, d$. But since f is of degree d to begin with, we must have $|T \cap I_j| = 1$ for all $j = 1, \ldots, d$.

The distribution $\mathcal{D}$. We describe how each chunk x_{I_j} of $x \sim \mathcal{D}$ is drawn. All chunks are drawn independently of the other chunks, except for $x_{I_{\text{extra}}}$ which is chosen last.

- x_{I_j} for $j = 1, \ldots, d$:

$$\Pr[x_{I_j} = (z_1, \ldots, z_m)] = \begin{cases} 1/m & \text{if } z_1 = \cdots = z_m = 1, \\ \frac{1}{2\binom{m-2}{m/2-1}} & \text{if } z_m = 1 \text{ and exactly } m/2 \text{ among } z_1, \ldots, z_{m-1} \text{ are equal } -1, \\ 0 & \text{otherwise.} \end{cases}$$

  Note that the sum over all probabilities is 1.
- $x_{I_{\text{rest}}}$: we set $x_{I_{\text{rest}}} = (1,1,\ldots,1)$ always.
- $x_{I_{\text{extra}}}$: Count the total number s of -1's in all chunks $x_{I_1}, x_{I_2}, \ldots, x_{I_d}$. Define

$$x_{I_{\text{extra}}} = \begin{cases} (1, \ldots, 1) & \text{if } m|s, \\ (-1, \ldots, -1) & \text{otherwise.} \end{cases} \tag{15}$$

[2]Note that here we define h as a function of y while the measure $d\mu$ is of $\{-1,1\}^n$. We claim that (12) implies that $y_1 \cdots y_d$ is uniquely determined from $y\circ x$ for all $y \circ x$ having positive probability. To see this, note that formula (12) does not depend on f, yet for $\widetilde{f}(z) = z^S$ it has 1 on the LHS while the RHS is bounded by 1 by the triangle inequality. This means $\widetilde{f}(y\circ x) = y_1 \cdots y_d$. Consequently, (3) holds with $h = y_1 \cdots y_d = \widetilde{f}(y \circ x)$, which is a function of $y \circ x$.

Observe that necessarily s is divisible by $m/2$, since each choice of x_{I_j} adds either 0 or $m/2$ to s. For this reason, (15) defines x is such a way that $x \in W(m)$. Furthermore, for all $y \in \{-1,1\}^d$ we have $y \circ x \in W(m)$ essentially because signs do not matter modulo 2.

Finally, in order to deduce (12) from (14), we must check that for all $T \subseteq \{1,\ldots,n\}$ with $\forall j \colon |T \cap I_j| = 1$ we have

$$\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[x^T] = \begin{cases} 1 & \text{if } T = S, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

The case $T = S$ is immediate, since by design $x_{jm} = 1$ for all $j = 1,\ldots,d$.

Suppose $T \neq S$. Focus on a particular coordinate $t \in T$ with $m \nmid t$ and let $j \in \{1,\ldots,d\}$ be the index with $t \in I_j$. Since $T \cap I_j = \{t\}$, we have that x^T is $x_t \cdot x^{T\setminus\{t\}}$. When $x \sim \mathcal{D}$, $x^{T\setminus\{t\}}$ is a random variable independent of x_t, as we draw the different chunks independently. Hence $\mathbb{E}_{x\sim\mathcal{D}}[x^T] = \mathbb{E}_{x\sim\mathcal{D}}[x_t] \cdot \mathbb{E}_{x\sim\mathcal{D}}[x^{T\setminus\{t\}}]$. In order to deduce (16) and finish the proof we just need to check that $\mathbb{E}_{x\sim\mathcal{D}}[x_t] = 0$. Indeed, by definition, the probability that $x_t = 1$ is $1/m + \frac{\binom{m-2}{m/2}}{2\binom{m-2}{m/2-1}}$, that is, either x_{I_j} is all 1's, or we need to choose $m/2$ locations for -1's in I_j out of $I_j \setminus \{t, mj\}$. This probability is 1/2, validating $\mathbb{E}_{x\sim\mathcal{D}}[x_t] = 0$, concluding the proof.

2.3. **The proof of Proposition 1: how to cover the hypercube efficiently.** Consider 2^m skewed hyperplanes

$$\sum_{j=1}^{2^m-1} x_j + \sum_{j=0}^{m-1} \pm 2^j x_{2^m+j} = 0.$$

Since any odd integer k, $-(2^m-1) \leq k \leq 2^m-1$ can be written as a sum $\sum_{j=0}^{m-1} \pm 2^j$ for some choice of signs $\pm$, it follows that these hyperplanes cover the cube $\{-1,1\}^n$ with $n = 2^m + m - 1$. □

There are other examples that are not produced by the construction above. In particular, for $n = 6$ the union of the following 5 skewed hyperplanes

$$\begin{aligned} x_1 - x_2 + 2x_3 + x_4 + x_5 + 2x_6 &= 0, \\ x_1 - x_2 + x_3 + x_4 + x_5 - x_6 &= 0, \\ x_1 - x_2 - x_3 + 2x_4 - 2x_5 + x_6 &= 0, \\ x_1 + x_2 + x_3 + x_4 + x_5 - x_6 &= 0, \\ x_1 - x_2 - 3x_3 + x_4 + x_5 - x_6 &= 0. \end{aligned}$$

cover the hypercube $\{-1,1\}^6$.

Acknowledgement

We are very grateful to the anonymous referee for the helpful suggestions and insightful comments, which have made the paper clearer and easier to read. P.I. acknowledges support from NSF grant CAREER-DMS-2152401. O.K. was supported in part by a grant from the Israel Science Foundation (ISF Grant No. 1774/20), and by a grant from the US-Israel Binational Science Foundation and the US National Science Foundation (BSF-NSF Grant No. 2020643). R.V. acknowledges support from NSF DMS-1954233, NSF DMS-2027299, U.S. Army 76649-CS, and NSF+Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning.

References

[1] N. Alon, E. E. Bergmann, D. Coppersmith, A. M. Odlyzko, *Balancing sets of vectors*. IEEE Transactions on Information Theory, 34(1):128–130, 1988

[2] A. E. Brouwer, A. M. Cohen, A. Neumaier, *Distance-Regular Graphs*. (1989) Berlin, Heidelberg: Springer Berlin Heidelberg

[3] D. H. Gottlieb, *A certain class of incidence matrices*. Proc. Amer. Math. Soc. 17 (1966) 1233-1237.

[4] O. Klein, *Slicing all Edges of an n-cube requires $n^{2/3}$ hyperplanes*. (2022) arXiv: 2102.05536

[5] N. Linial, J. Radhakrishnan, *Essential covers of the cube by hyperplanes*. Journal of Combinatorial Theory, Series A 109.2 (2005), pp. 331-338.

[6] L. Sauermann, Z. Xu, *Essential covers of the hypercube require many hyperplanes*. (2023) arXiv: 2310.05775

[7] G. Yehuda, A. Yehudayoff, *Slicing the hypercube is not easy*. (2021) arXiv: 2102.05536

(P. I.) Department of Mathematics, University of California, Irvine, Irvine, CA 92697, USA.
Email address: pivanisv@uci.edu

(O. K.) School of Computer Science and Engineering, Hebrew University of Jerusalem, Jerusalem, Israel
Email address: ohadkel@gmail.com

(R. V.) Department of Mathematics, University of California, Irvine, Irvine, CA 92697, USA.
Email address: rvershyn@uci.edu