On the Gap between Hereditary Discrepancy and the Determinant Lower Bound

Lily Li[*]   Aleksandar Nikolov[†]

January 18, 2024

Abstract

The determinant lower bound of Lovász, Spencer, and Vesztergombi [European Journal of Combinatorics, 1986] is a general way to prove lower bounds on the hereditary discrepancy of a set system. In their paper, Lovász, Spencer, and Vesztergombi asked if hereditary discrepancy can also be bounded from above by a function of the determinant lower bound. This was answered in the negative by Hoffman, and the largest known multiplicative gap between the two quantities for a set system of m subsets of a universe of size n is on the order of $\max\{\log n, \sqrt{\log m}\}$. On the other hand, building upon work of Matoušek [Proceedings of the AMS, 2013], Jiang and Reis [SOSA, 2022] showed that this gap is always bounded up to constants by $\sqrt{\log(m)\log(n)}$. This is tight when m is polynomial in n, but leaves open the case of large m. We show that the bound of Jiang and Reis is tight for nearly the entire range of m. Our proof amplifies the discrepancy lower bounds of a set system derived from the discrete Haar basis via Kronecker products.

1 Introduction

Let X be a finite universe of elements, $\mathcal{S}$ a collection of subsets of X called a *set system*, and $\chi : X \to \{\pm 1\}$ a ± 1 coloring of the elements of X. The discrepancy of $\mathcal{S}$ with respect to χ, denoted $\mathrm{disc}(\mathcal{S}, \chi)$, is defined as $\max_{S \in \mathcal{S}} ||\chi^{-1}(1) \cap S| - |\chi^{-1}(-1) \cap S||$ i.e. the largest difference between the number of elements colored differently in any set $S \in \mathcal{S}$. The *discrepancy of* $\mathcal{S}$, denoted $\mathrm{disc}(\mathcal{S})$, is then $\min_{\chi : X \to \{\pm 1\}} \mathrm{disc}(\mathcal{S}, \chi)$ i.e. among all ± 1 colorings of the elements of X, the least unequal we can make the most unequal set of $\mathcal{S}$. If $|X| = n$ and $|\mathcal{S}| = m$, let $X = [n]$ and $\mathcal{S} = \{S_1, ..., S_m\}$. For each set system $\mathcal{S}$, we can construct an associated $m \times n$ incidence matrix $\mathbf{A}_\mathcal{S}$: the entry in row i and column j of $\mathbf{A}_\mathcal{S}$ is equal to one if element j is in S_i and zero otherwise. We define the *discrepancy of a real-valued matrix* $\mathbf{A}$ as

$$\mathrm{disc}(\mathbf{A}) := \min_{\mathbf{x} \in \{\pm 1\}^n} \|\mathbf{A}\mathbf{x}\|_\infty,$$

where $\|\cdot\|_\infty$ denotes the L_∞-norm of a vector. Using this definition, we see that $\mathrm{disc}(\mathbf{A}_\mathcal{S}) = \mathrm{disc}(\mathcal{S})$. Throughout this exposition, we will use $\mathcal{S}$ and its indicator matrix $\mathbf{A}_\mathcal{S}$ interchangeably.

We would like discrepancy to be a robust quantity, but it can be sensitive to slight modifications to the incidence matrix e.g. the discrepancy of the matrix $[\mathbf{A}, \mathbf{A}]$ is always zero regardless of $\mathrm{disc}(\mathbf{A})$. This motivates the definition of the hereditary discrepancy of a matrix $\mathbf{A}$ as

$$\mathrm{herdisc}(\mathbf{A}) := \max_{\mathbf{B} \in \mathcal{S}} \mathrm{disc}(\mathbf{B}),$$

where $\mathcal{S}$ is the set of all sub-matrices of $\mathbf{A}$. Notice that adding rows to a matrix can never decrease its discrepancy so it suffices for $\mathcal{S}$ to consist of sub-matrices whose columns are a subset of the columns of $\mathbf{A}$. In

---

[*]University of Toronto, Department of Computer Science (xinyuan@cs.toronto.edu).

[†]University of Toronto, Department of Computer Science (anikolov@cs.toronto.edu).

a sense, this definition generalizes total unimodularity. It is easy to show that totally unimodular matrices (TUM)[1] have hereditary discrepancy at most one [Sch98]. Further, a result of Ghouila-Houri [GH62] states that the set of hereditary discrepancy one matrices with entries in $\{-1, 0, 1\}$ is exactly the set of TUM matrices.

An important early work in discrepancy theory of Lovász, Spencer, and Vesztergombi [LSV86] showed that the determinant lower bound of $\mathbf{A}$, defined as

$$\text{detlb}(\mathbf{A}) := \max_k \max_{\mathbf{B} \in \mathcal{S}_k} |\det(\mathbf{B})|^{1/k},$$

where $\mathcal{S}_k$ denotes the set of all $k \times k$ sub-matrices of $\mathbf{A}$, satisfies $2\text{herdisc}(\mathbf{A}) \geq \text{detlb}(\mathbf{A})$. This, again, generalizes what happens with totally unimodular matrices, for which both quantities are equal to one. The determinant lower bound has since become a powerful tool in proving nearly tight lower bounds on many natural and important set systems, e.g. axis-aligned boxes and point-line incidences [CL01, MNT18]. Given that the determinant lower bound often implies nearly tight discrepancy lower bounds, it is natural to ask how far it can be from the hereditary discrepancy. The first result in this direction is due to Matoušek [Mat13] who showed that the ratio between hereditary discrepancy and the determinant lower bound is bounded from above as

$$\frac{\text{herdisc}(\mathbf{A})}{\text{detlb}(\mathbf{A})} \lesssim \log(2mn) \cdot \sqrt{\log(2n)}.$$

Here we used the notation $a \lesssim b$ to denote the the existence of a universal constant c such that $a \leq c \cdot b$. Similarly, $a \gtrsim b$ denotes the existence of a universal constant $c > 0$ such that $a \geq c \cdot b$, and $a \cong b$ when $a \lesssim b$ and $a \gtrsim b$.

Matoušek's bound was not believed to be tight as the largest known value of $\frac{\text{herdisc}(\mathbf{A})}{\text{detlb}(\mathbf{A})}$ is on the order of $\log n$. Both the large discrepancy three permutations family of Newman, Neiman, and Nikolov [NNN12] (see also [Fra18]) and a construction due to Pálvölgyi [Pál10] achieve this gap. Matoušek's bound follows from a pair of inequalities

$$\text{herdisc}(\mathbf{A}) \lesssim \log(2mn) \cdot \text{hervecdisc}(\mathbf{A}), \tag{1}$$

$$\text{hervecdisc}(\mathbf{A}) \lesssim \sqrt{\log 2n} \cdot \text{detlb}(\mathbf{A}), \tag{2}$$

where the first inequality is implied by the seminal result of Bansal [Ban10], and the second inequality is proved using duality. Note that vecdisc is the *vector discrepancy* of a set system, or matrix. This quantity is similar to discrepancy but the elements of the universe are "colored" by vectors rather than by ± 1. In particular, for an $m \times n$ matrix $\mathbf{A}$,

$$\text{vecdisc}(\mathbf{A}) = \min_{\mathbf{v}_1, \ldots, \mathbf{v}_n \in S^{n-1}} \max_{j \in [m]} \left\| \sum_{i \in [n]} A_{j,i} \cdot \mathbf{v}_i \right\|_2,$$

where S^{n-1} is the unit sphere in $\mathbb{R}^n$. Note that vector discrepancy is a lower bound on discrepancy since a coloring $\chi : X \to \{\pm 1\}$ can be interpreted as a set of vectors where all vectors are parallel to each other. The *hereditary vector discrepancy of* $\mathbf{A}$, denoted hervecdisc, which appears in (1), is the maximum vector discrepancy of any subset of the columns of $\mathbf{A}$.

Recently Jiang and Reis [JR22] were able to improve Matoušek's result by showing that $\frac{\text{herdisc}(\mathbf{A})}{\text{detlb}(\mathbf{A})} \lesssim \sqrt{\log 2m \log 2n}$. Their work left open whether the $\sqrt{\log m}$ term can be replaced by $\sqrt{\log n}$ for large m. In the present work, we show that this is mostly not possible, and that the factor of $\sqrt{\log m}$ is necessary for all m in the range $n \leq m \leq 2^{n^{1-\epsilon}}$ for any constant $\epsilon > 0$. Note that when $m > 2^n$, the matrix contains duplicated rows whose removal will not change the value of herdisc($\mathbf{A}$) nor detlb($\mathbf{A}$). Thus, our lower bound covers nearly the whole range of values for m.

Our main result is stated in the next theorem. Its proof appears at the end of Section 3.

---

[1] A matrix $\mathbf{A}$ is TUM if every square submatrix of $\mathbf{A}$ has determinant in $\{-1, 0, 1\}$. A linear systems of the form $\mathbf{A}\mathbf{x} \geq \mathbf{b}$ for TUM $\mathbf{A}$, integral $\mathbf{b}$, and $0 \leq \mathbf{x}$ has an integral polyhedron as its feasible region.

Theorem 1. *For any real number $\epsilon \in (0,1)$, any integers $n \geq 2$ and $m \in \left[n, 2^{n^{1-\epsilon}}\right]$, there exists a matrix $\mathbf{A} \in \{0,1\}^{m \times n}$ such that*

$$\frac{\text{herdisc}(\mathbf{A})}{\text{detlb}(\mathbf{A})} \gtrsim \sqrt{\log m \log n}. \tag{3}$$

Note that the lower bound in Theorem 1 only holds for $m \leq 2^{n^{1-\epsilon}}$ for an arbitrarily small but fixed constant ε. This leaves open whether such a lower bound holds all the way to $m = 2^n$. The next theorem gives a new upper bound on herdisc($\mathbf{A}$) in terms of detlb($\mathbf{A}$), which implies that Theorem 1 cannot be extended to $m = 2^{\omega(n/\log n)}$.

Theorem 2. *For all positive integers m and n, and all matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$, we have*

$$\frac{\text{herdisc}(\mathbf{A})}{\text{detlb}(\mathbf{A})} \lesssim \sqrt{n}.$$

This upper bound is based on the relationship between the volume lower bound on discrepancy studied in [DNTTJ18], and the determinant lower bound. In particular, we show that the volume lower bound is bounded by a constant multiple of $\sqrt{n} \cdot \text{detlb}(\mathbf{A})$, and use a result from [DNTTJ18] characterizing the hereditary discrepancy of partial colorings in terms of the volume lower bound. We also give a simpler proof for the special case of $\mathbf{A} \in \{0,1\}^{m \times n}$ in the Appendix, using the theory of VC dimension.

Crucial to the proof of Theorem 1 is a recursively defined $2^k \times 2^k$ matrix $\mathbf{A}_k$, based on the Haar wavelet basis. We will also show that $\mathbf{A}_k$ is tight for Equation (2).

Theorem 3. *With $n = 2^k$ for an integer $k \geq 1$, hervecdisc($\mathbf{A}_k$) $\gtrsim \sqrt{\log n} \cdot \text{detlb}(\mathbf{A}_k)$.*

Its proof appears in Section 3.1. It is yet unknown whether Equation (1) is tight as Jiang and Reis improved upon Matoušek's bound by circumventing the inequality altogether. The resolution of this problem via an efficient algorithm would imply new and old constructive bounds, for example, the constructive version of Banaszczyk's upper bound for the Beck-Fiala problem [Ban98, BDG16], and a constructive version of Nikolov's upper bound for Tuśnady's problem [Nik17].

Our use of the matrix $\mathbf{A}_k$ is inspired by work of Kunisky [Kun23], who first used this matrix in the context of proving discrepancy lower bounds. We also observe that this matrix gives easier proofs of other known results in discrepancy theory: see Section 3.1.

Before we begin the proof proper, we define a few variants of discrepancy which will appear throughout this exposition. Just as disc($\mathbf{A}$) was defined in terms of L_∞, we can define discrepancy in terms of other norms. In particular, for L_1,

$$\text{disc}_1(\mathbf{A}) := \min_{\mathbf{x} \in \{\pm 1\}^n} \frac{\|\mathbf{A}\mathbf{x}\|_1}{m}$$

and generally for L_p,

$$\text{disc}_p(\mathbf{A}) := \min_{\mathbf{x} \in \{\pm 1\}^n} \left(\frac{\|\mathbf{A}\mathbf{x}\|_p^p}{m} \right)^{1/p}.$$

Note that $\text{disc}_p(\mathbf{A}) \leq \text{disc}_q(\mathbf{A})$ when $p \leq q$.

2   Proof Structure

In order to prove Theorem 1 we will find a family of matrices which satisfy equation (3). Our candidates will have the form $\mathbf{P}_N \otimes \mathbf{A}$ where $\mathbf{P}_N$ is the $2^N \times N$ incidence matrix of the power set, and $\mathbf{A}$ is some $p \times p$ matrix with a gap between detlb($\mathbf{A}$) and $\text{disc}_1(\mathbf{A})$. In particular, we let $\mathbf{A}$ be the Haar basis matrix used in the work of Kunisky [Kun23], and prove some properties of $\mathbf{A}$ in-order to obtain the present result.

We bound detlb($\mathbf{P}_N \otimes \mathbf{A}$) from above by showing that detlb($\mathbf{P}_N \otimes \mathbf{A}$) $\lesssim \sqrt{N} \cdot \text{detlb}(\mathbf{A})$ using standard linear algebra and Lemma 4 from [Mat13]. See Lemma 4. For our choice of $\mathbf{A}$, we will show that detlb($\mathbf{A}$) $\lesssim 1$.

We also bound $\text{disc}(\mathbf{P}_N \otimes \mathbf{A}) \gtrsim N \cdot \text{disc}_1(\mathbf{A})$ using a discrepancy amplification argument. See Lemma 5. By finding a tight lower bound on $\text{disc}_1(\mathbf{A})$, we obtain the lower bound $\text{disc}(\mathbf{P}_N \otimes \mathbf{A}) \gtrsim N \cdot \sqrt{\log p}$. Taken together, these bounds gives us a gap on the order of $\sqrt{N} \cdot \sqrt{\log p}$ between $\text{detlb}(\mathbf{P}_N \otimes \mathbf{A})$ and $\text{disc}(\mathbf{P}_N \otimes \mathbf{A})$.

Lemma 4. *For the power matrix $\mathbf{P}_N$, and any real matrix $\mathbf{A}$,*

$$\text{detlb}\,(\mathbf{P}_N \otimes \mathbf{A}) \leq \sqrt{eN} \cdot \text{detlb}(\mathbf{A}).$$

Proof. Let $\mathbf{u}_1, ..., \mathbf{u}_N$ be the columns of $\mathbf{P}_N$ and $\mathbf{A} \in \mathbb{R}^{p \times p}$. Divide the columns of $\mathbf{P}_N \otimes \mathbf{A}$ into N contiguous blocks of size $(2^N p) \times p$ each representing $\mathbf{u}_\ell \otimes \mathbf{A}$. Note that $\mathbf{u}_\ell \otimes \mathbf{A}$ consists of 2^N blocks of $\mathbf{A}$ or $\mathbf{0}$ stacked on top of one another. We claim that $\text{detlb}(\mathbf{u}_\ell \otimes \mathbf{A}) \leq \text{detlb}(\mathbf{A})$. Consider an $s \times s$ sub-matrix $\mathbf{B}$ of $\mathbf{u}_\ell \otimes \mathbf{A}$ with the rows indexed by I and columns indexed by J. Note that if any row of $\mathbf{B}$ is zero, then $\det(\mathbf{B}) = 0$ so in order for the determinant to be non-zero, the rows of $\mathbf{B}$ must be parts of rows of $\mathbf{A}$ with columns indexed by J. If there are multiple copies of the same row of $\mathbf{A}$, then again $\det(\mathbf{B}) = 0$. Thus $\mathbf{B}$ must come from distinct rows of $\mathbf{A}$ with columns indexed by J. It follows that $\mathbf{B}$ is actually a sub-matrix of $\mathbf{A}$ up to rearrangement of the rows, so $|\det(\mathbf{B})|^{1/s} \leq \text{detlb}(\mathbf{A})$. Since this is true for all choices of the submatrix $\mathbf{B}$, we have $\text{detlb}\,(\mathbf{u}_\ell \otimes \mathbf{A}) \leq \text{detlb}(\mathbf{A})$.

Recall from [Mat13] Lemma 4, that for real matrices $\mathbf{B}_1, ..., \mathbf{B}_t$ each with the same number of columns and $D := \max_{i=1,2,...t} \text{detlb}(\mathbf{B}_i)$, any matrix $\mathbf{B}$ whose rows are copies of the rows of the matrices $\mathbf{B}_i$ satisfies $\text{detlb}(\mathbf{B}) \leq D\sqrt{et}$. By applying Lemma 4 to $(\mathbf{P}_N \otimes \mathbf{A})^\top$ with $\mathbf{B}_i = (\mathbf{u}_i \otimes \mathbf{A})^\top$, and we have that

$$\text{detlb}(\mathbf{P}_N \otimes \mathbf{A}) \leq \sqrt{eN} \cdot \max_{\ell \in [N]} \text{detlb}(\mathbf{u}_\ell \otimes \mathbf{A}) \leq \sqrt{eN} \cdot \text{detlb}(\mathbf{A}).$$

□   □

Lemma 5. (Discrepancy Amplification). *For the power matrix $\mathbf{P}_N$ and any real matrix $\mathbf{A}$,*

$$\text{disc}(\mathbf{P}_N \otimes \mathbf{A}) \geq \frac{N \cdot \text{disc}_1(\mathbf{A})}{2}.$$

Proof. Let $\mathbf{A} \in \mathbb{R}^{p \times q}$ and $t := \text{disc}_1(\mathbf{A})$. Consider some vector $\mathbf{x} \in \{\pm 1\}^{qN}$ composed of vectors $\mathbf{x}^{(1)}, ..., \mathbf{x}^{(N)}$ stacked on top of each other containing p entries each. We compute $\|(\mathbf{P}_N \otimes \mathbf{A})\mathbf{x}\|_\infty$. Note that

$$\|(\mathbf{P}_N \otimes \mathbf{A})\mathbf{x}\|_\infty = \max_{S \subseteq [N]} \left\| \sum_{i \in S} \mathbf{A}\mathbf{x}^{(i)} \right\|_\infty = \max_{S \subseteq [N]} \max_{j \in [p]} \left| \sum_{i \in S} \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right|.$$

From the assumption, we have $\frac{1}{p} \|\mathbf{A}\mathbf{x}^{(i)}\|_1 \geq t$ for every $i \in [N]$. Taking an average over all choices of i,

$$t \leq \frac{1}{pN} \sum_{i=1}^{N} \sum_{j=1}^{p} \left| \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right| = \frac{1}{pN} \sum_{j=1}^{p} \sum_{i=1}^{N} \left| \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right| \implies \sum_{i=1}^{N} \left| \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right| \geq Nt$$

for some $j \in [p]$. With $S^+ = \{i : (\mathbf{A}\mathbf{x}^{(i)})_j > 0\}$ and $S^- = \{i : (\mathbf{A}\mathbf{x}^{(i)})_j < 0\}$,

$$
\begin{aligned}
Nt \leq \sum_{i=1}^{N} \left| \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right| &= \sum_{i \in S^+} \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j - \sum_{i \in S^-} \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \\
&= \left| \sum_{i \in S^+} \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right| + \left| \sum_{i \in S^-} \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right| \\
&\implies \max \left\{ \left| \sum_{i \in S^+} \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right|, \left| \sum_{i \in S^-} \left(\mathbf{A}\mathbf{x}^{(i)} \right)_j \right| \right\} \geq \frac{Nt}{2}.
\end{aligned}
$$

□

Lemmas 4 and 5 together imply that

$$\frac{\mathrm{herdisc}(\mathbf{P}_N \otimes \mathbf{A})}{\mathrm{detlb}(\mathbf{P}_N \otimes \mathbf{A})} \geq \frac{\sqrt{N}}{2\sqrt{e}} \cdot \frac{\mathrm{disc}_1(\mathbf{A})}{\mathrm{detlb}(\mathbf{A})}.$$

Note that, if $\mathbf{A} \in \mathbb{R}^{p \times q}$, then $\mathbf{P}_N \otimes \mathbf{A}$ is an $(2^N p) \times (Nq)$ matrix, so $\sqrt{N}$ is roughly $\sqrt{\log m}$ for small enough p, where $m := 2^N p$ is the number of rows of $\mathbf{P}_N \otimes \mathbf{A}$. To prove Theorem 1, we need to find a matrix $\mathbf{A}$ that exhibits a large gap between $\mathrm{disc}_1(\mathbf{A})$ and $\mathrm{detlb}(\mathbf{A})$. In the next section, we show that a matrix whose columns are the discrete Haar basis vectors has this property.

3 Discrete Haar Basis

The $2^k \times 2^k$ discrete Haar basis matrix $\mathbf{A}_k$ is defined recursively with $\mathbf{A}_0 = [1]$ and

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{A}_{k-1} & \mathbf{I}_{2^{k-1}} \\ \mathbf{A}_{k-1} & -\mathbf{I}_{2^{k-1}} \end{bmatrix}, \tag{4}$$

where $\mathbf{I}_{2^{k-1}}$ is the $2^{k-1} \times 2^{k-1}$ identity matrix. This matrix arises from the following tree structure. Construct a depth k perfect binary tree, and let r be an additional node. We make the root of the perfect binary tree the left child of r, and r becomes the root of our tree. Every non-leaf node represents a column in the matrix while every root-to-leaf path corresponds to a row in the matrix. Whenever the path proceeds down the left child from some node i, entry i of the corresponding row will have value $+1$. If instead the path proceeds down the right child of i, entry i of the corresponding row will have value -1. Thus every row will have k non-zero entries. It is also not hard to show that, for any ± 1 coloring of the columns, there is a row whose nonzero entries are equal to the corresponding column colors, and, therefore, $\mathrm{disc}(\mathbf{A}_k) = k$. Kunisky [Kun23] describes this in detail.

In addition, we define the $\{0,1\}^{2^k \times 2^k}$ matrices $\mathbf{A}_k^+$ and $\mathbf{A}_k^-$ to be the indicator matrices of the positive and negatives elements of $\mathbf{A}_k$ respectively. Here an *indicator matrix* will have one in some entry if and only the corresponding entry of $\mathbf{A}_k$ is non-zero and positive, in the case of $\mathbf{A}_k^+$, or negative, in the case of $\mathbf{A}_k^-$. Note that $\mathbf{A}_k = \mathbf{A}_k^+ - \mathbf{A}_k^-$. Finally define,

$$\mathbf{A}_k^{\pm} := \begin{pmatrix} \mathbf{A}_k^+ \\ \mathbf{A}_k^- \end{pmatrix}. \tag{5}$$

We bound the hereditary discrepancy to determinant lower bound ratio for both $\mathbf{P}_N \otimes \mathbf{A}_k$ and $\mathbf{P}_N \otimes \mathbf{A}_k^{\pm}$.

Theorem 6. *For the power matrix $\mathbf{P}_N$, the discrete Haar basis $\mathbf{A}_k$, and the stacked indicator matrix $\mathbf{A}_k^{\pm}$ as defined in equation (5),*

$$\frac{\mathrm{herdisc}\left(\mathbf{P}_N \otimes \mathbf{A}_k\right)}{\mathrm{detlb}\left(\mathbf{P}_N \otimes \mathbf{A}_k\right)} \gtrsim \sqrt{N \cdot k}, \tag{6}$$

$$\frac{\mathrm{herdisc}\left(\mathbf{P}_N \otimes \mathbf{A}_k^{\pm}\right)}{\mathrm{detlb}\left(\mathbf{P}_N \otimes \mathbf{A}_k^{\pm}\right)} \gtrsim \sqrt{N \cdot k}. \tag{7}$$

Proof. First we apply the proof structure described in the previous section to $\mathbf{A}_k$. In particular, we show that $\mathrm{detlb}(\mathbf{A}_k) = O(1)$ in Lemma 7 and that $\mathrm{disc}_1(\mathbf{A}_k) \gtrsim \sqrt{k}$ in Lemma 9. Applying Lemma 4 to the first result and Lemma 5 to the second, we have that $\mathrm{detlb}\left(\mathbf{P}_N \otimes \mathbf{A}_k\right) \lesssim \sqrt{N}$ and $\mathrm{disc}\left(\mathbf{P}_N \otimes \mathbf{A}_k\right) \gtrsim N \cdot \sqrt{k}$. It follows that

$$\frac{\mathrm{herdisc}\left(\mathbf{P}_N \otimes \mathbf{A}_k\right)}{\mathrm{detlb}\left(\mathbf{P}_N \otimes \mathbf{A}_k\right)} \geq \frac{\mathrm{disc}\left(\mathbf{P}_N \otimes \mathbf{A}_k\right)}{\mathrm{detlb}\left(\mathbf{P}_N \otimes \mathbf{A}_k\right)} \gtrsim \sqrt{N \cdot k}.$$

5

The process for $\mathbf{A}_k^{\pm}$ is similar. To show an upper bound on $\mathrm{detlb}(\mathbf{A}_k^{\pm})$, use Corollary 8 where $\mathbf{A}_k^{+}$ and $\mathbf{A}_k^{-}$ are shown to be TUM. Since the determinant of any square submatrix of either matrix is at most one in absolute value, we can apply Lemma 4 of [Mat13] to $\mathbf{A}_k^{+}$ and $\mathbf{A}_k^{-}$ to obtain $\mathrm{detlb}(\mathbf{A}_k^{\pm}) = O(1)$. To obtain the lower bound on $\mathrm{disc}_1(\mathbf{A}_k^{\pm})$, we will recall that $\mathrm{disc}_1(\mathbf{A}_k) \gtrsim \sqrt{k}$ from Lemma 9. Note that, for any $\mathbf{x} \in \{-1, +1\}^{2^k}$, by the triangle inequality

$$\frac{1}{2^k}\|\mathbf{A}_k\mathbf{x}\|_1 = \frac{1}{2^k}\|(\mathbf{A}_k^{+} - \mathbf{A}_k^{-})\mathbf{x}\|_1 \leq 2\left(\frac{1}{2^{k+1}}\|\mathbf{A}_k^{+}\mathbf{x}\|_1 + \frac{1}{2^{k+1}}\|\mathbf{A}_k^{-}\mathbf{x}\|_1\right).$$

Therefore, $\mathrm{disc}_1(\mathbf{A}_k^{\pm}) \gtrsim \sqrt{k}$ as well. Apply Lemma 4 and Lemma 5 to $\mathrm{detlb}(\mathbf{A}_k^{\pm}) = O(1)$ and $\mathrm{disc}_1(\mathbf{A}_k^{\pm 1}) \gtrsim \sqrt{k}$ respectively to obtain equation (7). $\qquad\square$

Lemma 7. $\mathrm{detlb}(\mathbf{A}_k) \leq 2$.

Proof. We show that any $i \times i$ square submatrix $\mathbf{B}$ of $\mathbf{A}_k$ satisfies $|\det(\mathbf{B})| \leq 2^i$. First, define $M_k(i) := \max_{\mathbf{B}}|\det(\mathbf{B})|$ where the maximum is taken over all $i \times i$ submatrices $\mathbf{B}$ of $\mathbf{A}_k$. We compute $M_k(i)$ recursively by considering the forms that all $i \times i$ submatrices of $\mathbf{A}_k$ can take:

1. $\mathbf{B}$ only contain elements from the first 2^{k-1} columns of $\mathbf{A}_k$,

2. $\mathbf{B}$ only contain elements from the second 2^{k-1} columns of $\mathbf{A}_k$, or

3. $\mathbf{B}$ contain elements from both the first and second 2^{k-1} columns of $\mathbf{A}_k$.

We use the recursive formula (4) to analyze these cases. In the first case the resulting submatrix is either entirely contained in $\mathbf{A}_{k-1}$ up to rearranging rows, or contains a duplicated row. The magnitude of the determinant of these submatrices can be bounded above by $M_{k-1}(i)$ and 0, respectively. In the second case we note that the submatrix is TUM. To see this, recall that the second 2^{k-1} columns of $\mathbf{A}_k$ consist of an identity matrix and its negation stacked on top of one another. Any square submatrix is entirely contained in the identity matrix or contains duplicated (and negated) rows. Thus the absolute value of the determinant of this kind of submatrix is at most one. It remains to consider the third case. Let $\mathbf{B}$ be a submatrix of $\mathbf{A}_k$ with some j columns coming from the second 2^{k-1} columns of $\mathbf{A}_k$ for $1 \leq j < i$. For any such column there is either one or two non-zero entries, equal to 1 or -1. If there is only one non-zero entry, then this reduces to computing $M_k(i-1)$ since we can perform a co-factor expansion on this column. If there are two non-zero entries, then we can permute the rows so that they are adjacent. This only changes the sign of the resulting determinant. Notice that the two rows are identical except for the sign of the non-zero entries. When performing a co-factor expansion on these two entries, the $(i-1) \times (i-1)$ submatrix that results when removing either row and the column is identical. Thus $|\det(\mathbf{B})|$ is at most twice the absolute value of the determinant of this $(i-1) \times (i-1)$ submatrix. After removing all the columns and associated rows of $\mathbf{B}$ from the second half of $\mathbf{A}_k$ in this way, we see that $|\det(\mathbf{B})| \leq 2^j M_{k-1}(i-j)$. Since, in the base case, $M_0(1) = 1$, the claim follows. $\qquad\square$

Using a similar argument as above, we can show that the matrices $\mathbf{A}_k^{+}$ and $\mathbf{A}_k^{-}$ are TUM. Note that, using (4), $\mathbf{A}_k^{+}$ and $\mathbf{A}_k^{-}$ can be recursively defined as $\mathbf{A}_0^{+} = [1]$, $\mathbf{A}_0^{-} = [0]$, and

$$\mathbf{A}_k^{+} = \begin{bmatrix} \mathbf{A}_{k-1}^{+} & \mathbf{I}_{2^{k-1}} \\ \mathbf{A}_{k-1}^{+} & \mathbf{0} \end{bmatrix} \qquad \mathbf{A}_k^{-} = \begin{bmatrix} \mathbf{A}_{k-1}^{-} & \mathbf{0} \\ \mathbf{A}_{k-1}^{-} & \mathbf{I}_{2^{k-1}} \end{bmatrix}, \qquad (8)$$

where $\mathbf{0}$ is the all zeros matrix of appropriate dimension, and $\mathbf{I}_{2^{k-1}}$ is the 2^{k-1} by 2^{k-1} identity matrix.

Corollary 8. $\mathbf{A}_k^{+}$ and $\mathbf{A}_k^{-}$ are TUM matrices where $\mathbf{A}_k^{+}$ and $\mathbf{A}_k^{-}$ are indicators of the positive and negatives entries of $\mathbf{A}_k$, respectively.

Proof. We only consider $\mathbf{A}_k^+$ as the proof that $\mathbf{A}_k^-$ is a TUM matrix is similar. The proof proceeds by induction on k. Consider some $i \times i$ submatrix $\mathbf{B}$ of $\mathbf{A}_k^+$. If $\mathbf{B}$ is entirely contained in first half of the columns of $\mathbf{A}_k^+$, then we are done by the inductive hypothesis; if $\mathbf{B}$ is entirely contained in the second half of the columns of $\mathbf{A}_k^+$, then $\mathbf{B}$ is a submatrix of the identity matrix or has a row of 0's, and the absolute value of its determinant is at most 1. Thus it suffices to consider the case where $\mathbf{B}$ has some columns from the first half of $\mathbf{A}_k^+$ and some columns from the second half of $\mathbf{A}_k^+$. Since any column from the second half has only one non-zero entry, equal to 1, performing co-factor expansions on the columns in the second half shows that the absolute value of the determinant will only be as large as the absolute value of the determinant of some smaller square sub-matrix in $\mathbf{A}_{k-1}^+$. Note that in the base case, $|\det(\mathbf{A}_0^+)| = 1$. $\qquad\square$

Lemma 9. $\mathrm{disc}_1(\mathbf{A}_k) = \frac{k+1}{2^k}\binom{k}{\lfloor (k+1)/2 \rfloor} \cong \sqrt{k}$.

Proof. The proof for $k = 0$ is trivial, so we focus on the case $k \geq 1$. Let $\tilde{\mathbf{A}}_k$ denote the $2^k \times (2^k - 1)$ matrix equal to $\mathbf{A}_k$ with the first column, all of whose entries are 1, removed. Note that this is equivalent to removing the root node r and keeping only the perfect binary tree of depth k in the tree structure of the Haar basis, as described at the beginning of the section. We have the following key claim.

Claim 10. *For any* $\mathbf{x} \in \{\pm 1\}^{2^k-1}$ *there exists a permutation which maps the entries of* $\tilde{\mathbf{A}}_k \mathbf{x}$ *to those of* $\tilde{\mathbf{A}}_k \mathbf{1}$.

Proof. Our proof is by induction on k. When $k = 1$, we have a root node with two children corresponding to the matrix

$$\tilde{\mathbf{A}}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

When $\mathbf{x} = [1]$, $\tilde{\mathbf{A}}_1 \mathbf{x} = \tilde{\mathbf{A}}_1 \mathbf{1}$ and the identity permutation suffices; when $\mathbf{x} = [-1]$, $\tilde{\mathbf{A}}_1 \mathbf{x} = -\tilde{\mathbf{A}}_1 \mathbf{1}$ and it suffices to swap the two entries.

Consider some height k perfect binary tree corresponding to $\tilde{\mathbf{A}}_k$. Let u be the root of the tree with left and right children u_+ and u_- respectively. Since every root-to-leaf path must go through u_+ or u_-, this forms a partition of the rows of $\tilde{\mathbf{A}}_k$. In particular, we can rearrange $\tilde{\mathbf{A}}_k$ as

$$\tilde{\mathbf{A}}_k = \begin{bmatrix} \mathbf{1} & \tilde{\mathbf{A}}_{k-1} & \mathbf{0} \\ -\mathbf{1} & \mathbf{0} & \tilde{\mathbf{A}}_{k-1} \end{bmatrix}.$$

Consider the $\tilde{\mathbf{A}}_{k-1}$ submatrices which appear in $\tilde{\mathbf{A}}_k$. The $\tilde{\mathbf{A}}_{k-1}$ submatrix in the first 2^{k-1} rows has rows which correspond to root-to-leaf paths with leaves in the subtree rooted at u_+. Its columns correspond to nodes in the same subtree. The $\tilde{\mathbf{A}}_{k-1}$ submatrix in the second 2^{k-1} rows is defined similarly on the subtree rooted at u_-. Write the vector $\mathbf{x}$ as $[x_u, \mathbf{x}_+, \mathbf{x}_-]^\top$ where x_u is the color of the node u and $\mathbf{x}_+$ and $\mathbf{x}_-$ are the colors of the nodes in the subtrees rooted at u_+ and u_-, respectively. Consider the value of x_u. If $x_u = 1$, then by the inductive hypothesis, there exists a permutation which takes the entries of $\tilde{\mathbf{A}}_{k-1}\mathbf{x}_+$ to the entries of $\tilde{\mathbf{A}}_{k-1}\mathbf{1}$ and another permutation which takes the entries of $\tilde{\mathbf{A}}_{k-1}\mathbf{x}_-$ to the entries of $\tilde{\mathbf{A}}_{k-1}\mathbf{1}$. These two permutations can be combined to form a permutation which maps the entries of $\tilde{\mathbf{A}}_k\mathbf{x}$ to the entries of $\tilde{\mathbf{A}}_k\mathbf{1}$. Otherwise $x_u = -1$. Again, there exists a permutation π_1 which takes $\tilde{\mathbf{A}}_{k-1}\mathbf{x}_+$ to $\tilde{\mathbf{A}}_{k-1}\mathbf{1}$ and another permutation π_2 which takes $\tilde{\mathbf{A}}_{k-1}\mathbf{x}_-$ to $\tilde{\mathbf{A}}_{k-1}\mathbf{1}$. We can construct a permutation which maps the elements of $\tilde{\mathbf{A}}_k\mathbf{x}$ to those of $\tilde{\mathbf{A}}_k\mathbf{1}$ by by first applying π_1 to the first 2^{k-1} entries of $\tilde{\mathbf{A}}_k\mathbf{x}$, and π_2 to the remaining 2^{k-1} entries, and then swapping the first 2^{k-1} entries with the second 2^{k-1} entries. $\qquad\square$

Note that, for any $\mathbf{x} \in \{\pm 1\}^{2^k}$, $\mathrm{disc}_1(\mathbf{A}_k, \mathbf{x}) = \mathrm{disc}_1(\mathbf{A}_k, -\mathbf{x})$. Then, we have

$$\mathrm{disc}_1(\mathbf{A}_k, \mathbf{x}) = \frac{1}{2}(\mathrm{disc}_1(\mathbf{A}_k, \mathbf{x}) + \mathrm{disc}_1(\mathbf{A}_k, -\mathbf{x}))$$

7

By Claim 10, and, since all entries of the first column of $\mathbf{A}_k$ are equal to 1,

$$\mathrm{disc}_1(\mathbf{A}_k, \mathbf{x}) = \frac{1}{2^k} \sum_{i=1}^{2^k} \frac{|\tilde{\mathbf{a}}_i^\top \mathbf{1} + 1| + |\tilde{\mathbf{a}}_i^\top \mathbf{1} - 1|}{2}$$

$$= \frac{1}{2^{k+1}} \sum_{i=1}^{2^k} |\tilde{\mathbf{a}}_i^\top \mathbf{1} + 1| + \frac{1}{2^{k+1}} \sum_{i=1}^{2^k} |\tilde{\mathbf{a}}_i^\top \mathbf{1} - 1|,$$

where $\tilde{\mathbf{a}}_i^\top$ is the i-th row of $\tilde{\mathbf{A}}_k$. Recall that each row of $\tilde{\mathbf{A}}_k$ has exactly k non-zero entries, and every sign pattern for these k entries appears exactly once, as each row in $\tilde{\mathbf{A}}_k$ corresponds to a root-to-leaf path in a depth k perfect binary tree, and the sign pattern corresponds to the sequence of left and right turns made by the path. In particular, there are exactly $\binom{k}{\ell}$ rows $\tilde{\mathbf{a}}_i$ for which $\tilde{\mathbf{a}}_i^\top \mathbf{1}$ equals $k - 2\ell$, since such rows have $k - \ell$ non-zero entries equal to $+1$, and ℓ non-zero entries equal to -1. Substituting above, we have

$$\frac{1}{2^{k+1}} \sum_{i=1}^{2^k} |\tilde{\mathbf{a}}_i^\top \mathbf{1} + 1| + \frac{1}{2^{k+1}} \sum_{i=1}^{2^k} |\tilde{\mathbf{a}}_i^\top \mathbf{1} - 1|$$

$$= \frac{1}{2^{k+1}} \sum_{\ell=0}^{k} \binom{k}{\ell} |k + 1 - 2\ell| + \frac{1}{2^{k+1}} \sum_{\ell=0}^{k} \binom{k}{\ell} |k - 1 - 2\ell|$$

$$= \frac{1}{2^{k+1}} \sum_{\ell=0}^{k} \binom{k}{\ell} |k + 1 - 2\ell| + \frac{1}{2^{k+1}} \sum_{\ell=1}^{k+1} \binom{k}{\ell - 1} |k + 1 - 2\ell|$$

$$= \frac{1}{2^{k+1}} \sum_{\ell=0}^{k+1} \left(\binom{k}{\ell} + \binom{k}{\ell - 1} \right) |k + 1 - 2\ell|$$

$$= \frac{1}{2^{k+1}} \sum_{\ell=0}^{k+1} \binom{k+1}{\ell} |k + 1 - 2\ell|.$$

The second equality above follows by a change of variables in the second sum, the third equality uses the convention $\binom{k}{k+1} = \binom{k}{-1} = 0$, and the last equality follows form Pascal's identity.

Now, by Lemma 11 below, we have that

$$\sum_{\ell=0}^{k+1} \binom{k+1}{\ell} |k + 1 - 2\ell| = 2(k+1) \binom{k}{\lfloor (k+1)/2 \rfloor},$$

which implies that $\mathrm{disc}_1(\tilde{\mathbf{A}}_k) \cong \sqrt{k}$ since $\binom{k}{\lfloor (k+1)/2 \rfloor} \cong 2^k/\sqrt{k}$ by Stirling's approximation. $\qquad \square$

The above proof also has a probabilistic interpretation. We show that, for a uniformly random row $\mathbf{a}^\top$ of $\tilde{\mathbf{A}}_k$ and a fixed coloring $\mathbf{x} \in \{\pm 1\}^{2^k - 1}$, $\mathbf{a}^\top \mathbf{x}$ is distributed like $X_1 + \cdots + X_k$ where the X_is are independent Rademacher random variables (i.e. random variables uniform in $\{-1, +1\}$). Recall that uniformly choosing a row of $\tilde{\mathbf{A}}_k$ corresponds to uniformly choosing a root-to-leaf path in the depth k perfect binary tree. Further, the non-leaf nodes of the tree correspond to the columns of $\tilde{\mathbf{A}}_k$. Thus we know that exactly k entries of the row will be non-zero. Let the indices of these entries be $U_1, ..., U_k$, and note that $\mathbf{a}^\top \mathbf{x} = x_{U_1} a_{U_1} + \cdots + x_{U_k} a_{U_k}$. The key observation is that, conditional on the values of $U_1, \ldots, U_\ell$, a_{U_ℓ} is equally likely to be -1 or $+1$, since the a uniformly random path in the binary tree going through $U_1, \ldots, U_\ell$ is equally likely to visit the left or the right child of U_ℓ. We can then show that $x_{U_1} a_{U_1} + \cdots + x_{U_k} a_{U_k}$ has the same distribution as a sum of k independent Rademacher random variables by induction on k. In the base case, a_{U_1} is uniform in $\{-1, +1\}$, and so is $x_{U_1} a_{U_1}$. Suppose $x_{U_1} a_{U_1} + \cdots + x_{U_{k-1}} a_{U_{k-1}}$ is distributed as the sum of $k - 1$ independent Rademacher random variables. Conditional on the choice of $U_1, \ldots, U_k$, a_{U_k}, and, therefore, $x_{U_k} a_{U_k}$ are equally likely to be -1 or $+1$. Taking expectation over the choice of $U_1, \ldots, U_k$ finishes the proof.

The next lemma is likely a well-known calculation. We include a proof due to Lavrov, for completeness.

Lemma 11 ([Lav18]). $\sum_{\ell=0}^{k} \binom{k}{\ell}|k - 2\ell| = 2k \cdot \binom{k-1}{\lfloor k/2 \rfloor}$.

Proof. Recall the identity $\binom{k}{\ell}\ell = \binom{k-1}{\ell-1}k$. We write

$$\sum_{\ell=0}^{k} \binom{k}{\ell}|k - 2\ell| = \sum_{\ell<k/2} \binom{k}{\ell}(k - 2\ell) - \sum_{\ell>k/2} \binom{k}{\ell}(k - 2\ell)$$

$$= k\left(\sum_{\ell<k/2} \binom{k}{\ell} - \sum_{\ell>k/2} \binom{k}{\ell}\right) - 2\left(\sum_{\ell<k/2} \binom{k}{\ell}\ell - \sum_{\ell>k/2} \binom{k}{\ell}\ell\right)$$

$$= 2k\left(\sum_{\ell>k/2} \binom{k-1}{\ell-1} - \sum_{\ell<k/2} \binom{k-1}{\ell-1}\right)$$

$$= 2k\binom{k-1}{\lfloor k/2 \rfloor}.$$

Here, the first equality follows since $\binom{k}{\ell}(k - 2(k/2)) = 0$ when k is even. The last equality follows by consider the parity of k; when k is even, we obtain a $\binom{k-1}{k/2}$ term after cancellation, and when k is odd, we obtain a $\binom{k-1}{\lfloor k/2 \rfloor}$ term after cancellation. $\square$

Note that when we divide the identity by 2^k, we obtain the expectation of a sum of k independent Rademacher random variables. The asymptotic version of this identity follows from Khintchine's inequality.

Proof of Theorem 1. Consider the range of m in $[n, n^2]$ and $\left(n^2, 2^{n^{(1-\epsilon)}}\right]$ separately. In the first interval, we let $\mathbf{A}$ be the matrix $\mathbf{A}_k$ padded with $m - n$ rows of zeros. Here, $\frac{\text{herdisc}(\mathbf{A})}{\text{detlb}(\mathbf{A})} \cong \log n \cong \sqrt{\log m \cdot \log n}$. When $m \in \left(n^2, 2^{n^{(1-\epsilon)}}\right]$, we consider the matrix $\mathbf{P}_N \otimes \mathbf{A}_k$ where $N = \lfloor \log_2(m/n) \rfloor$ and $k = \lfloor \log_2 n^\epsilon \rfloor$. Observe that $\mathbf{P}_N \otimes \mathbf{A}_k$ is an $m' \times n'$ matrix where $m' = 2^{N+k} \leq m/n^{1-\epsilon} < m$ and $n' = N \cdot 2^k \leq \log_2(m/n) \cdot n^\epsilon \leq n - n^\epsilon \log n$ since $m \leq 2^{n^{(1-\epsilon)}}$. We obtain $\mathbf{A}$ by padding $\mathbf{P}_N \otimes \mathbf{A}_k$ with zero vectors so that it has exactly m rows and n columns. Note that $\log m \cong N$ and $\log n \cong k$. By Theorem 6, $\frac{\text{herdisc}(\mathbf{A})}{\text{detlb}(\mathbf{A})} \gtrsim \sqrt{Nk} \gtrsim \sqrt{\log m \log n}$, as required.

The reader might object that the matrix $\mathbf{A}_k$ has negative entries which would not occur for incidence matrices of a set system, but we can remedy this by considering $\mathbf{A}_k^{\pm}$ as defined in Theorem 6 instead. $\square$

3.1 Other Notable Properties

We can make a few more observations about the properties of $\mathbf{A}_k$.

Claim 12. $|\det(\mathbf{A}_k)| = 2^{2^k - 1}$.

Proof. Note that the columns of $\mathbf{A}_k$ are orthogonal so $|\det(\mathbf{A}_k)|$ is equal to the product of the ℓ_2-norms of columns. Since the ith column has magnitude $2^{2^{i-1}}$, $|\det(\mathbf{A}_k)| = 2^{2^{k-1}+2^{k-2}+\cdots+2^0} = 2^{2^k - 1}$. $\square$

Next we prove Theorem 3, showing that $\mathbf{A}_k$ serves as an example that equation (2) of [Mat13] — mentioned in the introduction — is tight.

Proof of Theorem 3. To see this, it suffices to show that vecdisc $(\mathbf{A}_k)^2 = \Omega(k)$. Let $\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_q$ be the vector colors assigned to the 2^k columns of $\mathbf{A}_k$. Recall that $\mathbf{A}_k$ corresponds to a tree with root node r, where r has no right child, and the left child is the root of a perfect binary tree of depth k. The root to leaf paths of this tree represent rows in $\mathbf{A}_k$. For any path $r, t_1, ..., t_i$ from r to a node t_i, let $\overline{\mathbf{v}}_{t_i} = \mathbf{v}_r + \sum_{j=1}^{i-1} a_{t_j}\mathbf{v}_{t_j}$ where a_{t_j} is 1 if t_{j+1} is the left child of t_j, and -1 otherwise. We will show that there exists a root-to-leaf

t_k path $t_1, ..., t_k$ such that $\|\overline{\mathbf{v}}_{t_k}\|_2^2 \geq k$. In particular we show that at every internal node t, with children t_+ and t_-, must have $\|\overline{\mathbf{v}}_{t_+}\|^2 \geq 1 + \|\overline{\mathbf{v}}_t\|^2$ or $\|\overline{\mathbf{v}}_{t_-}\|^2 \geq 1 + \|\overline{\mathbf{v}}_t\|^2$. To see this, note that

$$\|\overline{\mathbf{v}}_{t_+}\|^2 = \|\overline{\mathbf{v}}_t + \mathbf{v}_t\|^2 = \|\overline{\mathbf{v}}_t\|^2 + \|\mathbf{v}_t\|^2 + 2\langle\overline{\mathbf{v}}_t, \mathbf{v}_t\rangle = \|\overline{\mathbf{v}}_t\|^2 + 1 + 2\langle\overline{\mathbf{v}}_t, \mathbf{v}_t\rangle.$$

Similarly, we have that $\|\overline{\mathbf{v}}_{t_-}\|^2 = \|\overline{\mathbf{v}}_t\|^2 + 1 - 2\langle\overline{\mathbf{v}}_t, \mathbf{v}_t\rangle$. The claim then follows since either $\langle\overline{\mathbf{v}}_t, \mathbf{v}_t\rangle \geq 0$ or $-\langle\overline{\mathbf{v}}_t, \mathbf{v}_t\rangle \geq 0$. The theorem then follows from the tree interpretation of $\mathbf{A}_k$. □

In [LSV86] there appears an open problem of Sós which asks if the hereditary discrepancy of a union of two sets systems is bounded above by the discrepancy of each individual set system i.e. for set systems $(X, \mathcal{S}_1)$ and $(X, \mathcal{S}_2)$ is it true that $\mathrm{herdisc}(\mathcal{S}_1 \cup \mathcal{S}_2) \leq f(\mathrm{herdisc}(\mathcal{S}_1), \mathrm{herdisc}(\mathcal{S}_2))$ for some function f? It turns out that no such bound exists. The Hoffman example[2], the example of Pálvölgyi [Pál10], and the three permutations family of Newman, Neiman, and Nikolov [NNN12] are instances of such pairs of set systems whose individual hereditary discrepancies are at most 1, but whose union on a universe of size n has discrepancy $\Omega(\log n / \log \log n)$ (for the Hoffman example) or $\Omega(\log n)$ (for the other two).

We see that $\mathbf{A}_k$ — with the decomposition into $\mathbf{A}_k^+$ and $\mathbf{A}_k^-$ — is another similar counter-example of Sós' conjecture. While it matches the $\Omega(\log n)$ discrepancy lower bound of the Pálvölgyi and Newman-Neiman-Nikolov constructions, it is simpler to analyze.

Claim 13. *With $\mathbf{A}_k^\pm$ as described above Claim 8,*

$$\mathrm{disc}\left(\mathbf{A}_k^\pm\right) \gtrsim k.$$

Proof. Recall that $\mathrm{disc}(\mathbf{A}_k) = k$. We claim that $\mathrm{disc}(\mathbf{A}_k^\pm) \geq \frac{1}{2}\mathrm{disc}(\mathbf{A}_k)$, and this proves the claim. Indeed, take any coloring $\mathbf{x}$. Let $\mathbf{a}^\top$ be the row of $\mathbf{A}_k$ achieving $|\mathbf{a}^\top\mathbf{x}| = \mathrm{disc}(\mathbf{A}_k, \mathbf{x})$ and let $\mathbf{a}_+^\top$ and $\mathbf{a}_-^\top$ be the corresponding rows in the copy of $\mathbf{A}^+$ and $\mathbf{A}^-$ in $\mathbf{A}^\pm$ respectively. Since $\mathrm{disc}(\mathbf{A}_k) \leq |\mathbf{a}^\top\mathbf{x}| = |\mathbf{a}_+^\top\mathbf{x} - \mathbf{a}_-^\top\mathbf{x}|$, by the triangle inequality we have that either $|\mathbf{a}_+^\top\mathbf{x}| \geq \mathrm{disc}(\mathbf{A}_k)/2$ or $|\mathbf{a}_-^\top\mathbf{x}| \geq \mathrm{disc}(\mathbf{A}_k)/2$. □

4 Upper Bound on Hereditary Discrepancy

In this section we prove Theorem 2. To this end we introduce the volume lower bound on hereditary discrepancy, introduced by Lovász and Vesztergombi [LV89], and, in a more general setting, by Banaszczyk [Ban93], and studied by Dadush, Nikolov, Talwar, and Tomczak-Jaegermann [DNTTJ18].

Let $\mathbf{A}$ be an $m \times n$ real matrix, and define the symmetric convex set $K_\mathbf{A} := \{x \in \mathbb{R}^n : \|\mathbf{A}x\|_\infty \leq 1\}$. Let us define the *volume lower bound of A*, denoted $\mathrm{volLB}(A)$, by

$$\mathrm{volLB}(A) = \max_{k \in [n]} \max_{S \subseteq [n], |S| = k} \frac{1}{\mathrm{vol}_k\left(K_\mathbf{A} \cap W_S\right)^{1/k}},$$

where W_S is the canonical subspace in the dimensions indexed by S (i.e. $W_S = \mathrm{span}\{\mathbf{e}_i, i \in S\}$) and vol_k is the k-dimensional volume within W_S, i.e., the Lebesgue measure restricted to this subspace. We also define a dual volume lower bound by

$$\mathrm{volLB}^*(A) = \max_{k \in [n]} \max_{S \subseteq [n], |S| = k} \frac{\mathrm{vol}_k\left(\mathrm{conv}(\pm\mathbf{\Pi}_S\mathbf{a}_1, \ldots \pm \mathbf{\Pi}_S\mathbf{a}_m)\right)^{1/k}}{c_k^{2/k}},$$

where $\mathbf{\Pi}_S$ is the orthogonal projection onto W_S, $\mathbf{a_i}^\top$ is the i-th row of $\mathbf{A}$, and $c_k = \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)}$ is the volume of the k-dimensional unit Euclidean ball.

---

[2] Hoffman's set system $\mathcal{F}$ is defined on a regular k-ary tree of depth k and obtains $\frac{\mathrm{herdisc}(\mathbf{A}_\mathcal{F})}{\mathrm{detlb}(\mathbf{A}_\mathcal{F})} = \Theta\left(\frac{\log n}{\log \log n}\right)$. Let T be a k-regular tree with height k. The universe consists of the nodes of T. Let $\mathcal{F}_1$ be the sets of all root-to-leaf paths in T, and $\mathcal{F}_2$ be the set of all sibling sets (all nodes with the same parent) of internal nodes in T. Then $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$. We have that $\mathrm{herdisc}(\mathcal{F}_1), \mathrm{herdisc}(\mathcal{F}_2) \leq 1$, $\mathrm{detlb}(\mathcal{F}) = O(1)$, and $\mathrm{disc}(\mathcal{F}) = \Omega(k/\log k)$. See [Mat09] Section 4.4.

We also need the concept of a polar set of a set $K \subseteq \mathbb{R}^n$, defined as

$$K^\circ := \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y}^\top \mathbf{x} \le 1 \ \forall \mathbf{x} \in K\}.$$

It is a consequence of the hyperplane separator theorem that for any closed convex K containing 0, $K^{\circ\circ} = K$ [Roc70, Section 14]. Moreover, it is clear from the definition that $K \subseteq L$ implies $L^\circ \subseteq K^\circ$.

We have the following relationship between volLB($\mathbf{A}$) and volLB*($\mathbf{A}$).

Claim 14. *For any matrix* $\mathbf{A} \in \mathbb{R}^{m \times n}$, volLB($A$) $\cong$ volLB*(A).

Proof. Let $K_\mathbf{A}$ be defined as above, and let $L_\mathbf{A} := \operatorname{conv}(\pm \mathbf{a}_1, \ldots \pm \mathbf{a}_m)$. We claim that, for any set $S \subseteq [n]$, $(K_\mathbf{A} \cap W_S)^\circ = \mathbf{\Pi}_S L_\mathbf{A}$, where the polar $(K_\mathbf{A} \cap W_S)^\circ$ is taken within the subspace W_S. It is sufficient to show this for $S = [n]$, as, otherwise, we can replace the matrix $\mathbf{A}$ by its submatrix consisting of the columns indexed by S. In the case $S = [n]$, we just need to show $K_\mathbf{A}^\circ = L_\mathbf{A}$. Notice that

$$\begin{aligned}
K_\mathbf{A} &= \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{A}\mathbf{x}\|_\infty \le 1\} \\
&= \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle \le 1 \text{ for all } \|\mathbf{y}\|_1 \le 1\} \\
&= \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{A}^\top \mathbf{y} \rangle \le 1 \text{ for all } \|\mathbf{y}\|_1 \le 1\}.
\end{aligned}$$

By the definition of polar, we see that $K_\mathbf{A} = L_\mathbf{A}^\circ$ as

$$L_\mathbf{A} = \{\mathbf{A}^\top \mathbf{y} : \mathbf{y} \in \mathbb{R}^m \text{ where } \|\mathbf{y}\|_1 \le 1\}.$$

Thus $K_\mathbf{A}^\circ = L_\mathbf{A}^{\circ\circ} = L_\mathbf{A}$ as required.

Once we have established that $(K_\mathbf{A} \cap W_S)^\circ = \mathbf{\Pi}_S L_\mathbf{A}$, we have, by the Santaló-Blaschke and the reverse Santaló inequalities (see Chapters 1 and 8 of [AAGM15]),

$$\operatorname{vol}_k(K_\mathbf{A} \cap W_S)^{1/k} \operatorname{vol}_k((K_\mathbf{A} \cap W_S)^\circ)^{1/k} \cong c_k^{2/k}.$$

This completes the proof. $\square$

The next lemma shows a relationship between volLB($\mathbf{A}$) and detlb($\mathbf{A}$) that, as far as we are aware, has not been observed before.

Lemma 15. *For any matrix* $\mathbf{A} \in \mathbb{R}^{m \times n}$, volLB*($\mathbf{A}$) $\lesssim \sqrt{n} \cdot$ detlb($\mathbf{A}$).

In the proof of Lemma 15 we use the following result of Nikolov. Closely related results were shown earlier by Dvoretzky and Rogers [DR50, Theorem 5B] and Ball [Bal89, Proposition 7].

Lemma 16. (Theorem 10 in [Nik15]). *Let* $m \ge n$ *and* $E \subseteq \mathbb{R}^n$ *be a minimum volume ellipsoid containing the points* $\pm \mathbf{a}_1, \ldots, \pm \mathbf{a}_m \in \mathbb{R}^n$. *There exists a set* $T \subseteq [m]$ *of size* n *such that*

$$|\det((\mathbf{a}_i)_{i \in T})| \ge \sqrt{\frac{n!}{n^n}} \frac{\operatorname{vol}_n(E)}{c_n} \cong n^{1/4} e^{-n/2} \frac{\operatorname{vol}_n(E)}{c_n},$$

where $(\mathbf{a}_i)_{i \in T}$ *is the matrix with columns* $\mathbf{a}_i$ *for* $i \in T$, *and* vol_n *is the* n-*dimensional Lebesgue measure.*

Note that Theorem 10 in [Nik15] in fact shows that there is a distribution on random multisets T for which $\mathbb{E}|\det((\mathbf{a}_i)_{i \in T})|^2 = \frac{n!}{n^n} \frac{\operatorname{vol}_n(E)^2}{c_n^2}$. Since the determinant is zero unless T is a set, this implies Lemma 16.

Proof of Lemma 15. Take some $S \subseteq [n]$ of size k such that

$$\text{volLB}^*(A) = \frac{\operatorname{vol}_k(\mathbf{\Pi}_S L_\mathbf{A})^{1/k}}{c_k^{2/k}},$$

11

where $\mathbf{a}_1^\top, \ldots, \mathbf{a}_m^\top$ are the rows of $\mathbf{A}$, and $L_\mathbf{A} := \mathrm{conv}(\pm\mathbf{a}_1, \ldots \pm\mathbf{a}_m)$. Applying Lemma 16 to $\pm\mathbf{\Pi}_S\mathbf{a}_1, \ldots, \pm\mathbf{\Pi}_S\mathbf{a}_m$, we have that, taking $E \subseteq W_S$ to be the smallest volume ellipsoid containing $\pm\mathbf{\Pi}_S\mathbf{a}_1, \ldots, \pm\mathbf{\Pi}_S\mathbf{a}_m$, there exists a set $T \subseteq [m]$ of size k for which

$$|\det((\mathbf{\Pi}_S\mathbf{a}_i)_{i\in T})| \gtrsim k^{1/4}e^{-k/2}\frac{\mathrm{vol}_k(E)}{c_k} \geq k^{1/4}e^{-k/2}\frac{\mathrm{vol}_k(\mathbf{\Pi}_S L_\mathbf{A})}{c_k}.$$

The last inequality follows because $L_\mathbf{A} \subseteq E$. Re-arranging and raising to the power $1/k$, this gives us that

$$\mathrm{volLB}^*(A) = \frac{\mathrm{vol}_k(\mathbf{\Pi}_S L_\mathbf{A})^{1/k}}{c_k^{2/k}} \lesssim \frac{|\det((\mathbf{\Pi}_S\mathbf{a}_i)_{i\in T})|^{1/k}}{c_k^{1/k}} \lesssim \sqrt{k} \cdot \mathrm{detlb}(A),$$

where, in the final inequality, we used that $(\mathbf{\Pi}_S\mathbf{a}_i)_{i\in T}$ is the transpose of a k by k submatrix of $\mathbf{A}$, and we also used the estimate $c_k^{-1/k} \lesssim \sqrt{k}$, which follows from Stirling's approximation. Since $k \leq n$, the result follows. $\qquad\square$

We remark in passing that the trivial inequality $\mathrm{vol}_k(E) \geq \mathrm{vol}_k(\mathbf{\Pi}_S L_\mathbf{A})$ for a k-dimensional symmetric convex polytope with $2m$ vertices $\mathbf{\Pi}_S L_\mathbf{A}$ and an ellipsoid E containing it can be improved to $\mathrm{vol}_k(E) \geq \sqrt{\frac{k}{\log(2m)}}\mathrm{vol}_k(\mathbf{\Pi}_S L_\mathbf{A})$ when m is small, using, e.g., results of Gluskin [Glu89]. Substituting this inequality in the proof of Lemma 15 gives the bound $\mathrm{volLB}^*(\mathbf{A}) \lesssim \sqrt{\log 2m} \cdot \mathrm{detlb}(\mathbf{A})$.

The final ingredient we need for the proof of Theorem 2 is an upper bound on the hereditary discrepancy of partial colorings in terms of the volume lower bound, due to Dadush, Nikolov, Tomczak-Jaegermann, and Talwar.

Lemma 17. (Lem.8 in [DNTTJ18]). *There exist universal constants $c \geq 1$ and $\epsilon_0 \in (0, 1)$ such that the following holds. For any closed convex set $K \subseteq \mathbb{R}^n$ satisfying $-K = K$ and*

$$\min_{k=1}^n \min_{S\subseteq[n]:|S|=k} \mathrm{vol}_k(K \cap W_S) \geq 1,$$

and any $\mathbf{y} \in (-1, 1)^n$, there exists an $\mathbf{x} \in [-1, 1]^n$ with $|\mathrm{fixed}(\mathbf{x})| \geq \lceil\epsilon_0 n\rceil$ and $\mathbf{x} - \mathbf{y} \in cK$, where $\mathrm{fixed}(\mathbf{x}) := \{i \in [n] : |x_i| = 1\}$.

We are now ready to complete the proof.

Proof of Theorem 2. It suffices to show that $\mathrm{disc}(\mathbf{A}) \lesssim \sqrt{n}\,\mathrm{detlb}(\mathbf{A})$, since this implies that for any submatrix $\mathbf{B}$ of $\mathbf{A}$ with k columns we also have

$$\mathrm{disc}(\mathbf{B}) \lesssim \sqrt{k}\,\mathrm{detlb}(\mathbf{B}) \leq \sqrt{n}\,\mathrm{detlb}(\mathbf{A}).$$

Using Lemma 17, we construct a sequence of partial colorings $\mathbf{x}_0 = \mathbf{0}, \ldots, \mathbf{x}_T \in \{-1, +1\}^n$, where $T \lesssim 1 + \log_{1/(1-\epsilon_0)}(n)$, each $\mathbf{x}_t \in [0, 1]^n$, and

$$\|\mathbf{A}(\mathbf{x}_t - \mathbf{x}_{t-1})\|_\infty \lesssim \sqrt{n(1-\epsilon_0)^{t-1}}\,\mathrm{detlb}(\mathbf{A}). \tag{9}$$

To construct $\mathbf{x}_1$, we apply Lemma 17 to $\mathbf{y} := \mathbf{0}$, and $K := \mathrm{volLB}(A) \cdot K_\mathbf{A}$. By the definition of $\mathrm{volLB}(A)$, this K satisfies the assumption of the lemma, and we let $\mathbf{x}_1$ equal the $\mathbf{x}$ guaranteed by the lemma. Since $\mathbf{x}_1 \in cK = c \cdot \mathrm{volLB}(A) \cdot K_\mathbf{A}$, by the definition of $K_\mathbf{A}$ we have that

$$\|\mathbf{A}\mathbf{x}_1\|_\infty \leq c \cdot \mathrm{volLB}(A) \cong \mathrm{volLB}^*(A) \lesssim \sqrt{n}\,\mathrm{detlb}(\mathbf{A}),$$

where the last two inequalities follow, respectively, by Claim 14 and by Lemma 15. In general, to get the bound (9) for $\mathbf{x}_t - \mathbf{x}_{t-1}$ for $t \geq 2$, we set $S := [n] \setminus \mathrm{fixed}(\mathbf{x}_{t-1})$, and apply Lemma 17 with $\mathbf{y} := \mathbf{\Pi}_S\mathbf{x}_1$, and $K := \mathrm{volLB}(\mathbf{A}_S) \cdot K_{\mathbf{A}_S}$, where $\mathbf{A}_S$ is the submatrix of $\mathbf{A}$ consisting of the columns indexed by S. If $\mathbf{x} \in [-1, +1]^S$ is the partial coloring guaranteed by the lemma, we define $\mathbf{x}_t$ by setting its coordinates in

12

S to equal the corresponding coordinates in $\mathbf{x}$, and the remaining coordinates to equal the corresponding coordinates in $\mathbf{x}_{t-1}$. It is straightforward to check that $\mathrm{fixed}(\mathbf{x}_t) \geq (1 - (1 - \epsilon_0)^t)n$ and (9) hold for all t. Moreover, once $t \geq T \geq 1 + \log_{1/(1-\epsilon_0)}(n)$, we must have $\mathbf{x}_t \in \{-1, +1\}^n$.

Having constructed $\mathbf{x}_1, \ldots, \mathbf{x}_T$, we observe that, by (9) and the triangle inequality,

$$\|\mathbf{A}\mathbf{x}\|_\infty \lesssim \sqrt{n} \cdot \mathrm{detlb}(\mathbf{B}) \left(1 + \sqrt{(1 - \epsilon_0)} + \sqrt{(1 - \epsilon_0)^2} + \cdots\right) \cong \sqrt{n} \cdot \mathrm{detlb}(\mathbf{A}).$$

This completes the proof. $\qquad\square$

Acknowledgements

References

[AAGM15]   Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D. Milman. *Asymptotic geometric analysis. Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[Bal89]   Keith Ball. Volumes of sections of cubes and related problems. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 251–260. Springer, Berlin, 1989.

[Ban93]   Wojciech Banaszczyk. Balancing vectors and convex bodies. *Studia Math.*, 106(1):93–100, 1993.

[Ban98]   Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4):351–360, 1998.

[Ban10]   Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 3–10. IEEE, 2010.

[BDG16]   Nikhil Bansal, Daniel Dadush, and Shashwat Garg. An algorithm for komlos conjecture matching banaszczyk's bound. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 788–799. IEEE, 2016.

[CL01]   Bernard Chazelle and Alexey Lvov. The discrepancy of boxes in higher dimension. *Discrete & Computational Geometry*, 25:519–524, 2001.

[DNTTJ18] Daniel Dadush, Aleksandar Nikolov, Kunal Talwar, and Nicole Tomczak-Jaegermann. Balancing vectors in any norm. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1–10. IEEE, 2018.

[DR50]   A. Dvoretzky and C. A. Rogers. Absolute and unconditional convergence in normed linear spaces. *Proc. Nat. Acad. Sci. U.S.A.*, 36:192–197, 1950.

[Fra18]   Cole Franks. A simplified disproof of beck's three permutations conjecture and an application to root-mean-squared discrepancy. *arXiv preprint arXiv:1811.01102*, 2018.

[GH62]   Alain Ghouila-Houri. Caractérisation des matrices totalement unimodulaires. *Comptes Redus Hebdomadaires des Séances de l'Académie des Sciences (Paris)*, 254:1192–1194, 1962.

[Glu89]   Efim Davydovich Gluskin. Extremal properties of orthogonal parallelepipeds and their applications to the geometry of banach spaces. *Mathematics of the USSR-Sbornik*, 64(1):85, 1989.

[JR22]   Haotian Jiang and Victor Reis. A tighter relation between hereditary discrepancy and determinant lower bound. In *Symposium on Simplicity in Algorithms (SOSA)*, pages 308–313. SIAM, 2022.

[Kun23]   Dmitriy Kunisky. The discrepancy of unsatisfiable matrices and a lower bound for the komlós conjecture constant. *SIAM Journal on Discrete Mathematics*, 37(2):586–603, 2023.

[Lav18]   Misha Lavrov. Finding sum of k from zero to n of the absolute value of twice k minus n times n choose k, 2018.

[LSV86]   László Lovász, Joel Spencer, and Katalin Vesztergombi. Discrepancy of set-systems and matrices. *European Journal of Combinatorics*, 7(2):151–160, 1986.

[LV89]   L. Lovász and K. Vesztergombi. Extremal problems for discrepancy. In *Irregularities of partitions (Fertőd, 1986)*, volume 8 of *Algorithms Combin. Study Res. Texts*, pages 107–113. Springer, Berlin, 1989.

[Mat95]   J. Matoušek. Tight Upper Bounds for the Discrepancy of Halfspaces. *Discrete and Computational Geometry*, 13(1):593–601, 1995.

[Mat09]   Jiří Matoušek. *Geometric discrepancy: An illustrated guide*, volume 18. Springer Science & Business Media, 2009.

[Mat13]   Jiří Matoušek. The determinant bound for discrepancy is almost tight. *Proceedings of the American Mathematical Society*, 141(2):451–460, 2013.

[MNT18]   Jiří Matoušek, Aleksandar Nikolov, and Kunal Talwar. Factorization Norms and Hereditary Discrepancy. *International Mathematics Research Notices*, 2020(3):751–780, 03 2018.

[Nik15]   Aleksandar Nikolov. Randomized rounding for the largest simplex problem. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 861–870, 2015.

[Nik17]   A. Nikolov. Tighter bounds for the discrepancy of boxes and polytopes. *Mathematika*, 63(3):1091–1113, 2017.

[NNN12]   Alantha Newman, Ofer Neiman, and Aleksandar Nikolov. Beck's three permutations conjecture: A counterexample and some consequences. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 253–262. IEEE, 2012.

[Pál10]   Dömötör Pálvölgyi. Indecomposable coverings with concave polygons. *Discrete & Computational Geometry*, 44(3):577–588, 2010.

[Roc70]   R. Tyrrell Rockafellar. *Convex analysis*, volume No. 28 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1970.

[Sch98]   Alexander Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, 1998.

[Ver18]   Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

A   Upper Bound on Hereditary Discrepancy for Set Systems

Here we give a different proof of the special case of Theorem 2 for matrices $\mathbf{A}$ with entries in $\{0, 1\}$, i.e., incidence matrices of set systems. This proof is more elementary, and has the interesting property that the upper bound on discrepancy is certified by a uniformly random coloring.

We need is a connection between the hereditary discrepancy of a set system $\mathcal{S}$ on a universe X, and its VC dimension. Recall that the VC dimension, denoted $\dim(\mathcal{S})$, is the largest size of a set $Y \subseteq X$ such that the restriction $\mathcal{S}|_Y \coloneqq \{S \cap Y : S \in \mathcal{S}\}$ equals the powerset of Y. Equivalently, we can define the VC dimension $\dim(\mathbf{A})$ of an incidence matrix $\mathbf{A} \in \{0, 1\}^{m \times n}$ as the largest N for which the power matrix $\mathbf{P}_N$ can be found as a submatrix of $\mathbf{A}$.

A.1 Upper Bound in Terms of VC Dimension

Let us introduce some terminology from stochastic processes that we use in our proof. For a metric space (T, d), let $\mathcal{N}(T, d, \epsilon)$ be the *covering number of T* i.e. $\mathcal{N}(T, d, \epsilon)$ is the smallest number of closed balls with centers in T and radii ϵ whose union covers T. Further, let $\|Y\|_{\psi_2}$ be the sub-gaussian norm of a real-valued random variable Y where $\|Y\|_{\psi_2} := \inf\{t \geq 0 : \mathbb{E}\exp\left(Y^2/t^2\right) \leq 2\}$. Finally, for a class of real-valued functions $\mathcal{F}$ defined on a probability space (Ω, μ), where Ω is a finite set, we define the $L^2(\mu)$ norm by $\|f\|_{L^2(\mu)} := \left(\sum_{\omega \in \Omega} |f(\omega)|^2 \mu(\omega)\right)^{1/2}$. If μ is the uniform measure on Ω, then we simple write L^2 rather than $L^2(\mu)$.

The following result is a consequence of well-known lemmas. We recount it here for completeness.

Lemma 18. *For any non-constant matrix $\mathbf{A} \in \{0, 1\}^{m \times n}$,*

$$\mathrm{herdisc}(\mathbf{A}) \lesssim \sqrt{n \cdot \dim(\mathbf{A})}.$$

Proof. It is enough to prove that $\mathrm{disc}(\mathbf{A}) \lesssim \sqrt{n \cdot \dim(\mathbf{A})}$ since, applying this inequality to any submatrix $\mathbf{B}$ consisting of a subset of k columns from $\mathbf{A}$ shows that, $\mathrm{disc}(\mathbf{B}) \lesssim \sqrt{k \cdot \dim(\mathbf{B})} \leq \sqrt{n \cdot \dim(\mathbf{A})}$.

We prove that $\mathrm{disc}(\mathbf{A}, \mathbf{x}) \lesssim \sqrt{n \cdot \dim(\mathbf{A})}$ is satisfied in expectation by a uniformly random coloring $\mathbf{x} \in \{-1, +1\}^n$ with entries $X_1, X_2, \ldots X_n$ which are independent Rademacher random variables (i.e., uniform in $\{-1, +1\}$). Let $\mathcal{F}$ denote the class of indicator functions defined by the rows of $\mathbf{A}$, i.e., for every $i \in [m]$, we define a function $f_i : [n] \to \{0, 1\}$ given by $f_i(j) = a_{i,j}$. We will show that

$$\mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i f(i) \right| \lesssim \sqrt{\dim(\mathcal{F})}.$$

For each indicator function f, let the random variable $Z_f := \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i f(i) \right|$. Note that this differs from the discrepancy of the row indicated by f by a multiple of $\sqrt{n}$, i.e., $\mathrm{disc}(\mathbf{A}, \mathbf{x}) = \sqrt{n} \cdot \sup_{f \in \mathcal{F}} Z_f$. Consider the random process $(Z_f)_{f \in \mathcal{F}}$. We will apply Dudley's inequality (Lemma 19) to show that

$$\mathbb{E} \sup_{f \in \mathcal{F}} Z_f \lesssim \int_0^1 \sqrt{\log \mathcal{N}\left(\mathcal{F}, L^2, \epsilon\right)} d\epsilon. \tag{10}$$

In order to apply Dudley's inequality we must show that $(Z_f)_{f \in \mathcal{F}}$ has sub-gaussian increments. Note that, since $\|X_i\|_{\psi_2} \lesssim 1$, we have

$$\|Z_f - Z_g\|_\psi = \frac{1}{\sqrt{n}} \left\| \left| \sum_{i=1}^{n} X_i f(i) \right| - \left| \sum_{i=1}^{n} X_i g(i) \right| \right\|_{\psi_2}$$

$$\leq \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{n} X_i (f - g)(i) \right\|_{\psi_2} \lesssim \left(\frac{1}{n} \sum_{i=1}^{n} (f - g)(i)^2 \right)^{1/2},$$

where the second step follows from the reverse triangle inequality, and the final step by Hoeffding's lemma. The right hand side is $\|f - g\|_{L^2}$ so when we apply Dudley's inequality as shown in Lemma 19, we obtain Equation 10.

Using Theorem 20, we can bound the covering number with respect to the normalized L_2 norm as

$$\log \mathcal{N}\left(\mathcal{F}, L^2, \epsilon\right) \lesssim \dim(\mathcal{F}) \log\left(\frac{2}{\epsilon}\right).$$

Plugging the right hand side into the integral in Equation 10 and integrating, we have $\mathbb{E} \sup_{f \in \mathcal{F}} Z_f \lesssim \sqrt{\dim(\mathcal{F})}$. Recall that the discrepancy of the row indicated by f is $\sqrt{n} \cdot Z_f$, thus the hereditary discrepancy is bounded above as $\mathrm{herdisc}(\mathbf{A}) \lesssim \sqrt{n \dim(\mathcal{F})}$, as was our goal. $\square$

Lemma 19. (Dudley's Inequality, Remark 8.1.5 [Ver18]). *Let $(X_t)_{t \in T}$ be a random process on a metric space (T, d) with sub-gaussian increments i.e. there exists a $K \geq 0$ such that $\|X_t - X_s\|_{\psi_2} \leq K d(t, s)$ for all $t, s \in T$. Then*

$$\mathbb{E} \sup_{t, s \in T} |X_t - X_s| \lesssim K \int_0^\infty \sqrt{\log \mathcal{N}(T, d, \epsilon)} d\epsilon.$$

Theorem 20. (Covering Numbers via VC Dimension, 8.3.18 [Ver18]). *Let $\mathcal{F}$ be a class of Boolean functions on a probability space (Ω, Σ, μ). Then, for every $\epsilon \in (0, 1)$, we have*

$$\mathcal{N}\left(\mathcal{F}, L^2(\mu), e\right) \leq \left(\frac{2}{\epsilon}\right)^{C \cdot \dim(\mathcal{F})}$$

for an absolute constant C.

We note that Lemma 18 can likely be improved further, for example by following the techniques of Matoušek [Mat95], and carefully tracking constants.

A.2 Connection to the Determinant Lower Bound

To finish the proof, it remains to show a connection between VC dimension and the determinant lower bound. To do so, we show that a matrix $\mathbf{A} \in \{0, 1\}^{m \times n}$ with large VC dimension must contain a submatrix with large determinant. This submatrix is a binary version of the Hadamard matrix, described next.

Let the 0-1 Hadamard matrix be the $\{0, 1\}$ matrix obtained by applying the linear map $a \mapsto (a + 1)/2$ to all of the entries in the standard ± 1 Hadamard matrix. Denote the $n \times n$ 0-1 and standard Hadamard matrices by $\tilde{\mathbf{H}}_n$ and $\mathbf{H}_n$ respectively. We prove the following.

Claim 21. $\left| \det \left(\tilde{H}_n \right) \right| \geq 2^{-n} \cdot n^{n/2}$.

Proof. Consider $\mathbf{H}_n$ and suppose w.l.o.g. that its first rows is the all ones row. Add this row to all the other rows. Observe that all the other rows now have entries in $\{0, 2\}$. Scale them down by a factor of two. Adding one row to another does not change the determinant. Scaling a row scales the determinant by the same amount. Since $|\det(\mathbf{H}_n)| = n^{n/2}$, $\left| \det \left(\tilde{\mathbf{H}}_n \right) \right| = 2^{-n} \cdot n^{n/2}$. $\square$

We can now finish the proof of Theorem 2 for $\mathbf{A} \in \{0, 1\}^{m \times n}$. If $\mathbf{A}$ is a constant matrix (i.e., all its entries are equal), the bound is trivial, so we assume otherwise. The upper bound arises from the pair of inequalities herdisc$(\mathbf{A}) \lesssim \sqrt{n \dim(\mathbf{A})}$ and $\sqrt{\dim(\mathbf{A})} \lesssim$ detlb$(\mathbf{A})$. The former inequality is achieved by a random coloring, as shown in Lemma 18. The latter follows by considering the power matrix $\mathbf{P}_{\dim(\mathbf{A})}$, which is a submatrix of $\mathbf{A}$. Since every $\dim(\mathbf{A}) \times \dim(\mathbf{A})$ 0-1 matrix is a submatrix of $\mathbf{P}_{\dim(\mathbf{A})}$, we can find also find $\tilde{\mathbf{H}}_{\dim(\mathbf{A})}$ as a submatrix of $\mathbf{P}_{\dim(\mathbf{A})}$, and, therefore, of $\mathbf{A}$. By Claim 21 we know that $\left| \det \left(\tilde{\mathbf{H}}_{\dim(\mathbf{A})} \right) \right| \geq 2^{-\dim(\mathbf{A})} \cdot (\dim(\mathbf{A}))^{\dim(\mathbf{A})/2}$. It follows that

$$\text{detlb}(\mathbf{A}) \geq \left| \det \left(\tilde{\mathbf{H}}_{\dim(\mathbf{A})} \right) \right|^{1/\dim(\mathbf{A})} \gtrsim \sqrt{\dim(\mathbf{A})}.$$

Thus, the two inequalities together give us detlb$(\mathbf{A}) \gtrsim$ herdisc$(\mathbf{A})/\sqrt{n}$ as required.