THE STRUCTURE OF EXTREMAL BAD SCIENCE MATRICES

SHRIDHAR SINHA

Abstract. The *bad science matrix problem* is concerned with identifying, among all matrices $A \in \mathbb{R}^{n\times n}$ whose rows are normalized in the ℓ_2 norm, the matrix that maximizes

$$\beta(A) = \frac{1}{2^n} \sum_{x\in\{-1,1\}^n} \|Ax\|_\infty .$$

Steinerberger [5] proved that the asymptotically optimal rate is given by $(1 + o(1))\sqrt{2\log n}$ and this rate is attained for matrices with i.i.d. random ± 1 entries. More recently, explicit constructions [3] have been obtained with $\beta(A) \geq \sqrt{\log_2(n+1)}$, which is only about 18% smaller than the asymptotic rate. We bridge these results by showing a geometric interpretation of these explicit constructions and demonstrating how to achieve the asymptotic rate through an explicit construction. The key ingredient is to establish a connection between a natural partition of the n-dimensional hypercube induced by the rows of A and isoperimetrically extremal partitions of the hypercube.

1. Introduction and Results

1.1. Introduction.

The *bad science matrix problem* models a “dishonest” testing scenario, where a researcher runs many fair statistical tests on random data in hopes of finding at least one atypical result. Concretely, let $A \in \mathbb{R}^{n\times n}$ be a matrix whose rows $a_i \in \mathbb{R}^n$ are normalized in ℓ_2, $\|a_i\|_2 = 1$. We consider the quantity

$$\beta(A) \;=\; \frac{1}{2^n} \sum_{x\in\{\pm 1\}^n} \|Ax\|_\infty ,$$

which is equivalently the expectation $\mathbb{E}[\|AX\|_\infty]$ for a random Rademacher vector $X \in \{\pm 1\}^n$. Here $\|Ax\|_\infty = \max_{1\leq i\leq n} |\langle a_i, x\rangle|$ is the largest absolute value among the linear tests a_i applied to the data vector x. In statistical terms, each row of A represents a (fair) linear test of the n random signs in x, and $\beta(A)$ measures the average largest test statistic over all 2^n possible outcomes. The problem is to understand how large $\beta(A)$ can be under the unit-norm constraint on the rows. This setup has a natural interpretation both in hypothesis testing and in geometry. From the testing viewpoint, a *bad scientist* pre-selects many unit-norm test directions and then looks at a sequence of coin flips. Even though the coin is fair, by chance, one of these tests will often yield a surprisingly large value, yielding an (incorrect) claim of significance. If any test is unusually large, the researcher obtains a small (but spurious) p-value and (incorrectly) concludes that the fair coin is biased. Measuring this effect, let . In geometric terms, the matrix A maps the vertices of the discrete hypercube $\{\pm 1\}^n$ into $\mathbb{R}^n$; one then asks whether a typical image point has a coordinate that is *significantly* larger than average. Steinerberger observes that such matrices correspond to affine images of the cube whose points "on average [have] at least one large coordinate". Equivalently, one can view $\beta(A)$ as measuring how far the hypercube can be "rotated" or embedded so that its vertices tend to lie outside the smaller cubes in ℓ_∞-norm. This dual perspective connects to classical discrepancy and vector-balancing problems (for instance, to the Komlós conjecture in discrepancy theory), but the bad science problem is a distinct "functional-balancing" variant. In summary, the bad science matrix problem captures the risk of false positives when many fair tests are run, and it raises fundamental questions about how a bounded-norm linear map can distort the discrete cube's geometry.

1.2. **Existing Results.** Steinerberger [5] established the first asymptotic bounds for this problem. He proved that, as $n \to \infty$, the maximum possible value of $\beta(A)$ (over all $n \times n$ matrices with unit-ℓ_2 rows) grows like

$$\max_{\|a_i\|_2=1} \beta(A) \;=\; (1+o(1))\sqrt{2\log n}\,.$$

This result shows that the worst-case average sup-norm is of order $\sqrt{2\log n}$. Moreover, the proof shows that this rate is attained (up to lower-order terms) by a random matrix with independent $\pm 1/\sqrt{n}$ entries. In other words, a matrix with i.i.d. Rademacher rows (properly scaled) typically achieves $\beta(A) = (1+o(1))\sqrt{2\log n}$, matching the theoretical maximum order given by Gaussian-maxima heuristics. However, all conjectured and verified extremal matrices in lower dimensions(till n=8) are highly structured and of low rank, very unlike the random asymptotic extremizers. An example for n=5 is

$$A \;=\; \frac{1}{2\sqrt{3}}\begin{pmatrix} 2 & 2 & 0 & 0 & 2 \\ -2 & 2 & 0 & 2 & 0 \\ -2 & 0 & 0 & -2 & 2 \\ 0 & -\sqrt{3} & \sqrt{3} & \sqrt{3} & \sqrt{3} \\ 0 & \sqrt{3} & \sqrt{3} & -\sqrt{3} & -\sqrt{3} \end{pmatrix}$$

Building on this, Albors, Bhatti, Ganjoo, Guo, Kunisky, Mukherjee, Stepin, and Zeng [3] provided explicit constructions and structural results. They exhibit concrete $n \times n$ matrices A achieving

$$\beta(A) \;\geq\; \sqrt{\log_2(n+1)}\,,$$

Improving upon trivial bounds and coming within a constant factor of the $\sqrt{2\log n}$ rate. Their construction uses combinatorial designs (e.g. Hadamard-type and tree-based constructions) to ensure that $\|Ax\|_\infty$ is large for many corners x of the cube. In addition, Albors *et al.* prove remarkable structure theorems for extremal matrices: every entry of an optimal bad science matrix must be the square root of a rational number. Using these insights, they completely solve the problem for small dimensions, determining exact maximizing matrices for $n \leq 4$. These results highlight the geometry of extremal examples and show that while random matrices are asymptotically optimal, the true maximizers in lower dimensions exhibit rich algebraic structure.

1.3. **Main results.** Throughout the paper we write

$$\beta(A) = \frac{1}{2^n}\sum_{x\in\{\pm1\}^n} \|Ax\|_\infty,$$

for an $n \times n$ real matrix A whose rows $a_1, \ldots, a_n$ satisfy $\|a_i\|_2 = 1$. For each row index i define the associated cell

$$S_i(A) = \{x \in \{\pm 1\}^n : \|Ax\|_\infty = \langle a_i, x\rangle\},$$

so that $\{S_i(A), -S_i(A) : 1 \leq i \leq n\}$ is the partition of the hypercube induced by the rows of A (see the proofs for the uniqueness/regularity considerations).

The first result gives a concise Fourier–analytic upper bound for $\beta(A)$ and connects the optimisation problem to level–one Fourier mass on cells of the hypercube.

Theorem (2.1 Fourier characterization). *Let $A \in \mathbb{R}^{n\times n}$ have rows normalized in ℓ_2, and let $S_i = S_i(A)$ as above. Denote by $W_1[\mathbf{1}_{S_i}]$ the level–one Fourier weight of the indicator $\mathbf{1}_{S_i}$. Then*

$$\beta(A) \;\leq\; 2\sum_{i=1}^{n} \sqrt{W_1[\mathbf{1}_{S_i}]}\,.$$

An immediate, useful consequence (combining Theorem 2.1 with the Level–1 inequality of Talagrand/O'Donnell) is a universal upper bound on $\beta(A)$ and the asymptotic tightness of the Fourier bound for extremal matrices.

Lemma (2.2 Structure of Extremal Matrices). *For every $n \times n$ matrix A with unit–ℓ_2 rows,*

$$\beta(A) \leq \sqrt{2\log(2n)}.$$

Moreover, any maximizer A of $\beta(A)$ satisfies

$$\beta(A) = 2(1 + o(1)) \sum_{i=1}^{n} \sqrt{W_1[\mathbf{1}_{S_i}]},$$

and the sets S_i have (up to multiplicative $1 + o(1)$ variation) equal volume and saturate the Level–1 inequality up to $1 + o(1)$ factors.

The next result explains why the explicit, highly structured low–dimensional constructions that appear in the literature (e.g. balanced tree / subcube constructions) fail to reach the asymptotic optimum: their induced partition is by subcubes, which are isoperimetrically nonoptimal for the level–one weight.

Lemma (2.3 Subcubes are Suboptimal). *Let $T \subset \{\pm 1\}^n$ be a subcube of co-dimension k (that is, k coordinates are fixed and $n - k$ are free). Then*

$$\sqrt{W_1[\mathbf{1}_T]} = \frac{\sqrt{k}}{n}.$$

In particular, the balanced co-dimension $\lfloor \log_2 n\rfloor$ subcube partition yields

$$\beta_{\text{subcube}} = \sqrt{\log_2 n + 1},$$

which is about 18% below the asymptotic optimal rate $\sqrt{2\log n}$.

The following theorem provides a precise asymptotic expansion for two concrete constructions that attain the optimal leading order: (i) normalized random sign matrices and (ii) deterministic orthonormal almost–Hadamard matrices obtained by truncating and orthonormalizing a Hadamard block. The expansion records the second–order term coming from classical Gaussian extreme–value theory and gives an explicit CLT error term.

Theorem (2.4 Asymptotics for explicit constructions)**.** *Let $n \geq 3$. For both the normalized random sign matrix S (rows i.i.d. uniform in $\{\pm 1\}^n$ and scaled by $1/\sqrt{n}$) and for any orthonormal almost–Hadamard matrix Q (top–left $n \times n$ block of an $m \times m$ Hadamard matrix, then orthonormalized), one has the expansion*

$$\beta(\,\cdot\,) = \sqrt{2\log(2n)} - \frac{\log\log(2n)}{2\sqrt{2\log(2n)}} + o\big(\log n^{-\frac{1}{2}}\big). \tag{1}$$

In particular both constructions attain the correct leading order $\sqrt{2\log(2n)}$, and the error from the Gaussian extreme–value prediction is at most $o\big(n^{-1/2}(\log n)^7\big)$.

This makes explicit the $(1 + o(1))$ statement in Steinerberger's original paper. Steinerberger proved the rate $\beta_{\text{opt}} = (1 + o(1))\sqrt{2\log n}$. Combining the refined CLT control used to prove Theorem 2.4, and the Gaussian extreme–value expansion, we obtain the rate in Theorem 2.4 as the correct scaling for the optimum(under the conditions that the matrix is invertible and the entries are bounded).

Theorems 2.1–2.4 tell the complete story developed in the proofs. An extremal matrix induces a nearly centroidal Voronoi partition of the hypercube: each cell must be simultaneously equal in volume and nearly isoperimetrically optimal for level–one Fourier mass. Explicit low–rank constructions correspond to subcube partitions and hence incur an unavoidable deficit. Finally, random signs and almost–Hadamard truncations realize the Gaussian maximum and match the second–order extreme–value expansion up to an explicit and negligible CLT error.

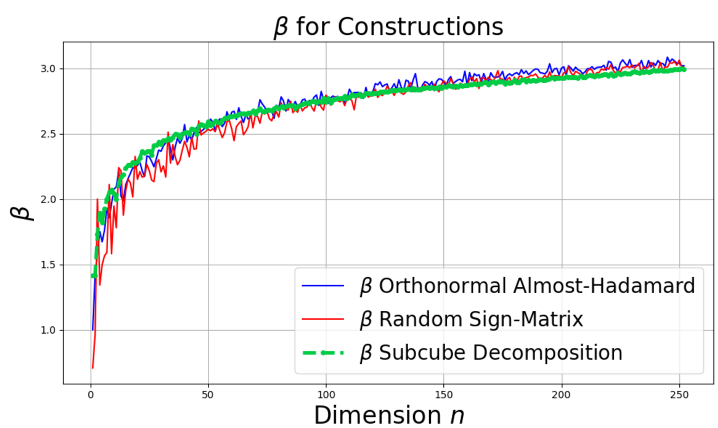


Figure 1. Empirical results with known Hadamard matrices show that our construction, as well as the random one, exceed the subcube construction fairly quickly.

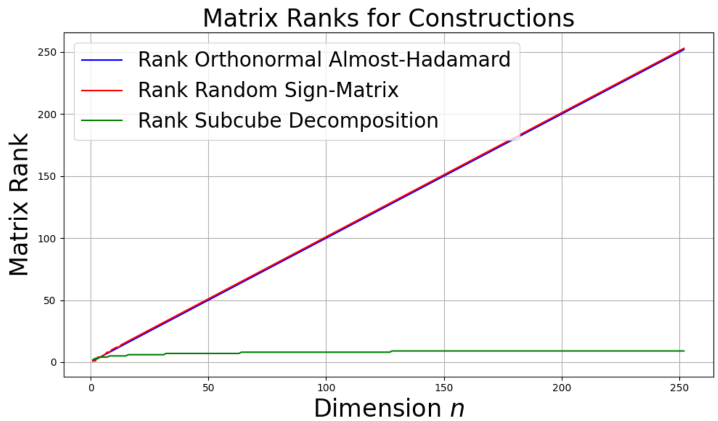


FIGURE 2. There is, however, a significant trade-off in the simplicity of the matrix as a transformation.

2. Proofs

We note here that for the rest of the paper, we can and will narrow our optimization argument to the following dense open subset of

$$\mathcal{A} = \left\{ A \in \mathbb{R}^{n\times n} : \|a_i\|_2 = 1,\ i = 1, \dots, n \right\},$$

namely

$$\mathcal{A}^\circ = \mathcal{A} \setminus \bigcup_{\substack{x\in\{\pm1\}^n \\ i\neq j}} \left\{ A : \langle a_i, x\rangle^2 - \langle a_j, x\rangle^2 = 0 \right\}.$$

On $\mathcal{A}^\circ$, for each hypercube vector $x \in \{\pm1\}^n$ the value $\|Ax\|_\infty = \max_i |\langle a_i, x\rangle|$ is attained at a unique row index.

Proof. Fix any $x \in \{\pm1\}^n$ and two distinct rows $i \neq j$. The "tie-condition"

$$|\langle a_i, x\rangle| \;=\; |\langle a_j, x\rangle| \quad\Longleftrightarrow\quad \langle a_i, x\rangle^2 \;-\; \langle a_j, x\rangle^2 = 0$$

is a nontrivial polynomial equation in the $2n$ entries of a_i and a_j. Hence its zero-set

$$H_{x,i,j} = \left\{ A \in \mathcal{A} : \langle a_i, x\rangle^2 - \langle a_j, x\rangle^2 = 0 \right\}$$

is a real hypersurface (codimension ≥ 1) in the compact manifold $\mathcal{A}$. There are only finitely many such triples (x, i, j), so $\mathcal{Z} = \bigcup H_{x,i,j}$ is a finite union of measure-zero hypersurfaces, and $\mathcal{A}^\circ = \mathcal{A} \setminus \mathcal{Z}$ is open and dense.

Meanwhile, the objective

$$\beta(A) = \sum_{x\in\{\pm1\}^n} \|Ax\|_\infty$$

is continuous (indeed Lipschitz) on $\mathcal{A}$. By compactness, it attains its maximum, and continuity implies any maximizer can be approximated arbitrarily well by points in $\mathcal{A}^\circ$. Thus, there is no loss of generality in restricting our search to $\mathcal{A}^\circ$, where for every x the maximizer of $\|Ax\|_\infty$ is unique.

$\square$

For the following results, we first outline the Fourier analysis of the indicator functions for subsets of the cube:

$$\mathbf{1}_B(x) = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{otherwise.} \end{cases}$$

The function $\mathbf{1}_B$ has the Fourier expansion

$$\mathbf{1}_B(x) = \sum_{S \subset [n]} \widehat{\mathbf{1}_B}(S)\chi_S(x),$$

where $\chi_S(x) = \prod_{i \in S} x_i$ and

$$\widehat{\mathbf{1}_B}(S) = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} \mathbf{1}_B(x)\chi_S(x).$$

Note that the sum of the inner products over pairs in B can be written as

$$\sum_{x,y \in B} \langle x, y \rangle = \sum_{x,y \in \{-1,1\}^n} \mathbf{1}_B(x)\mathbf{1}_B(y)\langle x, y \rangle.$$

Express each coordinate of the inner product and interchange summations:

$$\sum_{x,y \in \{-1,1\}^n} \mathbf{1}_B(x)\mathbf{1}_B(y)\langle x, y \rangle = \sum_{i=1}^{n} \left(\sum_{x \in \{-1,1\}^n} x_i \mathbf{1}_B(x) \right)^2.$$

Observing that by the definition of the Fourier coefficients, we have

$$\widehat{\mathbf{1}_B}(\{i\}) = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} x_i \mathbf{1}_B(x),$$

It follows that

$$\sum_{x,y \in B} \langle x, y \rangle = 2^{2n} \sum_{i=1}^{n} \widehat{\mathbf{1}_B}(\{i\})^2.$$

Define the *level-1 Fourier weight* by

$$W_1[\mathbf{1}_B] := \sum_{i=1}^{n} \widehat{\mathbf{1}_B}(\{i\})^2.$$

Thus, we have

$$\sum_{x,y \in B} \langle x, y \rangle = 2^{2n} \cdot W_1[\mathbf{1}_B].$$

We make extensive use of a well-known "Level-1 Inequality" as stated in O'Donnell's book[6]. They cite it as perhaps first originating from a paper of Talagrand [7]

Lemma 2.1 (Level-1 Inequality, O'Donnell [6] 5.28)**.** *Let $f : \{-1,1\}^n \to \{0,1\}$ be a Boolean function with mean $\mathbb{E}[f] = \alpha \leq \frac{1}{2}$. Then*

$$W_1[f] \leq 2\alpha^2 \log\left(\frac{1}{\alpha}\right),$$

which is asymptotically sharp for Hamming Balls.

We use this to give a structural result about the extremal Bad Science matrices.

Theorem 2.1 (Fourier Characterization)**.** *Let A be an $n \times n$ matrix that has rows normalized in ℓ_2. Define the subset*

$$S_i = \{x \in Q_n \quad \|Ax\|_\infty = \langle A_i, x\rangle\}.$$

Then, the value of $\beta(A)$ satisfies

$$\beta(A) \leq 2\sum_{i=1}^{n} \sqrt{W_1[\mathbf{1}_{S_i}]}.,$$

Proof.

$$\beta(A) := \frac{1}{2^n} \sum_{x \in Q_n} \|Ax\|_\infty.$$

For each row A_i of A, there exists the largest subset $S_i \subset Q_n$ such that

$$\forall x \in S_i, \quad \|Ax\|_\infty = \langle A_i, x\rangle,$$

By our note from earlier, we also know that these sets are disjoint, and cover half the hypercube, while the other half is covered by their negatives, which are also disjoint.

Enumerating elements of S_i as $v_1, ..., v_{|S_i|}$, we can write

$$\beta(A) = \frac{1}{2^n} \sum_{i=1}^{n} \sum_{j=1}^{|S_i|} \langle A_i, 2v_j\rangle = \frac{1}{2^n} \sum_{i=1}^{n} \left\langle A_i, 2\sum_{j=1}^{|S_i|} v_j \right\rangle.$$

Define

$$B_i := 2\sum_{j=1}^{|S_i|} v_j.$$

So we have

$$\beta(A) = \frac{1}{2^n} \sum_{i=1}^{n} \langle A_i, B_i\rangle.$$

By the equality case of Cauchy-Schwarz, to maximize each summand, we must have

$$A_i = \frac{B_i}{\|B_i\|}.$$

By definition

$$\|B_i\|^2 = 4 \sum_{x,y \in S_i} \langle x, y\rangle.$$

It follows from the Fourier analysis introduction that

$$\beta(A) \leq \frac{1}{2^n} \sum_{i=1}^{n} \|B_i\| = 2\sum_{i=1}^{n} \sqrt{W_1[\mathbf{1}_{S_i}]}.$$

□

In combination with a result from Steinerberger's original paper, we get a quick corollary which elucidates the connection of the Bad Science Matrix problem with an isoperimetric tiling problem on the Boolean hypercube.

Lemma 2.2 (Structure of Extremal Matrices)**.** *Let A be an $n \times n$ matrix with rows normalized in ℓ_2 that has optimal beta value,*

$$\beta(A) = 2(1 + o(1)) \sum_{i=1}^{n} \sqrt{W_1[\mathbf{1}_{S_i}]}.$$

Furthermore, the S_i have equal size up to an $1 + o(1)$ multiplicative difference, and The associated indicators $\mathbf{1}_{S_i}$ saturate the level-one inequality up to an $1 + o(1)$ multiplicative difference.

Proof. Applying the Level-1 inequality, we have

$$W_1[\mathbf{1}_{S_i}] \leq 2\alpha^2 \log \frac{1}{\alpha}$$

where

$$\alpha = \frac{|S_i|}{2^{n-1}}$$

the 'volume' of S_i.

Noting that the volumes must add up to $\frac{1}{2}$, define the upper bound function on the square root of the level-one weight

$$f(x) = x\sqrt{2 \log \frac{1}{x}}, \quad x \in (0, 1].$$

Then for a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$,

$$F(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i \sqrt{2 \log \frac{1}{\alpha_i}} = \sum_{i=1}^{n} f(\alpha_i).$$

Since f is strictly concave on $(0, 1]$, we use Jensen's inequality

$$\frac{1}{n} \sum_{i=1}^{n} f(\alpha_i) \leq f\Big(\frac{1}{n} \sum_{i=1}^{n} \alpha_i\Big) = f\Big(\frac{1}{2n}\Big).$$

Multiplying by n gives

$$F(\boldsymbol{\alpha}) = \sum_{i=1}^{n} f(\alpha_i) \leq n\, f\Big(\frac{1}{2n}\Big) = n \cdot \frac{1}{2n} \sqrt{2 \log 2n} = \frac{1}{2} \sqrt{2 \log 2n}.$$

Equality occurs iff $\alpha_i = \frac{1}{2n}$ for all i. Combining this with Theorem 1 we get:

$$\beta(A) \leq_{\text{Cauchy-Schwarz}} 2 \sum_{i=1}^{n} \sqrt{W_1[\mathbf{1}_{S_i}]} \leq_{\text{Jensen}} \sqrt{2 \log 2n}.$$

Steinerberger showed that

$$(1 + o(1))\sqrt{2 \log n} \leq \beta(A).$$

So the chain of upper bounds must be asymptotically tight. This immediately gives us

$$\beta(A) = 2(1 + o(1)) \sum_{i=1}^{n} \sqrt{W_1[\mathbf{1}_{S_i}]}.$$

We get the statements on size and level-1 weight by noting

$$f\big((1+\varepsilon)x\big) = (1+\varepsilon)x\sqrt{2\log\frac{1}{(1+\varepsilon)x}} = (1+\varepsilon)x\sqrt{2\log\frac{1}{x}}\,\sqrt{1-\frac{\log(1+\varepsilon)}{\log(1/x)}}.$$

Taylor expanding gives us

$$f\big((1+\varepsilon)x\big) = \, f(x)\,(1+\varepsilon)\left[1-\tfrac{1}{2}\frac{\log(1+\varepsilon)}{\log(1/x)} + O\Big(\big(\tfrac{\log(1+\varepsilon)}{\log(1/x)}\big)^2\Big)\right]$$

Letting $\varepsilon = o(1)$ we obtain

$$f\big((1+\varepsilon)x\big) = f(x)\Big(1+\varepsilon+O\big(\tfrac{\varepsilon}{\log(1/x)}\big)\Big) = (1+o(1))\,f(x),$$

If we let the sizes any bigger than $\frac{(1+o(1))}{2n}$ we will not be within $(1+o(1))$ of the upper bound, and thus be smaller than the lower bound for optimal matrices. □

From a structural point of view this result is saying something fascinating about extremal Bad Science Matrices. We may interpret each S_i as a Voronoi cell when we consider a Voronoi tesselation of the hypercube seeded at the rows A_i and their negatives. The tight Cauchy-Schwarz inequality is saying that this is a nearly centroidal Voronoi tessellation, meaning the centroids of each cell coincide with the seeds. The tight Jensen and Level-1 inequalities add that each cell is roughly equal volume and that each of them are isoperimetrically optimal in the sense that they maximize Level-1 weight. This represents a highly structured partition of the hypercube, which is favorable for many settings in coding theory.

We now turn to an example that resolves part of the central question in the existing papers: Why are the known optimal low-dimensional matrices so highly structured, but their natural generalization becomes suboptimal for large dimensions where random matrices take over?

We first introduce a well-known family of subsets of $\{-1,1\}^n$.

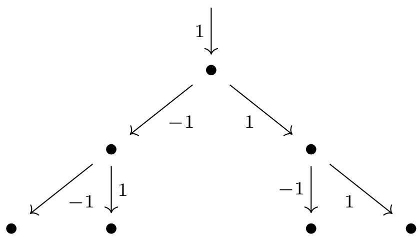
Definition 1 (Subcubes)**.** *Let $n \in \mathbb{N}$ and $k \in \{0,1,\dots,n\}$. Choose an index set $S \subseteq \{1,\dots,n\}$ with $|S| = k$ and a fixed assignment $a \in \{-1,1\}^S$. The subcube of co-dimension k determined by (S,a) is*

$$C_{S,a} = \big\{\, x \in \{-1,1\}^n : x_i = a_i \ \forall\, i \in S \big\}.$$

The paper [3] presents the following construction that has $\beta(\cdot)$ 18% smaller than the optimal rate:

Definition 2 (Highly balanced binary trees)**.** *For a fixed integer $n \geq 1$, fill up a binary tree with vertices from left to right until one has n leaves, and finally add an edge that points into the root.*

We label the edges of such a highly balanced binary tree in the following way: edges that point left have label -1, edges that point right have label 1, and the edge that points to the root has label 1. From here, for a leaf v, walk along the unique path from the root to v. Then the edge labels of this path becomes a row of the matrix



$$\begin{bmatrix} 1 & -1 & -1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

FIGURE 3. An unsatisfiable tree and the corresponding matrix.

this method generates, where if the length of the path is less than n, we make the rest of the entries 0. The case $n = 4$ is illustrated.

It is easy to see then that the root to leaf paths specify the fixed co-ordinates of a n-way subcube partition of the half of the hypercube $\{-1, 1\}^n$ with first coordinate $+1$, by subcubes of co-dimension $\lfloor \log_2 n \rfloor$ and $\lfloor \log_2 n + 1 \rfloor$. Since we leave all the remaining co-ordinates as 0, each row's Voronoi cell is the subcube specified by the fixed co-ordinates. Since the centroid of vectors in a subcube is trivially the vector with the fixed co-ordinates as they are and 0s everywhere else, this is indeed a centroidal Voronoi tessellation. The balanced nature of the tree also guarantees cell sizes within a factor of 2 of each other(not quite asymptotically equal, but in the case that n is a power of 2 we actually have exactly equal sizes). By our characterization of the optimum earlier, we see that the only thing holding us back is the level-1 inequality. Indeed, intuitively speaking, packing a cube with smaller cubes is not isoperimetrically optimal, something we now make precise.

Lemma 2.3 (Subcubes are Suboptimal)**.** *Let T_i be a $\lfloor \log_2 n + 1 \rfloor$ co-dimensional subcube of Q_n. Then,*

$$\sqrt{W_1[\mathbf{1}_{\mathbf{T_i}}]} = \frac{\sqrt{\lfloor \log_2 n + 1 \rfloor}}{n}$$

Proof. The Level-1 weight can be rewritten as

$$\sqrt{\sum_{j=1}^{n} \widehat{\mathbf{1}_{S_i}}(\{j\})^2}$$

On the fixed coordinates, the Fourier weights are simply the volume of S_i, and on the free coordinates, we have

$$\hat{\mathbf{1}_{\mathbf{S_i}}}(x) = 0$$

Since $\lfloor \log_2 n + 1 \rfloor$ of the coordinates are fixed, we plug in to get the required expression.

$$\frac{\sqrt{\lfloor \log_2 n + 1 \rfloor}}{n}$$

∎

This shows that the subcube decomposition in the best case(n is a power of 2) yields $\beta(A) = \sqrt{\log_2(n)+1}$, which is the same as the best derandomized construction found in the existing literature.

One may then easily check that the explicit matrices in [5] arise exactly from subcube partitions of $\{-1,1\}^n$ and thus have asymptotically suboptimal generalizations.

We will now develop a probabilistic viewpoint that shows the connection of the Bad Science Problem with well-studied Gaussian extreme value theory. The setting is as follows:

For a real $n \times n$ matrix A with each row having ℓ_2-norm $\sqrt{n}$, we are interested in

$$\beta(\hat{A}), \quad \text{where} \quad \hat{A} \text{ is } A \text{ with rows normalized in } \ell_2.$$

Let A_i^T denote the i-th row of A^T. Let $\{\varepsilon_i\}_{i=1}^n$ denote a collection of i.i.d. Rademacher random variables $\left(\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}\right)$. Then

$$X_i = A_i^T \, \varepsilon_i, \quad i = 1, \dots, n,$$

are independent mean-zero vectors in $\mathbb{R}^n$. If we consider

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i,$$

the jth coordinate of S_n is precisely the inner product of the row $\hat{A}_j$ with a uniformly random vector in $\{-1,1\}^n$. We are interested in computing

$$\beta(\hat{A}) \;=\; \mathbb{E}\big[\|S_n\|_\infty\big].$$

The reason we chose row-norm $\sqrt{n}$ is so that normalizing by $\sqrt{n}$ gives

$$\Big|\mathbb{P}(\|S_n\|_\infty > t) \;-\; \mathbb{P}(\|Z_n\|_\infty > t)\Big| \;\longrightarrow\; 0,$$

where $Z_n \sim \mathcal{N}(0, \mathbb{E}[S_n S_n^T])$, as studied in the High-Dimensional Central Limit Theorem [8].

Steinerberger already showed that $n \times n$ matrices with i.i.d. Rademacher entries attain the optimal bound for beta for large n. We investigate this random construction further, provide the precise asymptotics for the o(1) term in the optimal beta rate, and present a deterministic construction that converges to the optimal rate as well. We present empirical evidence that this method converges faster, as well as a heuristic explanation for this speed-up.

We first state a version of a 'High Dimensionl Central Limit Theorem' result from Fang, Koike, Liu, and Zhao [8], adjusted to our setting.

Theorem 2.2 (High-Dimensional CLT in the Degenerate Case)**.** *Let $n > 3$ be an integer. Let $X_1, \dots, X_n$ be independent mean-zero random vectors in $\mathbb{R}^n$. Define*

$$W \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i,$$

and let Σ denote its covariance matrix, satisfying, for all distinct $1 \le j, k, \ell \le n$,

$$\Sigma_{jj} = 1, \qquad \det(\Sigma_{j,k}) > \alpha^2 > 0, \qquad \frac{\det(\Sigma_{j,k,\ell})}{\det(\Sigma_{j,k})} > \beta^2 > 0,$$

where $\Sigma_{j,k}$ is the 2×2 sub-matrix of Σ by intersecting the jth and kth rows with the jth and kth columns and $\Sigma_{j,k,\ell}$ the 3×3 sub-matrix defined analogously in the obvious way. Let $G \sim N(0, \Sigma)$. Suppose there exists a constant $B > 0$ such that

$$\mathbb{E}\Big[\exp\big(|X_{ij}|/B\big)\Big] \;\le\; 2 \quad \textit{for all } 1 \le i, j \le n,$$

Then there is an absolute constant C such that

$$\sup_{a,b \in \mathbb{R}^n} \big|P(a \le W \le b) \;-\; P(a \le G \le b)\big| \le \frac{C\,B^3}{\alpha^2 \beta^2 \sqrt{n}} \;\log n^{6.5},$$

We note here that up to polylog(n) factors, this is the best known bound for this quantity. We choose this version as it is convenient to use and, as we shall see, polylog(n) factor improvements will not change the expected leading order of the infinity norm.

The main idea of the proof is that the random matrices from Steinerberger's construction have, upon being fixed, non-zero pairwise inner products with high probability. A classical result, which we now state, shows that identically zero pairwise inner products will provide the best infinity-norm.

Theorem 2.3 (Gaussian Correlation Inequality [9])**.** *Let $Z \sim \mathcal{N}(0, \Sigma)$ be an n-dimensional centered Gaussian, and let $A, B \subset \mathbb{R}^n$ be symmetric convex sets. Then*

$$\Pr[\,Z \in A \cap B\,] \;\ge\; \Pr[\,Z \in A\,]\;\Pr[\,Z \in B\,].$$

Corollary 2.1. *Among all centered Gaussian vectors $Z = (Z_1, \dots, Z_n)$ with $\mathrm{Var}(Z_i) = 1$ for each i, the quantity*

$$\mathbb{E}\big[\|Z\|_\infty\big] \;=\; \mathbb{E}\big[\max_{1 \le i \le n} |Z_i|\big]$$

is maximized when the Z_i are independent (i.e. $\mathrm{Cov}(Z_i, Z_j) = 0$ for all $i \ne j$).

Proof. Fix $t > 0$. Observe

$$\Pr\big\{\|Z\|_\infty \le t\big\} \;=\; \Pr\Big\{\bigcap_{i=1}^{n} \{-t \le Z_i \le t\}\Big\} \;=\; \Pr\Big(\bigcap_{i=1}^{n} A_i\Big),$$

where each

$$A_i = \{\, x \in \mathbb{R}^n : -t \le x_i \le t\}$$

is symmetric and convex. Now apply the Gaussian Correlation Inequality repeatedly: for any covariance matrix Σ with unit diagonal, one shows by induction on n that

$$\Pr_\Sigma\big\{\|Z\|_\infty \le t\big\} \;\ge\; \prod_{i=1}^{n} \Pr_\Sigma\big\{|Z_i| \le t\big\} \;=\; \big(\Pr\{|g| \le t\}\big)^n$$

where $g \sim \mathcal{N}(0, 1)$. Equality holds when all off-diagonal entries of Σ vanish, i.e. the Z_i are independent. Hence for every t,

$$\Pr_\Sigma\{\|Z\|_\infty > t\} \;\le\; \Pr_{\Sigma = \mathrm{Id}}\{\|Z\|_\infty > t\},$$

and integrating over $t \ge 0$ yields

$$\mathbb{E}_\Sigma\big[\|Z\|_\infty\big] \;\le\; \mathbb{E}_{\mathrm{Id}}\big[\|Z\|_\infty\big],$$

as claimed. $\square$

This maximum is actually well understood asymptotically from Extreme Value Theory. We state a short proof here for completeness.

Lemma 2.4. *Let $X_1, \dots, X_n$ be independent $\mathcal{N}(0,1)$ random variables, and set*

$$M_n = \max_{1 \le i \le n} |X_i|.$$

Then as $n \to \infty$,

$$\mathbb{E}[M_n] = \sqrt{2\log(2n)} \; - \; \frac{\log\log(2n) + \log(4\pi)}{2\sqrt{2\log(2n)}} \; + \; \frac{\gamma}{\sqrt{2\log(2n)}} \; + \; o\big((\log n)^{-1/2}\big),$$

where γ is the Euler–Mascheroni constant.

Proof. By definition,

$$\Pr(M_n \le t) = \big[\Pr(|X_1| \le t)\big]^n = \big(2\Phi(t) - 1\big)^n,$$

with Φ the standard-normal CDF. Hence

$$\mathbb{E}[M_n] = \int_0^\infty \Pr(M_n > t)\, dt = \int_0^\infty \big[1 - (2\Phi(t) - 1)^n\big]\, dt.$$

Define $u_n > 0$ by

$$1 - \Phi(u_n) = \frac{1}{2n}.$$

By Mill's ratio,

$$u_n = \sqrt{2\log(2n)} \; - \; \frac{\log\log(2n) + \log(4\pi)}{2\sqrt{2\log(2n)}} \; + \; o\big((\log n)^{-1/2}\big).$$

Set

$$Z_n \; = \; (M_n - u_n)\, u_n.$$

Classical extreme-value theory (Leadbetter *et al.* [10, Ch. 1]) shows

$$Z_n \; \xrightarrow{d} \; G,$$

where G is a standard Gumbel law with $\Pr(G \le x) = \exp(-e^{-x})$ and $\mathbb{E}[G] = \gamma$. Moreover, Hall [11] gives the rate

$$\sup_{x \in \mathbb{R}} \Big| \Pr(Z_n \le x) - \exp(-e^{-x}) \Big| \le O(u_n^{-2}) = O\big((\log n)^{-1}\big).$$

Thus the distributional convergence error is $o(u_n^{-1})$, which is negligible compared to the $O(u_n^{-1})$ bias in the mean.

Since for any sequence $Y_n \xrightarrow{d} G$ with $\mathbb{E}|Y_n| < \infty$, $\mathbb{E}Y_n = \mathbb{E}G + o(1)$, one deduces

$$\mathbb{E}[M_n] = u_n + \frac{\gamma}{u_n} + o\big(u_n^{-1}\big).$$

Substituting the expansion of u_n and expanding in powers of $(\log n)^{-1/2}$ completes the proof. $\square$

Combining this with the comparatively negligible error in the Central Limit Theorem(Given that the matrix is invertible and each entry is bounded, both of which are typical conditions), we see that the second order approximation for the Gaussian absolute maximum is the correct o(1) term, resolving another question from Steinerberger's original paper.

Now that we have all the tools we state our construction and provide an empirical comparison with the random matrices.

Definition 3 (Normalized Random Sign Matrix [5])**.** *Let $n \in \mathbb{N}$. A* normalized random sign matrix *of dimension n is an $n \times n$ real matrix S constructed as follows:*

- *For each $1 \le i \le n$, draw independently a vector $s_i \in \{-1, 1\}^n$ with each coordinate having equal probability $1/2$.*
- *Form the row–stacked matrix $S_0 = [\, s_1^\top;\ s_2^\top;\ \dots;\ s_n^\top \,] \in \{-1, 1\}^{n \times n}$.*
- *Normalize each row to unit Euclidean length:*

$$S_{i,*} = \frac{s_i^\top}{\|s_i\|_2} = \frac{s_i^\top}{\sqrt{n}}, \qquad i = 1, \dots, n.$$

The resulting matrix $S \in \mathbb{R}^{n \times n}$ has independent, isotropic rows satisfying $\|S_{i,*}\|_2 = 1$ and $S_{ij} \in \{\pm 1/\sqrt{n}\}$ for all i, j.

Lemma 2.5. *Let $S \in \mathbb{R}^{n \times n}$ be a normalized random sign matrix. Then there exists an absolute constant C such that*

$$\Pr\Big\{\max_{i \ne j} \big|\langle S_{i,*}, S_{j,*}\rangle\big| > \sqrt{\frac{C \log n}{n}}\Big\} \;\le\; o(1),$$

Proof. For any fixed $i < j$, write

$$\langle S_{i,*}, S_{j,*}\rangle = \frac{1}{n} \sum_{k=1}^{n} X_k,$$

where $X_k = s_{ik} s_{jk}$ are independent Rademacher variables. By Hoeffding's inequality,

$$\Pr\big\{\big|\tfrac{1}{n} \sum_{k=1}^{n} X_k\big| > t\big\} \le 2 \exp\!\big(-\tfrac{nt^2}{2}\big).$$

Taking the union bound over the $\binom{n}{2} < n^2/2$ pairs (i, j) gives

$$\Pr\Big\{\max_{i<j} |\langle S_{i,*}, S_{j,*}\rangle| > t\Big\} \le n^2 \exp\!\big(-\tfrac{nt^2}{2}\big).$$

Choosing $t = \sqrt{\frac{2c \log n}{n}}$ for $c > 2$ yields

$$n^2 \exp\!\big(-\tfrac{n\,(2c \log n)/n}{2}\big) = n^2\, n^{-c} = n^{2-c},$$

and the result follows. $\square$

Definition 4 (Orthonormal Almost–Hadamard Matrix)**.** *For a fixed integer n let $m \ge n$ be the smallest integer such that Hadmard matrix of order m exists, then an $n \times n$ matrix $Q \in \mathbb{R}^{n \times n}$ is called an* orthonormal almost–Hadamard matrix *if:*

(1) $H \in \{\pm 1\}^{m\times m}$ *is a Hadamard matrix with* $HH^\top = mI_m$*.*

(2) $U = H_{[1:n,\,1:n]}/\sqrt{m} \in \mathbb{R}^{n\times n}$ *is the top–left* $n\times n$ *block of* H*.*

(3) Q *is obtained by the QR-factorization* $U = Q\,R$*, so that* $QQ^\top = I_n$*.*

The reason we need orthonormality is so the high-dimensional Central Limit Theorem has us converge to the standard normal. The reason we chose a Hadamard Matrix to start is that to optimize the rate in the High-Dimensional Central Limit Theorem, we need strong moment bounds on individual entries of the matrix, which are best facilitated by uniform matrix entries.

Lemma 2.6 (Flatness of Truncated Hadamard under Hadamard's Conjecture)**.** *Assume Hadamard's conjecture*[1] *holds, i.e. that for every positive multiple of 4 there exists a Hadamard matrix of that order. Fix* $n \in \mathbb{N}$*. Let* Q *be an* $n\times n$ *Orthonormal Almost–Hadamard Matrix then*

$$\left|Q_{ij}\right| = \frac{1+O(1/n)}{\sqrt{n}}, \qquad 1 \le i, j \le n.$$

Proof. Let $u_1, \ldots, u_n \in \mathbb{R}^n$ be the columns of U. Since $HH^\top = mI_m$, for $i \ne j$ we have

$$\langle u_i, u_j\rangle = \frac{1}{m}\sum_{r=1}^{n} H_{ri}H_{rj} = -\frac{1}{m}\sum_{r=n+1}^{m} H_{ri}H_{rj},$$

and $m-n < 4$ implies $\langle u_i, u_j\rangle = O(n^{-1})$. Moreover $\|u_i\|^2 = n/m$. Hence the Gram matrix satisfies

$$U^\top U \;=\; \frac{n}{m}I_n + E, \qquad \|E\|_{\max} = O(n^{-1}).$$

Since $U = Q\,R$ and $R^\top R = U^\top U$, the diagonal entries obey

$$R_{jj}^2 = \frac{n}{m} + E_{jj} = \frac{n}{m}\big(1 + O(n^{-1})\big),$$

so

$$R_{jj} = \sqrt{\tfrac{n}{m}}\,(1 + O(n^{-1})) = \Theta(1).$$

We now prove by induction on i that for each $1 \le i < j \le n$, $R_{ij} = O(n^{-1})$.

- Base case ($i = 1$): The Cholesky/QR recurrence gives

$$R_{1j} = \frac{(R^\top R)_{1j}}{R_{11}} = \frac{E_{1j}}{R_{11}} = \frac{O(n^{-1})}{\Theta(1)} = O(n^{-1}).$$

- Inductive step: Suppose for some $i \ge 2$ that $R_{\ell j} = O(n^{-1})$ for all $\ell < i$ and $j > \ell$. Then for each $j > i$,

$$R_{ij} = \frac{1}{R_{ii}}\Big((R^\top R)_{ij} - \sum_{k=1}^{i-1} R_{k\,i}\,R_{k\,j}\Big).$$

  Here $(R^\top R)_{ij} = E_{ij} = O(n^{-1})$, and each product $R_{k\,i}R_{k\,j} = O(n^{-2})$ by the inductive hypothesis, so the sum over $k = 1, \ldots, i-1$ is $O(n^{-1})$. Since $R_{ii} = \Theta(1)$, it follows that $R_{ij} = O(n^{-1})$.

Thus

$$R = \sqrt{\tfrac{n}{m}}\, I_n + F, \qquad \|F\|_{\max} = O(n^{-1}).$$

Inverting R gives an upper-triangular R^{-1} with $(R^{-1})_{jj} = \sqrt{m/n}\,(1+O(n^{-1}))$ and $(R^{-1})_{ij} = O(n^{-1})$ for $i < j$. Finally,

$$Q = U\,R^{-1}, \quad Q_{ij} = \sum_{k=1}^{n} \frac{H_{ik}}{\sqrt{m}}\,(R^{-1})_{kj} = \frac{1}{\sqrt{m}}\Big(\sqrt{\tfrac{m}{n}} + O(n^{-1})\Big) = \frac{1+O(n^{-1})}{\sqrt{n}},$$

as claimed. □

Theorem 2.4 (Asymptotics for explicit constructions)**.** *Let $n \geq 3$. For both the normalized random sign matrix S (rows i.i.d. uniform in $\{\pm 1\}^n$ and scaled by $1/\sqrt{n}$) and for any orthonormal almost–Hadamard matrix Q (top–left $n \times n$ block of an $m \times m$ Hadamard matrix, then orthonormalized), one has the expansion*

$$\begin{aligned} \beta(\,\cdot\,) = {} & \sqrt{2\log(2n)} \\ & - \frac{\log\log(2n)}{2\sqrt{2\log(2n)}} \\ & + o\big(\log n^{-\frac{1}{2}}\big). \end{aligned} \tag{2}$$

In particular both constructions attain the correct leading order $\sqrt{2\log(2n)}$, and the error from the Gaussian extreme–value prediction is at most $o\big(n^{-1/2}(\log n)^7\big)$.

Proof. We now apply Theorem 2.2 to compute $\beta(\hat{A})$ for the two constructions described above. Throughout, we fix the matrix A and take the randomness only over the Rademacher vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$.
General covariance structure. For any fixed $n \times n$ matrix A whose rows have ℓ_2–norm $\sqrt{n}$, define

$$S_n \;=\; \frac{1}{\sqrt{n}}\,A\,\varepsilon, \qquad (S_n)_j \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_{ji}\,\varepsilon_i.$$

The covariance matrix of S_n is

$$\Sigma \;=\; \mathrm{Cov}(S_n) \;=\; \frac{1}{n}\,A\,A^{\top},$$

so that $\Sigma_{jj} = 1$ for all j, and for $j \neq k$,

$$\Sigma_{jk} \;=\; \frac{1}{n}\sum_{i=1}^{n} A_{ji}A_{ki} \;=\; \big\langle \hat{A}_{j,*},\, \hat{A}_{k,*} \big\rangle,$$

where $\hat{A}$ is A with each row normalized to unit length. The theorem thus applies with this covariance.

We now treat the two constructions separately.

1. Normalized Random Sign Matrix. Let $S \in \mathbb{R}^{n\times n}$ be a normalized random sign matrix. By Lemma 2.5, with probability $1-o(1)$,

$$\max_{j\neq k}\bigl|\Sigma_{jk}\bigr| \;=\; \max_{j\neq k}\bigl|\langle \hat{S}_{j,*}, \hat{S}_{k,*}\rangle\bigr| \;=\; O\Bigl(\sqrt{\tfrac{\log n}{n}}\Bigr).$$

Consequently,

$$\det(\Sigma_{j,k}) \;=\; 1-\Sigma_{jk}^2 \;\geq\; 1-O\Bigl(\tfrac{\log n}{n}\Bigr) \;=\; 1-o(1), \qquad \frac{\det(\Sigma_{j,k,\ell})}{\det(\Sigma_{j,k})} \;\geq\; 1-o(1),$$

so that we may take $\alpha^2 = 1-o(1)$ and $\beta^2 = 1-o(1)$. Moreover, since every entry of X_{ij} satisfies $|X_{ij}| = 1/\sqrt{n}$, we may fix any constant $B>0$ (e.g. $B=1$) for all sufficiently large n to ensure

$$\mathbb{E}\bigl[e^{|X_{ij}|/B}\bigr] \;\leq\; 2.$$

Applying Theorem 2.2, we therefore obtain

$$\sup_{a,b\in\mathbb{R}^n}\Bigl|P(a\leq S_n\leq b)-P(a\leq Z\leq b)\Bigr| \;\leq\; \frac{C\,B^3}{\alpha^2\beta^2\sqrt{n}}\,(\log n)^{6.5} \;=\; O\bigl(n^{-1/2}(\log n)^{6.5}\bigr),$$

where $Z\sim N(0,\Sigma)$.

(3)

$$\begin{aligned}\mathbb{E}\bigl[\|S_n\|_\infty\bigr] = \int_0^\infty \Pr\bigl(\|S_n\|_\infty > t\bigr)\,dt - \int_0^q \bigl[\Pr(\|Z\|_\infty > t) - \Pr(\|S_n\|_\infty > t)\bigr]\,dt \\ - \int_q^\infty \Pr\bigl(\|Z\|_\infty > t\bigr)\,dt + \int_q^\infty \Pr\bigl(\|S_n\|_\infty > t\bigr)\,dt\end{aligned}$$

$$\begin{aligned}\mathbb{E}\bigl[\|S_n\|_\infty\bigr] \leq \mathbb{E}\bigl[\|Z\|_\infty\bigr] + q\,\sup_t\Bigl|\Pr(\|Z\|_\infty > t) - \Pr(\|S_n\|_\infty > t)\Bigr| \\ - \int_q^\infty \Pr\bigl(\|Z\|_\infty > t\bigr)\,dt + \int_q^\infty \Pr\bigl(\|S_n\|_\infty > t\bigr)\,dt\end{aligned} \tag{4}$$

So we have

$$\mathbb{E}\bigl\|S_n\bigr\|_\infty \;\leq\; \mathbb{E}\bigl\|Z\bigr\|_\infty \;+\; q\,\delta \;-\; \int_q^\infty P\bigl(\|Z\|_\infty > t\bigr)\,dt \;+\; \int_q^\infty P\bigl(\|S_n\|_\infty > t\bigr)\,dt,$$

where

$$\delta \;=\; \sup_{t\in\mathbb{R}}\Bigl|P(\|Z\|_\infty > t) - P(\|S_n\|_\infty > t)\Bigr| \;=\; O\bigl(n^{-1/2}(\log n)^{6.5}\bigr).$$

We now choose

$$q \;=\; \sqrt{C\,\log n}, \qquad C>3.$$

and bound the individual pieces

(1) **Sup–error term.**

$$q\,\delta \;=\; \sqrt{C\log n}\;\cdot\;O\bigl(n^{-1/2}(\log n)^{6.5}\bigr) \;=\; O\bigl(n^{-1/2}(\log n)^{7}\bigr).$$

(2) **Gaussian-tail integral.** Since

$$P\big(\|Z\|_\infty > t\big) \;=\; P\Big(\bigcup_{i=1}^{n}\{|Z_i| > t\}\Big) \;\le\; \sum_{i=1}^{n} P\big(|Z_i| > t\big) \;=\; 2n\big[1-\Phi(t)\big],$$

we have

$$\int_q^\infty P\big(\|Z\|_\infty > t\big)\,dt \;\le\; \int_q^\infty \frac{2n}{\sqrt{2\pi}}\,\frac{e^{-t^2/2}}{t}\,dt.$$

Observe that for $t \ge q$, $\frac{1}{t} \le \frac{1}{q}$, so

$$\int_q^\infty \frac{e^{-t^2/2}}{t}\,dt \le \frac{1}{q}\int_q^\infty e^{-t^2/2}\,dt = \frac{1}{q}\int_q^\infty t\,e^{-t^2/2}\,\frac{dt}{t} \le \frac{1}{q}\int_q^\infty t\,e^{-t^2/2}\,dt = \frac{e^{-q^2/2}}{q}.$$

Hence

$$\int_q^\infty P(\|Z\|_\infty > t)\,dt \;\le\; \frac{2n}{\sqrt{2\pi}}\,\frac{e^{-q^2/2}}{q} = O\Big(\frac{n}{q}e^{-q^2/2}\Big).$$

Substituting $q = \sqrt{C\log n}$ gives

$$\frac{n}{q}e^{-q^2/2} = \frac{n}{\sqrt{C\log n}}\,n^{-C/2} = n^{1-C/2}(\log n)^{-1/2},$$

which—for any fixed $C > 3$—is $O(n^{-\varepsilon})$ with $\varepsilon = C > 0.5$.

(3) **Sign-matrix tail integral.**

By Hoeffding's inequality, each coordinate $S_{n,i}$ satisfies

$$P\big(|S_{n,i}| > t\big) \;\le\; 2\exp\big(-nt^2/2\big),$$

so

$$P\big(\|S_n\|_\infty > t\big) \;\le\; 2n\exp\big(-nt^2/2\big).$$

Hence

$$\int_q^\infty P\big(\|S_n\|_\infty > t\big)\,dt \le 2n\int_q^\infty e^{-nt^2/2}\,dt \;\le\; 2n\,\frac{e^{-nq^2/2}}{nq} = \frac{2}{q}\,e^{-nq^2/2}$$
$$= O\Big(q^{-1}\,e^{-\frac{nC\log n}{2}}\Big) = O\big(n^{-nC/2}\big)$$

Combining these three bounds, we obtain

$$\mathbb{E}\big\|S_n\big\|_\infty = \mathbb{E}\big\|Z\big\|_\infty \;+\; O\big(n^{-1/2}(\log n)^7\big).$$

Finally, by Lemma 2.4,

$$\mathbb{E}\big\|Z\big\|_\infty < \sqrt{2\log(2n)} \;-\; \frac{\log\log(2n)+\log(4\pi)}{2\sqrt{2\log(2n)}} \;+\; \frac{\gamma}{\sqrt{2\log(2n)}} \;+\; o\big((\log n)^{-1/2}\big),$$

so the same expansion holds for $\mathbb{E}\|S_n\|_\infty$, up to the additional $O\big(n^{-1/2}(\log n)^7\big)$–error.

Hence

$$\boxed{\beta(S) \;=\; \mathbb{E}\|S_n\|_\infty \;\le\; \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}} + O\big(n^{-1/2}(\log n)^7\big).} \tag{5}$$

2. Orthogonal Almost–Hadamard Matrix. Let $Q \in \mathbb{R}^{n\times n}$ be an orthogonal almost–Hadamard matrix. By definition, $QQ^\top = I_n$. Consequently,

$$\Sigma \;=\; \frac{1}{n}\,Q\,Q^\top \;=\; \frac{1}{n}\,I_n,$$

and upon scaling by $\sqrt{n}$ in the CLT normalization, we have $\Sigma = I_n$. Thus

$$\alpha^2 \;=\; 1, \qquad \beta^2 \;=\; 1, \qquad B \;=\; 1,$$

exactly. Theorem 2.2 therefore gives

$$\sup_{a,b\in\mathbb{R}^n}\Big|P(a \le S_n \le b) \;-\; P(a \le Z \le b)\Big| \;\le\; \frac{C}{\sqrt{n}}(\log n)^{6.5},$$

where now $Z \sim N(0, I_n)$. As above,

$$\mathbb{E}\|Z\|_\infty \;=\; \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}} + o\big((\log n)^{-1/2}\big),$$

Continuing from the decomposition and the choice

$$q = \sqrt{C\log n}, \qquad C > 3,$$

the sup-error term and the Gaussian-tail integral remain unchanged in order. It remains only to bound

$$\int_q^\infty P\big(\|S_n\|_\infty > t\big)\,dt.$$

By Lemma 2.6, each entry of Q satisfies $|Q_{ji}| = (1 + O(1/n))/\sqrt{n}$. Hence

$$(S_n)_j = \frac{1}{\sqrt{n}}\sum_{i=1}^n Q_{ji}\,\varepsilon_i$$

is a sum of n independent, mean-zero terms each bounded by $(1 + O(1/n))/n$. By Hoeffding's inequality,

$$P\big(|(S_n)_j| > t\big) \;\le\; 2\exp\Big(-\frac{2\,t^2}{\sum_{i=1}^n (Q_{ji}/\sqrt{n})^2}\Big) \;=\; 2\exp\big(-2t^2 + o(1)\big).$$

A union bound over $j = 1, \dots, n$ then yields

$$P\big(\|S_n\|_\infty > t\big) \;\le\; 2n\,e^{-2t^2+o(1)}.$$

Therefore

$$\int_q^\infty P(\|S_n\|_\infty > t)\,dt \;\le\; 2n\int_q^\infty e^{-2t^2+o(1)}\,dt \;=\; O\Big(\frac{n}{q}e^{-2q^2}\Big) \;=\; O\big(n^{1-2C}(\log n)^{-1/2}\big)$$

This completes the estimate of the matrix tail integral, which is still negligible compared to the CLT error.

and so

$$\boxed{\beta(Q) \;=\; \mathbb{E}\|Q\,\|_\infty \;=\; \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}} + O\big(n^{-1/2}(\log n)^7\big).} \tag{6}$$

Conclusion. Combining (5) and (6), we conclude that for both constructions,

$$\boxed{\beta(\hat{A}) \;=\; \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}} + O\big(n^{-1/2}(\log n)^7\big).}$$

In fact we have the chain of inequalities:

$$\boxed{\beta(S) \;<\; \mathbb{E}\big\|Z\big\|_\infty + O\big(n^{-1/2}(\log n)^7\big) \;\le\; \beta(Q) + O\big(n^{-1/2}(\log n)^7\big).}$$

So we see that they are within some $\frac{\mathrm{polylog}(n)}{\sqrt{n}}$ factor of each other. Although from the covariance matrix heuristic, we can see that the Orthonormal Almost-Hadamard Matrix should do better, there is no way to explicitly prove this without some stronger results or conditions that we are unaware of. $\square$

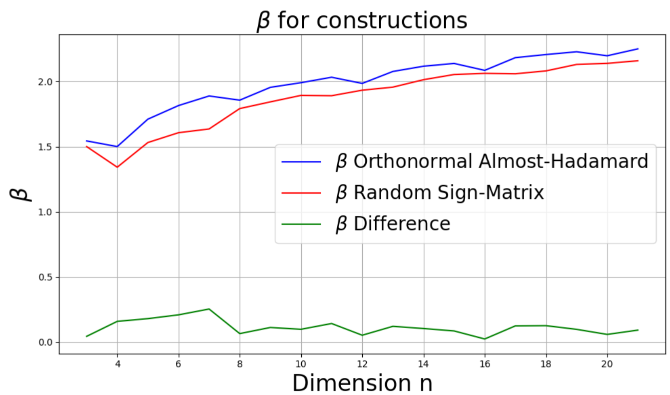


FIGURE 4. Running exact simulations of the expectation, which is computationally inviable for n>20, show a strict gap between the deterministic and random constructions.

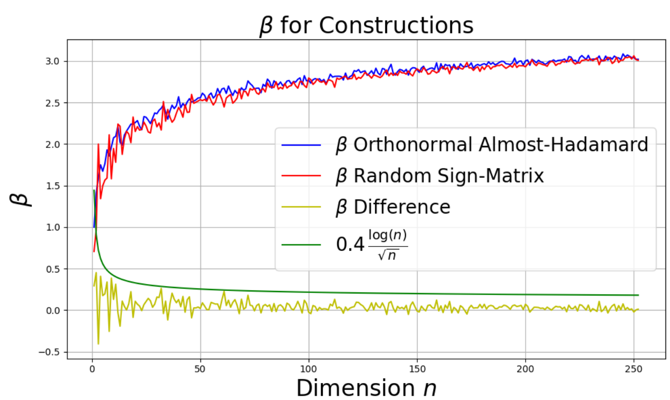


FIGURE 5. Running Monte-Carlo simulations of the expectation, for larger values of n, show a gap visually matching our asymptotic bound.

References

[1] J. Hadamard, *Résolution d'une question relative aux déterminants*, Bull. Sci. Math. (2) **17**, 240–246 (1893).

[2] R. O'Donnell, Analysis of Boolean Functions, Cambridge University Press, 2014.

[3] Alex Albors, Hisham Bhatti, Lukshya Ganjoo, Raymond Guo, Dmitriy Kunisky, Rohan Mukherjee, Alicia Stepin, Tony Zeng, On the Structure of Bad Science Matrices, arXiv:2408.00933

[4] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[5] S. Steinerberger, Bad science matrices, arXiv:2402.03205

[6] R. O'Donnell, Analysis of Boolean Functions, arXiv:2105.10386

[7] Michel Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996. https://doi.org/10.1007/BF01844850.

[8] Xiao Fang, Yuta Koike, Song-Hao Liu, and Yi-Kun Zhao. High-dimensional central limit theorems by Stein's method in the degenerate case. *arXiv preprint arXiv:2305.17365*, May 2023.

[9] L. D. Pitt, A Gaussian correlation inequality for symmetric convex sets, *Ann. Probab.*, 10(4):629–637, 1982.

[10] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer, 1983.

[11] P. Hall. On the rate of convergence of normal extremes. *J. Appl. Probab.*, 16(2):433–439, 1979.

Email address: `ssinha19@uw.edu`

University of Washington, Seattle, WA 98195, USA