The block complexity conjecture

Marcel K. Goh

29 April 2026

1. Cohen's idempotent theorem

We begin with some background on Cohen's idempotent theorem [*American Journal of Mathematics* **82** (1960), 191–212]. Let G be a locally compact abelian group, and let $M(G)$ be the set of all bounded measures on G (non-negative, countably additive set functions μ on G with $\mu(\emptyset) = 0$). We can view $M(G)$ as a topological semigroup under the convolution operation

$$\mu * \nu(A) = \int_G \int_G \mathbf{1}_A(x+y)\, d\mu(x)\, d\nu(y),$$

where $\mathbf{1}_A$ is the characteristic function of the set $A \subseteq G$. An *idempotent* in this algebra is a measure μ with $\mu * \mu = \mu$; in the middle of the 20th century, there was a programme surrounding the problem of characterising the idempotent elements of $M(G)$.

Something we can see right away is that for any measure μ with $\mu * \mu = \mu$, the Fourier–Stieltjes transform $\widehat{\mu}$ given by

$$\widehat{\mu}(\chi) = \int_G \overline{\chi(x)}\, d\mu(x)$$

satisfies $\widehat{\mu}^2 = \widehat{\mu}$ by the convolution law and must hence be boolean. We are then left with the task of determining which boolean functions on $\widehat{G}$ (or equivalently, which subsets of $\widehat{G}$) are the Fourier–Stieltjes transforms of idempotent measures on G. A starting point is the following.

Proposition 1.1. *Let $E = \psi\Gamma$ be a coset of the dual group $\widehat{G}$ of a locally compact abelian group G. Then $\mathbf{1}_E = \widehat{\mu}$ for some idempotent measure μ on G with $\|\mu\| = 1$.*

Proof. Let H be the annihilator subgroup

$$H = \{x \in G : \chi(x) = 1 \text{ for all } \chi \in \Gamma\}.$$

It can be shown that H is compact, since Γ is open (hence $\Gamma/\Lambda = \widehat{H}$ is discrete), and we can normalise the Haar measure m_H on H so that $m_H(H) = 1$, which yields a measure (which we also call m_H) on G as well. Define μ by $d\mu(x) = \psi(x)\, dm_H(x)$.

Note that if $\overline{\psi}\chi \notin \Gamma$, then there is some $y \in H$ with $\overline{\psi(y)}\chi(y) \neq 1$, and we have

$$\begin{aligned}\widehat{\mu}(\chi) &= \int_G \overline{\chi(x)}\psi(x)\, dm_H(x)\\ &= \overline{\chi(y)}\psi(y) \int_H \overline{\chi(x-y)}\psi(x-y)\, dm_H(x)\\ &= \chi(y)\widehat{\mu}(\chi)\end{aligned}$$

for the Fourier–Stieltjes transform $\widehat{\mu}$ of μ, by the invariance of Haar measure under translation. This gives the orthogonality relation

$$\widehat{\mu}(\chi) = \begin{cases} 1, & \text{if } \chi \in \psi\Gamma;\\ 0, & \text{otherwise}\end{cases}.$$

In other words, $\widehat{\mu} = \mathbf{1}_{\psi\Gamma}$, and μ is idempotent by the convolution law. Lastly, note that

$$\|\mu\| = |\mu|(G) = \int_G \left|\overline{\chi(x)}\right| dm_H(x) = m_H(G) = 1. \quad ▮$$

Let Ω be the set of all subsets of $\widehat{G}$ whose indicator functions are Fourier–Stieltjes transforms of idempotent measures on G.

Proposition 1.2. *The set Ω contains the ring of cosets of $\widehat{G}$; that is, the smallest collection that contains all cosets of $\widehat{G}$ and is closed under finite intersections, unions, and complements.*

Proof. If E and F are cosets of $\widehat{G}$ with $\mathbf{1}_E = \widehat{\mu}$ and $\mathbf{1}_F = \widehat{\nu}$ for some idempotent measures μ and ν on G, then we have $\widehat{\mu * \nu} = \mathbf{1}_{E\cap F}$ by the convolution law. Then since the Fourier–Stieltjes transform is linear, we see that

$$\widehat{\mu \vee \nu} = \widehat{\mu} + \widehat{\nu} - \widehat{\mu}\cdot\widehat{\nu} = \mathbf{1}_{E\cup F},$$

where $\mu \vee \nu = \mu + \nu - \mu * \nu$. Finally, letting δ_0 be the Dirac measure that assigns 1 to precisely those sets containing 0, we have

$$\widehat{\delta_0}(\chi) = \int_G \overline{\chi(x)}\, d\delta_0(x) = \overline{\chi(0)} = 1$$

for all $\chi \in \widehat{G}$, so

$$\widehat{\delta_0 - \mu} = 1 - \mathbf{1}_E = \mathbf{1}_{\widehat{G}\setminus E}.$$

The full proposition now follows by induction. ▮

Cohen's idempotent theorem states that Ω is in fact equal to the ring of cosets of $\widehat{G}$.

Theorem 1.3 (*Cohen,* 1960). *A measure μ is idempotent if and only if $\widehat{\mu}$ belongs to the coset ring of $\widehat{G}$.*

2. The Green–Sanders theorem

Another way of stating (the difficult direction of) Cohen's idempotent theorem is that the Fourier–Stieltjes transform $\widehat{\mu}$ of an idempotent measure μ on a locally compact abelian group G can be written as a finite signed sum

$$\widehat{\mu} = \sum_{i=1}^{L} \pm \mathbf{1}_{E_i},$$

where each E_i is a coset of $\widehat{G}$. When G is finite, this statement has no content, since *any* function on $\widehat{G}$ can be expressed as a finite sum in that manner.

Hence we want to develop a quantitative version of Cohen's idempotent theorem, that bounds the number L of summands in terms of some parameter of μ. The natural parameter turns out to be its norm $\|\mu\|$. To see this, note that the *contractive* idempotent measures (that is, those with norm at most 1) are exactly the idempotents μ with $\widehat{\mu} = \mathbf{1}_E$ for some open coset E of $\widehat{G}$. (We saw one direction of this earlier, and the converse is not terribly difficult, but we omit it in this exposition.)

The quantitative idempotent theorem of Green and Sanders [*Annals of Mathematics* **168** (2008), 1025–1054] tells us that any idempotent measure can be written as a signed sum of contractive idempotents. It turns out to be possible to prove this theorem by first doing so for finite abelian groups; we shall state the theorem in the latter case. When writing out the idempotent theorem for finite groups it is customary to exchange the roles of G and $\widehat{G}$. Hence one writes $f = \widehat{\mu}$ and assumes that f is boolean (which is equivalent to the idempotence of μ). Then the norm $\|\mu\|$ is exactly the l_1 norm of the Fourier transform of f. This is sometimes denoted $\|f\|_A$ and called the *algebra norm* of f, because for any $f, g : G \to \mathbf{C}$ one has

$$\|fg\|_A \le \|f\|_A \cdot \|g\|_A .$$

In any case we have dawdled enough. Let us now state the finitary form of the Green–Sanders quantitative idempotent theorem.

Theorem 2.1 (*Green–Sanders,* 2008). *Let G be a finite abelian group and let $f : G \to \{0, 1\}$ have $\|f\|_A \le M$. Then we may write*

$$f = \sum_{i=1}^{L} \pm \mathbf{1}_{a_i + H_i}, \tag{1}$$

where each $a_i + H_i$ *is a coset of* G *and* $L \leq \exp(\exp(CM^4))$ *for some absolute constant* C*.*

Of course, if f can be expressed as a signed sum (1), then $\|f\|_A \leq L$ by the triangle inequality. The Green–Sanders theorem is a (quantitatively weaker) converse to this.

3. Schur multipliers

We now shift gears entirely and introduce a different problem in matrix analysis, which at first seems only related to the idempotent theorem by analogy. Consider the Hilbert space l_2 of square-summable complex sequences, and let $B(l_2)$ be the space of bounded linear operators on l_2, endowed with the operator norm

$$\|T\|_{\mathrm{op}} = \sup_{y \neq 0} \frac{\|Ty\|_2}{\|y\|_2}.$$

Each operator T in $B(l_2)$ uniquely determines an associated matrix $(t_{i,j})_{i,j \in \mathbf{N}}$, where $t_{i,j} = \langle Te_i, e_j \rangle$ for the standard orthonormal basis $\{e_i\}_{i \in \mathbf{N}}$ of l_2, and vice versa.

So every infinite matrix $T : \mathbf{N} \times \mathbf{N} \to \mathbf{C}$ represents an element of $B(l_2)$, but we can also view infinite matrices $A : \mathbf{N} \times \mathbf{N} \to \mathbf{C}$ as operators on $B(l_2)$, via the mapping $T \mapsto A \circ T$, where $\circ$ denotes the Schur (entrywise) product. The matrix A is called a *Schur multiplier* if $A \circ T \in B(l_2)$ for all $T \in B(l_2)$, or equivalently, if its *Schur multiplier norm*, defined by

$$\|A\|_{\mathrm{m}} = \sup_{T \neq 0} \frac{\|A \circ T\|_{\mathrm{op}}}{\|T\|_{\mathrm{op}}},$$

is finite. The set of Schur multipliers is closed under addition, and the inequality

$$\|A \circ B\|_{\mathrm{m}} \leq \|A\|_{\mathrm{m}} \cdot \|B\|_{\mathrm{m}},$$

proved by Schur [*Crelle* **140** (1911), 1–28], verifies that it is closed under Schur product. Hence this set with these two operations forms a Banach algebra.

As before, we might like to characterise the idempotents in this algebra. Any matrix A that satisfies $A \circ A = A$ must be boolean, but not all infinite boolean matrices are Schur multipliers. One example is the infinite upper triangular matrix (the normalised trace norm of an $n \times n$ upper triangular matrix is roughly $\log n$, and the Schur multiplier norm is bounded from below by the normalised trace norm). Hence the question can be restricted to: *Which boolean matrices have finite Schur multiplier norm?*

The simplest idempotent Schur multipliers are the ones with norm 1 (since any idempotent element must have norm at least 1). This leads us to define the following class of matrices. A matrix $B : X \times Y \to \{0, 1\}$ is called *blocky* if there exist pairwise disjoint subsets S_i of X and pairwise disjoint subsets T_i of Y such that the support of B is exactly

$$\bigcup_{i=1}^{\infty} S_i \times T_i.$$

Simple examples of blocky matrices are the zero matrix, $m \times n$ all-ones matrices, and identity matrices. The following proposition of Livshits [*Linear Algebra and its Applications* **222** (1995), 15–22] shows that blocky matrices are precisely the contractive idempotent Schur multipliers.

Proposition 3.1. *A nonzero boolean matrix satisfies* $\|A\|_{\mathrm{m}} = 1$ *if and only if* A *is blocky.*

It is an open problem, dating back at least to a 2003 preprint of Katavolos and Paulsen (the journal version is [*Canadian Mathematical Bulletin* **48** (2005), 97–111]), whether any idempotent Schur multiplier can be written as a finite sum of contractive idempotents. A compactness argument of Hambardzumyan, Hatami, and Hatami [*Israel Journal of Mathematics* **253** (2023), 555–616] shows that a positive resolution to this problem, while only being meaningful for infinite matrices, is equivalent to the following structural conjecture concerning finite matrices with bounded Schur multiplier norm.

Conjecture 3.2 (*Hambardzumyan–Hatami–Hatami,* 2023). *Suppose that* A *is a finite boolean matrix with* $\|A\|_{\mathrm{m}} \le \gamma$. *Then we may express* A *as the signed sum*

$$A = \sum_{i=1}^{L} \pm B_i,$$

where each B_i *is a blocky matrix and* L *depends only on* γ.

Given a finite abelian group G and a boolean function $f : G \to \{0,1\}$, we may define the matrix $A_f : G \times G \to \{0,1\}$ by letting $A_f(x,y) = f(x-y)$. It can be shown that $\|M_f\|_{\mathrm{m}} = \|f\|_A$, so the Green–Sanders theorem verifies Conjecture 3.2 for this special case of "convolution" matrices.

To simplify the discussion below, we define the *block complexity* $\mathrm{block}(A)$ of a matrix A (it turns out to be just as easy to allow this matrix to take integer values rather than boolean ones) to be the smallest integer L such that there exist blocky matrices $B_1, \ldots, b_L$ and signs $\sigma_1, \ldots, \sigma_L$ such that $A = \sum_{i=1}^{L} \sigma_i B_i$. It is immediate from the definition (and the triangle inequality) that $\|A\|_{\mathrm{m}} \le \mathrm{block}(A)$. Conjecture 3.2 claims that conversely, the block complexity can be bounded by a function of the Schur multiplier norm.

4. Consequences of the block complexity conjecture

In this section we note two consequences that follow from Conjecture 3.2. The first one, regarding the ranges of Schur idempotents, we shall record only very briefly, as even giving a full definition of the term “hyperreflexive” is outside the scope of these notes. In any case, a paper of Eleftherakis, Levene, and Todorov [*Israel Journal of Mathematics* **215** (2016), 317–337] shows that the range $A\big(B(l_2)\big)$ is hyperreflexive whenever the Schur multiplier A is a finite sum of contractive idempotents, and asks whether all Schur idempotents have hyperreflexive ranges. Conjecture 3.2 would imply a positive answer to this question.

The second consequence is in the realm of communication complexity, and this we shall describe more fully. Let $f(x, y)$ be a boolean function taking two n-bit strings as input. Suppose that Alice is in possession of the bitstring x and Bob is in possession of the bitstring y. The *communication complexity* $D(f)$ of the function f is the minimum number of bits that Alice and Bob need to exchange in order for either of them to work out the value of the function $f(x, y)$.

A simple example of a function with low communication complexity is the parity function p, which asks whether there are odd number of 1s in the bitstring $x \oplus y$, where $\oplus$ denotes vector addition modulo 2. We have $D(p) = 1$, since Alice just need to send Bob a 1 if the number of bits in x is odd, and 0 otherwise. Then Bob computes the same with y and adds the two bits together (modulo 2) to obtain $p(x, y)$.

But there are very simple functions with very high communication complexity as well. Take for instance the equality function $e(x, y) = \mathbf{1}_{[x=y]}$. It is not hard to see that $e(x, y) = n$. There is another communication complexity model that in some sense “defines” this example to be simple. We suppose that Alice and Bob have access to $e(x, y)$ (of up to n bits at a time) as an oracle. The *communication complexity with access to an equality oracle* of a boolean function f, denoted $D^{\mathrm{EQ}}(f)$, is the number of calls to this oracle required to compute f. Since Alice and Bob can use this oracle to communicate individual bits as they were able to before (for example, Alice sends the pertinent bit of x, and Bob sends the bit 1), it is always true that $D^{\mathrm{EQ}}(f) \leq D(f)$. We also now have $D^{\mathrm{EQ}}(e) = 1$.

One can express any function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ as a boolean matrix A and by an abuse of notation write block(f) for the block complexity of this boolean matrix (we might similarly write $D^{\mathrm{EQ}}(A)$). Under this representation, the function $e(x, y)$ is simply the $n \times n$ identity matrix; hence block$(e) = 1$. It turns out that all blocky matrices B have $D^{\mathrm{EQ}}(B) = 1$. (If x belongs to the set $S_i \subseteq \{0, 1\}^n$, then Alice can send $\mathbf{1}_{T_i}$, and Bob sends $\mathbf{1}_{T_j}$ for the set $T_j \subseteq \{0, 1\}^n$ containing y. These two are equal if and only if the (x, y) entry of the blocky matrix is 1.) Hence $D^{\mathrm{EQ}}(f) \leq \mathrm{block}(A_f)$. It was also shown by Hambardzumyan, Hatami, and Hatami that

$$\frac{1}{2} \log \mathrm{block}(A) \leq D^{\mathrm{EQ}}(A),$$

so these two notions are qualitatively equivalent.

Conjecture 3.2 would imply that $D^{\mathrm{EQ}}(A)$ is constant whenever $\|A\|_{\mathrm{m}}$ is.

5. A polylogarithmic upper bound

A recent paper of Hatami and the author [to appear in *International Mathematics Research Notices*] gives a polylogarithmic upper bound on the block complexity of a matrix in terms of its Schur multiplier norm.

Theorem 5.1 (*G.–Hatami,* 2025). *Let A be an integer $n \times n$ matrix with $\|A\|_{\mathrm{m}} \le \gamma$. Then*

$$\mathrm{block}(A) \le 2^{O(\gamma^7)}(\log n)^2.$$

While not a strong enough result for either of the applications discussed in the previous section, this already provides a case for matrices with small Schur multiplier norm having atypical block complexity, as evidenced by this probabilistic result in the other direction, due to Avraham and Yehudayoff [*Computational Complexity* **33** (2024), 1–18].

Theorem 5.2 (*Avraham–Yehudayoff,* 2024). *Let A be an $n \times n$ boolean matrix chosen uniformly at random. Then*

$$\mathbf{P}\bigl\{\mathrm{block}(A) \ge n/(4\log(2n))\bigr\} \ge 1 - 2^{-n^2/2}.$$

In other words, with high probability a boolean matrix has block complexity $\Omega(n/\log n)$.

The proof of Theorem 5.1 is inductive and very much mirrors the proof of Green and Sanders in the boolean case $G = \mathbf{F}_2^n$ [*Geometric and Functional Analysis* **18** (2008), 144–162]. We find a suitable matrix $\widehat{A}$ with $\mathrm{block}(\widehat{A}) \ll_\gamma (\log n)^2$, and express $A = \widehat{A} + (A - \widehat{A})$. By showing that $\|A - \widehat{A}\|_{\mathrm{m}}^2$ drops by a constant, we only need to iterate a constant number of times (in terms of γ) before we must have $\|A\|_{\mathrm{m}} < 1$, in which case $A = 0$.

Let us outline the proof in more detail. First, we note that our original definition of the Schur multiplier norm, as a norm for operators on Hilbert spaces, does not admit easy analysis. Instead, we define the *γ_2-norm* of a matrix $A : X \times Y \to \mathbf{C}$ to be

$$\|A\|_{\gamma_2} = \min_{UV=A} \|U\|_{\mathrm{row}} \|V\|_{\mathrm{col}},$$

where $\|U\|_{\mathrm{row}}$ is the maximum l_2 norm of any row in U, and $\|V\|_{\mathrm{col}}$ is the maximum l_2 norm of any column in V. It is clear from this definition that duplicating any row or column does not increase the γ_2-norm, and neither can passing to a submatrix. By an old theorem of Grothendieck [*Annales de linstitut Fourier* **4** (1954), 73–112], the γ_2 norm is in fact equal to the Schur multiplier norm; i.e., $\|A\|_{\gamma_2} = \|A\|_{\mathrm{m}}$. The way we shall use the definition of the γ_2 norm is to express A with $\|A\|_{\mathrm{m}} \le \gamma$ as $A = UV$ with all the rows u_x of U satisfying $\|u_x\|_2 \le 1$ and all the columns v_y of V satisfying $\|v_y\|_2 \le \gamma$. This is called a *γ-factorisation* of A.

Littlestone dimension. We shall also require the notion of the Littlestone dimension of a matrix. A *mistake tree* of depth d over a domain X is a complete binary tree of depth d in which

i) each internal node ν is labelled with an element $x(\nu) \in X$; and
ii) each edge e is labelled with a sign $\sigma(e) \in \{-1, 1\}$, where $\sigma(e) = -1$ indicates a left child and $\sigma(e) = 1$ indicates a right child.

For the purposes of the Littlestone dimension, a sign matrix $A : X \times Y \to \{-1, 1\}$ is said to *shatter* a mistake tree over X if for every root-to-leaf path $(\nu_1, \ldots, \nu_{d+1})$, there exists a column $y \in Y$ such that $A\big(x(\nu_i), y\big) = \sigma(\nu_i \nu_{i+1})$ for all $i \in [d]$. Then the *Littlestone dimension* of a matrix A, denoted by $\mathrm{Ldim}(A)$, is the largest integer d for which there exists a mistake tree of depth d that is shattered by A.

The first proposition we record bounds the Littlestone dimension of a sign matrix in terms of its Schur multiplier norm.

Proposition 5.3. *Every matrix* $A : X \times Y \to \{-1, 1\}$ *satisfies*

$$\mathrm{Ldim}(A) \leq \|A\|_{\mathrm{m}}^2.$$

The property of matrices with small Littlestone dimension that we need is the following. For any sign matrix A with $\mathrm{Ldim}(A) = d$ and any small ϵ, one can find a vector that matches an ϵ^d proportion of the columns of A, with error at most ϵ.

Proposition 5.4. *Let* $A : X \times Y \to \{-1, 1\}$ *be a matrix with* $\mathrm{Ldim}(A) = d$ *and let* $0 < \epsilon < 1/2$. *There exists a function* $\sigma : X \to \{-1, 1\}$ *and some subset* $S \subseteq Y$ *iwth* $|S| \geq \epsilon^d |Y|$ *such that*

$$\mathbf{P}_{y \in S}\big\{A(x, y) \neq \sigma(x)\big\} \leq \epsilon.$$

We are being slightly misleading here; it is not exactly the Littlestone dimension that is used in the real proof. More on this later.

Three lemmas. We now list three miscellaneous lemmas that will be used to prove Theorem 5.1. Earlier, we claimed that we were going to prove the theorem by expressing $A = \widehat{A} + (A - \widehat{A})$, bounding the block complexity of $\widehat{A}$, and then inducting on the multiplier norm of $A - \widehat{A}$. Well, the following lemma is the tool with which we bound the block complexity of $\widehat{A}$.

Lemma 5.5. *Every integer matrix* $A : X \times Y \to \mathbf{Z}$ *satisfies*

$$\mathrm{block}(A) \leq 2 \max_{x \in X} \sum_{y \in Y} \big|A(x, y)\big|.$$

Next, we observe that if a set of bounded-length vectors has a large average vector $\widehat{v}$, then subtracting $\widehat{v}$ from many of the individual vectors significantly reduces their lengths. This will help us obtain a bound on $\|A - \widehat{A}\|_{\mathrm{m}}$.

Lemma 5.6. *Let* $v_1, \ldots, v_r$ *be vectors in a Hilbert space, and let* $\widehat{v} = \mathbf{E}_{i\in[r]}\, v_i$ *denote their average. If* $\|v_i\|_2 \le \gamma$ *for all* $1 \le i \le r$*, and* $\|\widehat{v}\| = c$*, then*

$$S = \bigl\{ i \in [r] : \|v_i - \widehat{v}\|_2^2 \le \|v_i\|_2^2 - c^2/2 \bigr\}$$

satisfies $|S| \ge c^2 r/(2\gamma^2)$*, and for every* $i \in S$*,*

$$\|v_i - \widehat{v}\|_2^2 \le \|v_i\|_2^2 - \frac{c^2}{2}.$$

Lastly, we have the following lemma which, in some sense, extracts an approximately blocky matrix out of any matrix.

Lemma 5.7. *Let* $A : X \times Y \to \{-1, 1\}$ *be a sign matrix with* $|Y| = n$*. There exists a partition* $Y = \bigcup_{i=1}^{k} S_i$ *such that*

- *i) for every* i *there is a row* x_i *such that* $A(x_i, y) = 1$ *for all* $y \in S_i$*; and*
- *ii) for every row* x *and every* $\delta > 0$*, there are at most* $O_\delta(\log n)$ *sets* S_i *with*

$$\mathbf{P}_{y\in S_i}\bigl\{A(x,y) = 1\bigr\} \ge \delta.$$

A simplified account of the proof. We now give an account of the proof of Theorem 5.1, suppressing some very important technical details. The downside of this is that we will be forced at some stage simply to admit defeat and handwave the modifications needed to make the argument sound. We do this in order to illuminate the skeleton of the real proof, which may be obscured by the technical machinery used therein; our approach here also highlights the need for this machinery in the first place.

Let $A : X \times Y \to \{0, 1\}$ be a boolean matrix with $\|A\|_{\mathrm{m}} \le \gamma$. By applying the $A \mapsto 2A - J$ where J is the all-1s matrix, we can view A as a sign matrix, with multiplier norm at most $2\gamma + 1$. So we may reasonably assume that A takes values in $\{-1, 1\}$ instead, and let γ now denote the multiplier norm of this sign matrix. We will let $A_{X'\times Y'}$ denote the matrix A restricted to the rows in the subset $X' \subseteq X$ and the columns $Y' \subseteq Y$.

First we invoke Lemma 5.7 to obtain a partition $Y = \bigcup_{i=1}^{k} S_i$ for every i there is a row x_i such that $A(x_i, y) = 1$ for all $y \in S_i$. (The second statement in that lemma will not be used until later.) We then observe the bound $\mathrm{Ldim}(A) \le \gamma^2$ on the Littlestone dimension given by Proposition 5.3. By then applying Proposition 5.4, with the parameter $\epsilon_1 > 0$ to be chosen later (in fact, our proof will stop working before we get a chance to choose ϵ_1), on each of the matrices $A_{X\times S_i}$, we obtain subsets $S_i' \subseteq S_i$ with

$$|S_i'| \ge \epsilon_1^{\gamma^2} |S_i|$$

and functions $g_i : X \to \{-1, 1\}$ with

$$\mathbf{P}_{y\in S_i'}\bigl\{A(x,y) \ne g_i(x)\bigr\} \le \epsilon_1$$

for all $1 \leq i \leq k$.

For each i, let $\widehat{v}_i$ denote the average $\mathbf{E}_{y \in S_i'} v_y$, and observe that

$$\left|\langle u_{x_i}, \widehat{v}_i\rangle\right| = \left|\mathbf{E}_{y \in S_i'} \langle u_{x_i}, v_y\rangle\right| = \left|\mathbf{E}_{y \in S_i'} \langle A(x_i, y)\right| = 1,$$

since $A(x_i, y) = 1$ for all $y \in S_i$. This gives us the bound $\|\widehat{v}_i\|_2 \geq 1$ by the Cauchy–Schwarz inequality. We may then apply Lemma 5.6 with $c = 1$ to obtain subsets $S_i'' \subseteq S_i'$ with

$$|S_i''| \geq \frac{|S_i'|}{2\gamma^2}$$

and

$$\|v_i - \widehat{v}\|_2^2 \leq \|v_i\|_2^2 - \frac{1}{2}.$$

Hence if we define $Y' = \bigcup_{i=1}^k S_i'$ and $\widetilde{Y} = \bigcup_{i=1}^k S_i''$, then we have

$$|\widetilde{Y}| \geq \frac{|Y'|}{2\gamma^2} \geq \frac{\epsilon_1^{\gamma^2}}{2\gamma^2}|Y|. \tag{2}$$

We are then able to define the matrix $\widehat{A}$ on the domain $X \times \widetilde{Y}$ by setting

$$\widehat{A}(x, y) = \mathbf{E}_{y' \in S_i'} A(x, y') = \langle u_x, \widehat{v}_i\rangle$$

for every $x \in X$ and $y \in S_i''$. (Note that the average is taken over S_i' and not S_i''.) Letting $\widetilde{v}_y = v_y - \widehat{v}_i$ for all $y \in S_i''$, we have $\|\widetilde{v}_y\|_2^2 \leq \gamma^2 - 1/2$ for all $y \in \widetilde{Y}$. Then the identity

$$A(x, y) - \widehat{A}(x, y) = \langle u_x, v_y - \widehat{v}_i\rangle = \langle u_x, \widetilde{v}_y\rangle,$$

which holds for every pair $(x, y) \in X \times S_i''$, allows us to conclude that

$$\|A - \widehat{A}\|_{\mathrm{m}}^2 \leq \|A\|_{\mathrm{m}}^2 - \frac{1}{2}.$$

But we have reached a snag. (The careful reader would have realised this as soon as we defined $\widehat{A}$ in terms of averages.) The matrix $\widehat{A}$ is not ±1-valued or even integer valued! It was promised earlier that we would bound block$(\widehat{A})$, but we cannot reasonably do that for a real-valued matrix. The solution is to consider its *integer rounding* $\widehat{A}_{\mathbf{Z}}$. Using Lemma 5.5 and part (ii) of Lemma 5.7, one will be able to bound block$(\widehat{A}_{\mathbf{Z}})$, which will turn out to be enough for our purposes.

In order to repeat the procedure with $A - \widehat{A}$, we must change our inductive step to admit real-valued matrices. In fact, we only need to consider matrices whose values are close to integers. A matrix A shall be called ϵ-almost integer-valued if $\|A - A_{\mathbf{Z}}\|_\infty \leq \epsilon$. (This ϵ is related to the parameter ϵ_1 above that we left woefully unspecified.) The inductive step in the *bona fide* proof of Theorem 5.1 is the following.

Lemma 5.8. *Let* $A : X \times Y \to \mathbf{R}$ *be a real-valued matrix with* $\|A\|_{\mathrm{m}} = \gamma$. *Suppose further that* A *is* ϵ*-almost integer-valued for* $\epsilon = 2^{-20\gamma^2}$. *If* $\widehat{A}_{\mathbf{Z}}$ *is not an all-*0*s matrix, then there exists a* 2ϵ*-almost integer-valued matrix* $A' : X \times Y \to \mathbf{R}$ *such that*

$$\|A - A'\|_{\mathrm{m}}^2 \le \gamma^2 - \frac{1}{8}$$

and

$$\mathrm{block}(A'_{\mathbf{Z}}) \le 2^{O(\gamma^7)}\bigl(\log |Y|\bigr)^2.$$

The proof of this lemma proceeds as in the sketch above, but everything is much uglier. In addition to the ϵ-almost integer-valued business, one needs to define a generalised version of the Littlestone dimension for real-valued matrices, called the *α-weighted Littlestone dimension*, and prove analogues of Propositions 5.3 and 5.4 accordingly. At this stage the reader who is still unsatisfied with the details that are only vaguely alluded to here should simply consult the original paper.

Note that this proof is in spirit very related to the proof of the Green–Sanders theorem for boolean functions with small spectral norm. There, one also subtracts averages and consequently has to deal with ϵ-almost integer-valued functions.

We close this section by noting that even in this bargain-bin account of the proof, it is already possible to trace where the two logarithmic factors in Theorem 5.1 come from. One is due to the appearance of the logarithm in Lemma 5.7; this does not appear to be easily fixable, since under general hypotheses that lemma is tight. (An obvious thing to try is to use the hypothesis $\|A\|_{\mathrm{m}} \le \gamma$ to improve it.) The other logarithmic factor is a result of the bound (2), when we throw away a constant fraction of the columns to define $\widetilde{Y}$ and consequently $\widehat{A}$. Getting around this also seems difficult, since discarding columns is very necessary in our construction of the matrix $\widehat{A}$.