

Blocky Matrices and Group Idempotents

Gaia Carenini *

August 22, 2026

Abstract

We prove a common generalization of two structure theorems: the dimension-free decomposition theorem for idempotent Schur multipliers and the idempotent theorem in harmonic analysis. Roughly speaking, our result shows that an invariant integer-valued kernel with Hilbert-space factorization norm γ admits a signed decomposition into at most $2^{O(\gamma^4)}$ elementary pieces. In the matrix setting these pieces are blocky matrices, while in the group setting they are indicators of cosets. In the locally compact abelian setting this quantitatively strengthens the theorem of Green and Sanders, while in the non-abelian setting it gives a quantitative strengthening of Host's idempotent theorem and, for finite groups, of Sanders's quantitative result. It also improves the exponent in the dimension-free matrix decomposition from γ^6 to γ^4 .

1 Introduction

1.1 A common decomposition problem

We begin with two pieces of mathematics that at first sight belong to different subjects, with the aim of making clear why the same decomposition principle should govern both.

The first comes from the study of Boolean matrices and communication complexity. Call a 0/1 matrix B , with rows indexed by a set X and columns by a set Y , *blocky* if, after independently permuting its rows and its columns, it becomes a block-diagonal matrix each of whose blocks is entirely 1s, together with some all-zero rows and columns left over. A moment's thought shows that this is the same as saying that there is a partition of (a subset of) X into classes X_λ , a partition of (a subset of) Y into classes Y_λ indexed by the *same* label set, and that $B(x, y) = 1$ exactly when x and y have been placed in the same class. In other words, if we define two labelling functions ϕ, ψ , sending each row and column either to its class label or to a dummy symbol $\perp$ for the leftover rows and columns, then

$$B(x, y) = 1 \iff \phi(x) = \psi(y) \in \Lambda.$$

So a blocky matrix is nothing but an *equality test*, dressed up by first relabelling the rows and columns. This is a very simple idea, but it is worth stating carefully, because it is about to reappear in a completely different guise.

The second piece of mathematics comes from harmonic analysis. Let G be a group and let $H \leq G$ be a subgroup, and consider a left coset aH . Its indicator function, viewed through the translation-invariant kernel

$$M(x, y) := \mathbf{1}_{aH}(x^{-1}y),$$

satisfies

$$\mathbf{1}_{aH}(x^{-1}y) = 1 \iff x^{-1}y \in aH \iff y \in xaH \iff (xa)H = yH.$$

Look at the right-hand side: it says that the coset $(xa)H$, viewed as a function of x , equals the coset yH , viewed as a function of y . That is exactly an equality test again, except that now the “labels” are cosets of H in the quotient set G/H , rather than an abstract set Λ chosen ad hoc.

*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge.

So both a blocky matrix and a coset kernel are equality kernels; the only difference is what plays the role of the label set, and, crucially, whether there is a group acting compatibly on everything. This suggests that *both* of the structure theorems built from these atoms – the dimension-free theorem of Beke, Goh, Hatami, Jaffe and Naylor [1] for Schur multipliers, the quantitative idempotent theorem of Green and Sanders in the abelian setting [4], Host’s non-abelian idempotent theorem [5], and Sanders’s quantitative finite-group refinement [7] – ought to be instances of one master statement: an integer-valued kernel that is “simple” in a Hilbert-space sense (small factorization norm), and that respects a group symmetry, should be expressible with few signed equality kernels that respect the same symmetry. Proving exactly this statement, with a clean quantitative bound, is the point of this paper.

The proof builds on the decomposition framework developed by Goh–Hatami [3] and Beke–Goh–Hatami–Jaffe–Naylor [1]. In particular, the near-constancy argument and the local norm-decreasing split adapt key ingredients from those works, while the subsequent binary recursion follows the architecture of [1]. The new ingredients are the martingale estimate in Theorem 4.1, which improves the quantitative exponent, and the equivariant orbitwise implementation, which yields the non-abelian consequence.

1.2 Equality kernels and the master theorem

We now make the notion of equality kernel precise, since everything else in the paper is built out of it.

Definition 1.1 (Equality kernel). *Let X and Y be sets. A kernel $B : X \times Y \rightarrow \{0, 1\}$ is an equality kernel if there exist a set Λ (the label set) and maps*

$$\phi : X \rightarrow \Lambda \cup \{\perp\}, \quad \psi : Y \rightarrow \Lambda \cup \{\perp\}$$

such that for all $x \in X$ and $y \in Y$,

$$B(x, y) = 1 \iff \phi(x) = \psi(y) \in \Lambda.$$

If a group Γ acts on X and on Y , we say B is Γ -equivariant if $B(gx, gy) = B(x, y)$ for every $g \in \Gamma$.

The role of $\perp$ is simply bookkeeping: it lets a row or column be “unmatched”, contributing an all-zero row or column to B , without forcing us to throw it out of X or Y .

We also need a way of measuring how complicated a kernel is, and this is where Hilbert space enters. The relevant notion, standard in operator theory and in the theory of Schur multipliers, is the factorization norm.

Definition 1.2 (Factorization norm). *For a real kernel $A : X \times Y \rightarrow \mathbb{R}$, its γ_2 -norm is*

$$\|A\|_{\gamma_2} := \inf \left\{ \left(\sup_x \|u_x\| \right) \left(\sup_y \|v_y\| \right) : A(x, y) = \langle u_x, v_y \rangle \text{ for some Hilbert space vectors } u_x, v_y \right\}.$$

A factorization realizing (or nearly realizing) this infimum with $\sup_x \|u_x\| \leq 1$ and $\sup_y \|v_y\| \leq \gamma$ is called a γ -factorization.

This is a genuinely flexible notion of complexity: it does not care how large X and Y are, only about how many dimensions of “genuine variation” are needed to represent A as a Gram-type matrix. In particular it makes sense to ask for a γ_2 -factorization that is compatible with a group action, and this compatibility is what turns the analytic statement into a structural one.

Definition 1.3 (Equivariant factorization). *Let a group Γ act on sets X and Y . An equivariant γ -factorization of a kernel $A : X \times Y \rightarrow \mathbb{R}$ consists of a real Hilbert space $\mathcal{H}$, an orthogonal representation $R : \Gamma \rightarrow O(\mathcal{H})$, and vectors $u_x, v_y \in \mathcal{H}$ such that*

$$A(x, y) = \langle u_x, v_y \rangle, \quad \|u_x\| \leq 1, \quad \|v_y\| \leq \gamma,$$

and the vectors transform correctly under the action:

$$u_{gx} = R_g u_x, \quad v_{gy} = R_g v_y \quad \text{for all } g \in \Gamma.$$

The requirement $u_{gx} = R_g u_x$ is the natural thing to ask for: it says that moving a row by the group action corresponds to applying a single fixed orthogonal transformation to its associated vector, uniformly across the whole space, rather than something that could vary erratically from orbit to orbit.

We can now state the theorem around which the whole paper is organized.

Theorem 1.4 (Equivariant integral-kernel decomposition). *There is an absolute constant $C > 0$ with the following property. Let a group Γ act freely on sets X and Y , and suppose that the column orbit space Y/Γ is finite. Let $K : X \times Y \rightarrow \mathbb{Z}$ be diagonally Γ -invariant, meaning $K(gx, gy) = K(x, y)$ for all g , and suppose K has an equivariant γ -factorization. Then*

$$K = \sum_{i=1}^L \sigma_i B_i, \quad L \leq 2^{C\gamma^4},$$

where each $\sigma_i \in \{-1, 1\}$ and each B_i is a Γ -equivariant equality kernel.

The roles of the freeness and finiteness hypotheses are as follows. Freeness on the row side is used in the terminal equality decomposition, where choices made on row-orbit representatives are transported equivariantly. Freeness on the column side is used in the splitting step, where a split constructed at one column is transported uniquely around its orbit. Finiteness is needed only for the column orbit space Y/Γ , so that the orbit-processing loop terminates after finitely many steps. No finiteness assumption on X/Γ is used.

The two motivating applications are obtained by specializing the group action, as we now explain.

1.3 The trivial action: matrices

If Γ is the trivial group, then freeness is automatic and the column orbit space is just Y , so the only finiteness hypothesis in the master theorem is that Y be finite. In the finite-matrix application both X and Y are of course finite, and Γ -invariance is no constraint at all. Moreover a Γ -equivariant equality kernel is simply an equality kernel, which by the discussion above is the same thing as a blocky matrix. So Theorem 1.4 immediately specializes to a theorem purely about matrices.

Corollary 1.5 (Integer matrices). *There is an absolute constant $C > 0$ such that every finite integer matrix A with $\|A\|_{\gamma_2} \leq \gamma$ can be written*

$$A = \sum_{i=1}^L \sigma_i B_i, \quad L \leq 2^{C\gamma^4},$$

where each B_i is blocky and $\sigma_i \in \{-1, 1\}$. In particular this applies to every Boolean matrix of bounded Schur-multiplier norm.

This strengthens the dimension-free theorem of Beke, Goh, Hatami, Jaffe and Naylor [1], whose bound has γ^6 in the exponent rather than γ^4 ; we explain in Section 8 exactly where the saving comes from.

1.4 The regular action: groups

At the opposite extreme, let $X = Y = G$ for a locally compact group G , and let G act on itself by left translation, $g \cdot x = gx$. This action is free (if $gx = x$ then g is the identity) and has a single orbit, so the orbit-space finiteness hypothesis is satisfied trivially, however large or infinite G may be. An element of the Fourier–Stieltjes algebra $B(G)$ – recall that this consists of the coefficient functions $f(g) = \langle \pi(g)\xi, \eta \rangle$ of strongly continuous unitary representations π of G , with $\|f\|_{B(G)}$ the infimum of $\|\xi\| \|\eta\|$ over such representations – gives rise to exactly the kind of equivariant Hilbert factorization that Theorem 1.4 wants, once one unwinds the definitions (we carry this out in Section 9). The extra piece of input needed on top of the abstract theorem is topological: because coefficient functions of continuous representations are uniformly continuous, the subgroups that appear at the end of the recursion are automatically open. This gives:

Theorem 1.6 (Quantitative non-abelian idempotent theorem). *There is an absolute constant $C > 0$ with the following property. Let G be a locally compact group and let $f \in B(G)$ be integer-valued with $\|f\|_{B(G)} \leq \gamma$. Then there exist open subgroups $H_1, \dots, H_L \leq G$, elements $a_1, \dots, a_L \in G$, and signs $\sigma_i \in \{-1, 1\}$, with repetitions allowed, such that*

$$f = \sum_{i=1}^L \sigma_i \mathbf{1}_{a_i H_i}, \quad L \leq 2^{C\gamma^4}.$$

The passage from the matrix theorem to Theorem 1.6 is not formal. The decomposition in Beke–Goh–Hatami–Jaffe–Naylor [1] produces arbitrary blocky matrices; even when the input has the translation-invariant form $M_f(x, y) = f(x^{-1}y)$, the individual summands need not retain that form, so a coset decomposition does not follow directly. The additional structural ingredient here is an *equivariant recursion*: a local split of one column is transported around its entire group orbit using the underlying orthogonal representation, so every child kernel remains equivariant. For the regular action this means that every child still has the form $A(x, y) = g(x^{-1}y)$ for a coefficient function g . At termination, the rounded coefficient function is constant on cosets of its period subgroup, and strong continuity forces that subgroup to be open. This structural ingredient is logically distinct from the martingale estimate: equivariance is what makes the non-abelian consequence possible, while the improved weighted Littlestone bound is what sharpens the quantitative exponent from six to four.

In the Boolean special case for finite groups, this replaces the tower-type bound in Sanders’s quantitative non-abelian theorem [7] by a single exponential. Sanders’s explicit quantitative theorem is stated for finite groups; the present conclusion gives a uniform single-exponential bound for arbitrary locally compact groups.

Sanders also gave a related structural description in terms of coset decision trees: if G is finite and $f : G \rightarrow \{0, 1\}$ has $\|f\|_{A(G)} \leq M$, then f is computed by a coset decision tree with at most $\exp(\exp(\exp(O(M^2))))$ leaves [9]; we compare this with our coset-decomposition statement in Section 10.

Restricting to a locally compact *abelian* group K and taking Fourier transforms turns Theorem 1.6 into the classical quantitative idempotent theorem.

Corollary 1.7 (Abelian quantitative idempotent theorem). *Let K be a locally compact abelian group and $\mu \in M(K)$ an idempotent measure. Then*

$$\hat{\mu} = \sum_{i=1}^L \sigma_i \mathbf{1}_{\gamma_i + \Gamma_i}, \quad L \leq 2^{O(\|\mu\|^4)},$$

where each Γ_i is an open subgroup of $\hat{K}$.

Green and Sanders originally proved the bound $L \leq \exp \exp(C\|\mu\|^4)$ [4]. Sanders later summarized the finite-abelian bound as $\exp(M^{4+o(1)})$ [8]. Thus an $\exp(O(M^4))$ estimate removes the $M^{o(1)}$ loss in the exponent. On $\mathbb{F}_2^n$, Sanders’s $\exp(M^{3+o(1)})$ bound [6] remains sharper. In addition, the same argument gives a clean single-exponential bound simultaneously for arbitrary integer matrices and for arbitrary locally compact groups, abelian or not.

1.5 Proof strategy

Before beginning the proof, we describe the role of its main ingredients. The argument has several moving parts, and the following overview will be useful when they are assembled in the recursion.

Think of K as a huge integer-valued table, with a Hilbert-space vector v_y of norm at most γ attached to each column and a unit vector u_x attached to each row, so that $K(x, y) = \langle u_x, v_y \rangle$. The problem is that we want to find columns with *exactly* equal entries – that is what an equality kernel needs – while all we are handed is a Euclidean picture, in which “closeness” rather than “equality” is the natural

notion. So the argument has to do two things: it has to convert closeness into equality (by rounding), and it has to find enough closeness to begin with (this is where the Hilbert-space hypothesis is used).

The proof has four steps.

- (1) **A dimension bound for a combinatorial game.** We first show that if A has a γ -factorization, then a certain adaptive guessing game that one can play against the columns of A – formalized as a weighted version of Littlestone dimension – cannot last for more than roughly γ^2 rounds. This is Theorem 4.1, and it is proved by a short martingale computation: along a random play of the game, the “mistakes” made by an adversary correspond to increments of a random walk in Hilbert space, and predictability of the game (the row played at each round cannot depend on the future) is exactly the condition that makes the increments orthogonal in expectation, so that the total squared displacement after d rounds is only d , not d^2 .
- (2) **From a dimension bound to near-constant columns.** A finite bound on this combinatorial dimension is then turned, by an induction and pigeonhole argument in the spirit of Sauer–Shelah type proofs, into the existence of a *large* set of columns on which every row is nearly constant. This is Proposition 5.1 and Corollary 5.2.
- (3) **Averaging produces a norm-decreasing split.** Given such a large near-constant family of columns, we average their Hilbert-space vectors. The common nonzero rounded value in one row forces the average vector to have nontrivial norm, and Lemma 6.1 then makes one difference $v_{y_0} - \widehat{v}$ smaller than the ambient norm bound. Independently, pairwise separation of the distinct rounded columns and the bias–variance identity force $\widehat{v}$ itself to be smaller than the ambient bound. Thus $v_{y_0} = (v_{y_0} - \widehat{v}) + \widehat{v}$ is a split into two pieces of strictly smaller norm, both still nearly integer-valued against every row. This is Lemma 6.2, the real combinatorial heart of the paper.
- (4) **Recurring, and stopping.** We keep applying step (3), which strictly decreases a squared-norm potential each time, so it can only happen boundedly many times (at most $O(\gamma^2)$, since each split costs at least $1/16$). Eventually either the potential is exhausted, or the reduced row mass $D^\#$ falls below the splitting threshold. At that point we stop and use a greedy equivariant subtraction argument: each subtraction removes one unit of reduced row mass from every nonzero row, so an integer kernel of reduced row mass D is a signed sum of at most $2D$ equality kernels (Lemma 2.3).

To implement the equivariance in step (3), we do not split columns independently. A split constructed at one representative y_0 is transported around the whole orbit by the representation R . Freeness of the action on Y makes this unambiguous, since every point in the orbit has a unique form gy_0 . For the trivial action the orbits are singletons, while for the regular action of G there is only one orbit, so the transported split is already a global convolution split. Section 7 carries out this orbitwise construction precisely.

The rest of the paper carries out steps (1)–(4) in order (Sections 2–6), assembles them into the equivariant recursion that proves Theorem 1.4 (Section 7), and then specializes to matrices (Section 8) and to groups (Section 9), before closing with a comparison to earlier quantitative bounds in the matrix and abelian/non-abelian settings (Sections 8 and 10).

2 Equality kernels and reduced row mass

We start with the easy, purely combinatorial half of the story: understanding equality kernels themselves, and setting up the bookkeeping device – *reduced row mass* – that will tell us, later, when to stop splitting and simply read off the answer.

2.1 Equality kernels are exactly blocky matrices

We promised in the introduction that equality kernels and blocky matrices are the same notion; we now confirm this formally, since it is used at once in Corollary 1.5.

Proposition 2.1. *A Boolean matrix is blocky if and only if it is an equality kernel.*

Proof. Suppose B is blocky, so that its support decomposes as a disjoint union $\text{supp}(B) = \bigcup_{\lambda} X_{\lambda} \times Y_{\lambda}$, with the X_{λ} pairwise disjoint and the Y_{λ} pairwise disjoint. Label every element of X_{λ} and of Y_{λ} by λ , and label every row or column not appearing in any block by $\perp$. Then $B(x, y) = 1$ exactly when x and y received the same non- $\perp$ label, which is the definition of an equality kernel.

Conversely, if $B(x, y) = 1 \iff \phi(x) = \psi(y) \in \Lambda$, then for each $\lambda \in \Lambda$ let $X_{\lambda} = \phi^{-1}(\lambda)$ and $Y_{\lambda} = \psi^{-1}(\lambda)$; these are the desired blocks. $\square$

This is a triviality, but it is the triviality that makes the whole paper possible: it says that proving a decomposition theorem with equality kernels as atoms, in the case of the trivial group action, is the theorem about blocky matrices.

2.2 Reduced row mass

Now suppose a group Γ acts on X and Y , and $K : X \times Y \rightarrow \mathbb{Z}$ is diagonally invariant. We are going to want to measure, at any stage of the argument, how far K is from being expressible with a bounded number of equality kernels. The right measure turns out not to be the naive row sum $\sum_y |K(x, y)|$, which can be enormous simply because many columns happen to agree with one another; instead we should count each *distinct* column only once.

Definition 2.2 (Column equivalence and reduced row mass). *For an integer kernel $K : X \times Y \rightarrow \mathbb{Z}$, define an equivalence relation on Y by*

$$y \sim_K y' \iff K(x, y) = K(x, y') \text{ for every } x \in X,$$

and write $\mathcal{C}(K) = Y/\sim_K$ for the resulting set of distinct columns. Because K is Γ -invariant, the relation $\sim_K$ is Γ -invariant too, so Γ acts on $\mathcal{C}(K)$ by $g[y] = [gy]$, and for a class $C \in \mathcal{C}(K)$ the value $K(x, C)$ is unambiguous. Define the reduced row mass of K to be

$$D^{\sharp}(K) := \sup_{x \in X} \sum_{C \in \mathcal{C}(K)} |K(x, C)|,$$

interpreting a sum of nonnegative terms over a possibly infinite index set as the supremum of its finite subsums. For a real kernel A that is close to an integer kernel (made precise in Section 3), we set $D^{\sharp}(A) := D^{\sharp}(A_{\mathbb{Z}})$, where $A_{\mathbb{Z}}$ is its entrywise rounding.

For a finite matrix, $D^{\sharp}(K)$ is simply the largest row sum you get after you have thrown away duplicate columns – exactly the quantity relevant to the following observation, which says that once $D^{\sharp}$ is finite, we are already done, and moreover done cheaply.

Lemma 2.3 (Terminal equality decomposition). *Let Γ act freely on X , and let $K : X \times Y \rightarrow \mathbb{Z}$ be Γ -invariant with $D^{\sharp}(K) = D < \infty$. Then*

$$K = \sum_{i=1}^L \sigma_i B_i, \quad L \leq 2D,$$

with each B_i a Γ -equivariant equality kernel.

Proof. We first treat the case $K \geq 0$; the general case follows by applying this to the positive and negative parts K^+ and K^- separately. Passing from K to either K^+ or K^- can only merge column classes, so each has reduced row mass at most D ; moreover $K = K^+ - K^-$.

Pick one representative row x from each orbit of X . For each such representative whose row is not identically zero, choose *some* column class $C_x \in \mathcal{C}(K)$ with $K(x, C_x) > 0$: such a class exists because the row is nonzero. We now extend this single choice, made once per orbit, to the whole orbit equivariantly, by declaring $C_{gx} := gC_x$. Freeness of the action on X guarantees that every point of X is gx for a unique g (once we have fixed the representative x of its orbit), so this is a genuine, well-defined function of all of X , not just of the representatives.

Now define

$$B(x, y) = 1 \iff [y] = C_x$$

on rows with nonzero K , and $B \equiv 0$ on rows belonging to an orbit where K vanishes identically. This B is manifestly an equality kernel: the labelling map on rows sends x to C_x (or to $\perp$ on zero rows), and the labelling map on columns sends y to its class $[y]$. It is Γ -equivariant by construction, since $C_{gx} = gC_x$ was arranged precisely so that the definition of B commutes with the action.

The point of subtracting B from K is that, on every row x where the row was nonzero, the value on the selected old column class C_x decreases by exactly 1, and it stays nonnegative. Moreover B is constant on every old column class of K , so passing from K to $K - B$ can only merge old column classes; it cannot split one. Hence, for each row, the reduced row mass of $K - B$ is at most the sum of the absolute values on the old classes, and that sum is exactly one smaller than before on every nonzero row. Repeating this operation can therefore happen at most D times before every row is zero, i.e. before K itself is exhausted. This proves the lemma with $L \leq D$ in the case $K \geq 0$, and hence $L \leq 2D$ after combining K^+ and K^- . $\square$

We will meet $D^\sharp$ again as the quantity that decides, at every node of the recursion in Section 7, whether we should stop (invoking Lemma 2.3) or split further.

3 Hilbert factorizations and the weighted mistake tree

We now set up the analytic side of the argument: what it means for a kernel to be “almost” integer-valued, and the combinatorial game – the weighted mistake tree – whose depth we are going to bound using the Hilbert-space factorization.

Throughout, recall the factorization norm $\|\cdot\|_{\gamma_2}$ from the introduction. We record for later use the elementary but useful fact that

$$\|A\|_{\max} \leq \|A\|_{\gamma_2},$$

which follows from Cauchy–Schwarz applied to $A(x, y) = \langle u_x, v_y \rangle$. For finite matrices, $\|\cdot\|_{\gamma_2}$ coincides with the usual Schur-multiplier norm.

Rounding will be used repeatedly, so we fix conventions. For $t \in \mathbb{R}$, write $t_{\mathbb{Z}}$ for a choice of nearest integer to t (fixed once and for all at half-integers, say by rounding down). A real kernel A is ϵ -almost integer-valued if $\|A - A_{\mathbb{Z}}\|_{\max} \leq \epsilon$, where $A_{\mathbb{Z}}(x, y) := (A(x, y))_{\mathbb{Z}}$.

Now to the combinatorial game. Fix $\alpha > 0$; think of α as a resolution below which we are not able, or not willing, to distinguish values. The following notion is a weighted analogue of Littlestone dimension, the parameter from online learning theory that measures how deep a binary decision tree of “queries” one can shatter.

Definition 3.1 (Weighted mistake tree and its dimension). *A weighted mistake tree of depth d over X is a complete binary tree of depth d whose internal nodes ν are labelled by pairs $(x(\nu), w(\nu)) \in X \times \mathbb{R}$: a row to query, and a threshold. We think of the tree as a strategy for an adaptive guesser who, at each internal node, asks “is the entry in row $x(\nu)$ large or small relative to $w(\nu)$?” and moves left or right accordingly.*

A matrix $A \in \mathbb{R}^{X \times Y}$ α -shatters the tree if every root-to-leaf path can be “realized” by some column: that is, for every path there is a $y \in Y$ such that at each internal node ν on the path,

$$A(x(\nu), y) \geq w(\nu) + \alpha/2 \quad \text{if the path turns left at } \nu,$$

$$A(x(\nu), y) \leq w(\nu) - \alpha/2 \quad \text{if the path turns right at } \nu.$$

The largest depth of a tree that A α -shatters is the weighted Littlestone dimension $\text{Ldim}_\alpha(A)$.

The intuition is that if $\text{Ldim}_\alpha(A)$ is large, then A contains a rich combinatorial structure: for every one of the exponentially many strategies the adversary (choosing left/right moves) might follow, there really is a column of A that is consistent with every single query along that particular path, always by a margin of at least $\alpha/2$. This is a strong demand, and our first real theorem says that a bounded factorization norm makes it impossible to sustain for very long.

One more piece of notation before we begin: we will run this game along a *random* path, generated by independent uniform signs $\varepsilon_1, \dots, \varepsilon_d \in \{-1, 1\}$, going left when $\varepsilon_t = 1$ and right when $\varepsilon_t = -1$. Write $\mathcal{F}_t = \sigma(\varepsilon_1, \dots, \varepsilon_t)$ for the natural filtration. The crucial structural feature of the tree is that the row x_t and threshold w_t queried at level t , having been determined by the tree structure and by $\varepsilon_1, \dots, \varepsilon_{t-1}$ alone, are $\mathcal{F}_{t-1}$ -measurable: in probabilists’ language, they are *predictable*. Predictability is the key property used in the martingale estimate below.

4 The martingale bound on weighted Littlestone dimension

Theorem 4.1 (Martingale weighted Littlestone bound). *For every real matrix A and every $\alpha > 0$,*

$$\text{Ldim}_\alpha(A) \leq \frac{4\|A\|_{\gamma_2}^2}{\alpha^2}.$$

We first describe the idea, since the short proof is easiest to read with the probabilistic picture in mind. Fix a γ -factorization of A and suppose A α -shatters a tree of depth d . We are going to follow a *random* root-to-leaf path. At each level t we pick up a vector $\varepsilon_t u_{x_t}$, where x_t is the row queried at that level and $\varepsilon_t = \pm 1$ tells us which way the path turned. If we sum these vectors over $t = 1, \dots, d$, we get a random vector in Hilbert space; and the shattering hypothesis is designed so that this random vector, paired against the vector v_y of a column consistent with the whole path, is forced to be large – roughly $\alpha d/2$. On the other hand, the *typical* size of a sum of d vectors of norm at most 1, each multiplied by an independent random sign, is only about $\sqrt{d}$, not d – *provided* the vectors themselves do not conspire with the signs. This is exactly where predictability comes in: because the row x_t picked at time t cannot look ahead at ε_t , the cross terms $\varepsilon_s \varepsilon_t \langle u_{x_s}, u_{x_t} \rangle$ for $s < t$ average out to zero, just as in the classical fact that a martingale’s quadratic variation adds rather than multiplies. Comparing the “forced to be at least $\alpha d/2$ ” lower bound with the “typically about $\gamma\sqrt{d}$ ” upper bound and solving for d gives the theorem.

Proof. Fix a γ -factorization $A(x, y) = \langle u_x, v_y \rangle$, $\|u_x\| \leq 1$, $\|v_y\| \leq \gamma$, and suppose A α -shatters a weighted mistake tree of depth d . Let $\varepsilon_1, \dots, \varepsilon_d$ be independent uniform random signs, and follow the path that turns left when $\varepsilon_t = 1$ and right when $\varepsilon_t = -1$. Write x_t, w_t for the (random, but predictable) row and threshold encountered at level t .

By definition of shattering, for every realization of the path there is a column y (depending on the realization) such that for every t ,

$$\varepsilon_t (\langle u_{x_t}, v_y \rangle - w_t) \geq \frac{\alpha}{2}.$$

Summing over $t = 1, \dots, d$ and using linearity of the inner product in its first argument,

$$\left\langle \sum_{t=1}^d \varepsilon_t u_{x_t}, v_y \right\rangle \geq \frac{\alpha d}{2} + \sum_{t=1}^d \varepsilon_t w_t.$$

Since $\|v_y\| \leq \gamma$, the left-hand side is at most $\gamma \left\| \sum_t \varepsilon_t u_{x_t} \right\|$, regardless of which y realizes the path; so, taking the supremum over the (path-dependent) choice of y and then expectation over the random path, and using that $\mathbb{E}[\varepsilon_t w_t] = 0$ because w_t is $\mathcal{F}_{t-1}$ -measurable while ε_t is independent of $\mathcal{F}_{t-1}$ with mean zero,

$$\frac{\alpha d}{2} \leq \gamma \mathbb{E} \left\| \sum_{t=1}^d \varepsilon_t u_{x_t} \right\|.$$

By Cauchy–Schwarz, $\mathbb{E} \|\cdot\| \leq (\mathbb{E} \|\cdot\|^2)^{1/2}$, so it remains to bound the second moment. Expand the square:

$$\mathbb{E} \left\| \sum_{t=1}^d \varepsilon_t u_{x_t} \right\|^2 = \sum_{t=1}^d \mathbb{E} \|u_{x_t}\|^2 + 2 \sum_{s < t} \mathbb{E} [\varepsilon_s \varepsilon_t \langle u_{x_s}, u_{x_t} \rangle].$$

Fix $s < t$. Conditioning on $\mathcal{F}_{t-1}$, the vectors u_{x_s} and u_{x_t} (which depend only on $\varepsilon_1, \dots, \varepsilon_{t-1}$, since x_t is predictable), and also ε_s , are all $\mathcal{F}_{t-1}$ -measurable, while ε_t is independent of $\mathcal{F}_{t-1}$ and has mean zero. Hence

$$\mathbb{E} [\varepsilon_s \varepsilon_t \langle u_{x_s}, u_{x_t} \rangle] = \mathbb{E} [\varepsilon_s \langle u_{x_s}, u_{x_t} \rangle \cdot \mathbb{E} [\varepsilon_t \mid \mathcal{F}_{t-1}]] = 0.$$

So every cross term vanishes, and, using $\|u_{x_t}\| \leq 1$,

$$\mathbb{E} \left\| \sum_{t=1}^d \varepsilon_t u_{x_t} \right\|^2 = \sum_{t=1}^d \mathbb{E} \|u_{x_t}\|^2 \leq d.$$

Combining, $\alpha d/2 \leq \gamma \sqrt{d}$, i.e. $d \leq 4\gamma^2/\alpha^2$. Finally, taking the infimum over γ -factorizations of A (i.e. letting $\gamma \rightarrow \|A\|_{\gamma_2}$) proves the theorem. $\square$

It is instructive to compare this with the more naive route, used in earlier work on Schur multipliers, of encoding the threshold w_t directly into an enlarged factorization (for instance by appending a coordinate). That route gives a bound of the shape $\text{Ldim}_\alpha(A) = O_\alpha(\gamma^4)$, quadratically worse in γ than Theorem 4.1. The martingale argument above avoids this loss entirely, because it never needs to fold the thresholds w_t into the geometry at all – they only ever interact with the mean-zero random signs ε_t , and disappear upon taking expectation. This quadratic saving is exactly what turns the final exponent of the whole paper from six down to four; see Section 8.

5 From bounded dimension to near-constant columns

We now perform step (2) of the plan in Section 1.5: convert the dimension bound on Ldim_α into a statement that a matrix of bounded γ_2 -norm has, in every row simultaneously, a *large* set of columns on which its value barely moves. This kind of implication – bounded combinatorial dimension forces large uniform, or near-uniform, substructures – is the same flavour of argument that appears in the Sauer–Shelah lemma and its many descendants, and the proof below is a weighted, real-valued version of the standard argument.

Proposition 5.1 (Near constancy from weighted Littlestone dimension). *Let $A \in \mathbb{R}^{X \times Y}$ with Y finite and nonempty, let $d = \text{Ldim}_\alpha(A) < \infty$, let $M = \|A\|_{\max}$, and let $\eta \in (0, 1]$. Put $q := \lceil 2M/\alpha \rceil + 2$. Then there exist $S \subseteq Y$ and $g : X \rightarrow [-M, M]$ such that*

$$|S| \geq |Y| \left(\frac{\eta}{q} \right)^d$$

and, for every $x \in X$,

$$\Pr_{y \in S} [|A(x, y) - g(x)| > 2\alpha] \leq \eta.$$

This is a quantitative reformulation of the weighted-Littlestone near-constancy argument of Goh and Hatami [3, Proposition 3.1]; we include the proof both for completeness and to keep track of the parameters needed below.

Proof. We argue by strong induction on d .

If $d = 0$: for every row x the values $\{A(x, y) : y \in Y\}$ must have diameter less than α . Indeed, if two columns y, y' had $|A(x, y) - A(x, y')| \geq \alpha$, then labelling the single-node tree by (x, w) , where w is the midpoint of $A(x, y)$ and $A(x, y')$, would be α -shattered by A (one of y, y' realizes the left branch, the other the right), giving $\text{Ldim}_\alpha(A) \geq 1$, a contradiction. So $S = Y$ and any $g(x)$ in this tiny range works, in particular with the stronger conclusion that the exceptional probability is 0.

Now suppose $d \geq 1$ and the proposition holds for every dimension strictly smaller than d . There are two cases.

Case A. Suppose that for every row x there is an interval of length 4α containing at least $(1 - \eta)|Y|$ of the values $\{A(x, y) : y \in Y\}$. Then take $g(x)$ to be the centre of such an interval (projected into $[-M, M]$ if necessary), and take $S = Y$. Every value in the chosen interval is at distance at most 2α from $g(x)$, so the stated strict-tail bound holds.

Case B. Otherwise there is a row x for which no interval of length 4α captures a $(1 - \eta)$ -fraction of Y . Partition the range $[-M, M]$ into consecutive intervals of length α ; there are at most q of them, so they define at most q column classes $Y_1, \dots, Y_q$ (grouping y by which interval $A(x, y)$ lands in). By pigeonhole some class, say Y_i , has $|Y_i| \geq |Y|/q$.

Now, the union $Y_{i-1} \cup Y_i \cup Y_{i+1} \cup Y_{i+2}$ (indices truncated to the valid range) lies inside a single interval of length 4α by construction; and by the assumption of Case B, this union cannot capture a $(1 - \eta)$ -fraction of Y . So at least $\eta|Y|$ of the columns lie *outside* it, and by pigeonhole again, some single class Y_j among those outside classes has $|Y_j| \geq \eta|Y|/q$.

Because Y_j lies outside $Y_{i-1} \cup Y_i \cup Y_{i+1} \cup Y_{i+2}$, its interval index satisfies $j \leq i - 2$ or $j \geq i + 3$. Thus the value intervals supporting Y_i and Y_j are separated by a gap of at least α . Choosing a threshold w at the midpoint of this gap puts one class entirely above $w + \alpha/2$ and the other entirely below $w - \alpha/2$.

Consider the two restrictions $A|_{X \times Y_i}$ and $A|_{X \times Y_j}$. If *both* had weighted Littlestone dimension at least d , we could take a shattered tree of depth d for each, attach them as the two subtrees below a new root labelled (x, w) , and get a tree of depth $d + 1$ shattered by A (a path through the left subtree is realized by a column in Y_i , consistent with the root because Y_i lies above $w + \alpha/2$, and similarly on the right) – contradicting $\text{Ldim}_\alpha(A) = d$. So one of the two restrictions, say to $Y' \in \{Y_i, Y_j\}$, has $\text{Ldim}_\alpha(A|_{X \times Y'}) \leq d - 1$, and $|Y'| \geq \eta|Y|/q$.

Let

$$d' := \text{Ldim}_\alpha(A|_{X \times Y'}), \quad M' := \|A|_{X \times Y'}\|_{\max}, \quad q' := \lceil 2M'/\alpha \rceil + 2.$$

Then $d' \leq d - 1$ and $q' \leq q$. By the strong inductive hypothesis applied to $A|_{X \times Y'}$ with the same η , there are $S \subseteq Y'$ and g such that

$$|S| \geq |Y'| \left(\frac{\eta}{q'} \right)^{d'} \geq |Y'| \left(\frac{\eta}{q} \right)^{d-1} \geq |Y| \left(\frac{\eta}{q} \right)^d.$$

Here we used $0 < \eta/q \leq 1$, together with $d' \leq d - 1$, in the second inequality. The required near-constancy conclusion is inherited from the inductive hypothesis. This completes the induction. $\square$

Specializing the parameters gives the version we will actually use.

Corollary 5.2 (Large near-constant set). *There is an absolute constant C such that if $A \in \mathbb{R}^{X \times Y}$ with Y finite and nonempty, $\|A\|_{\gamma_2} \leq \gamma$, $\|A\|_{\max} \leq M$, and $\eta \in (0, 1]$, then there exist $S \subseteq Y$ and $g : X \rightarrow [-M, M]$ with*

$$|S| \geq |Y| \left(\frac{\eta}{C(1 + M)} \right)^{C\gamma^2}$$

and $\Pr_{y \in S}[|A(x, y) - g(x)| > 1/4] \leq \eta$ for every $x \in X$.

Proof. Apply Proposition 5.1 with $\alpha = 1/8$, so that “ 2α ” there becomes $1/4$, and bound $d = \text{Ldim}_{1/8}(A) \leq 256\gamma^2$ using Theorem 4.1; also bound $q = \lceil 16M \rceil + 2 \leq C(1 + M)$. Substituting these into Proposition 5.1 gives the stated exponent, after absorbing constants into C . $\square$

6 The local norm-decreasing split

We come now to step (3): the combinatorial heart of the paper. Everything so far has produced, from a bound on $\|A\|_{\gamma^2}$, a large set of columns S on which every row is nearly constant. We now show how to exploit this to peel a genuine chunk of squared norm off a single column vector, while keeping every row of the two resulting pieces close to an integer. Fix, throughout this section, absolute constants $C_0, C_1 \geq 1$, chosen large enough for the proof to work, and set

$$T(\gamma, \epsilon) := 2 \left(\frac{C_0 \gamma^2}{\epsilon} \right)^{C_1 \gamma^2}.$$

This threshold function will reappear constantly: it is the size below which the reduced row mass $D^\#$ is small enough that we should stop splitting and instead invoke Lemma 2.3.

We first isolate an elementary Hilbert-space fact, essentially just the parallelogram law in disguise, which is what forces a genuine norm decrease whenever we average several vectors that are not too spread out.

Lemma 6.1 (A vector close to its average). *Let $v_1, \dots, v_r$ be vectors in a Hilbert space with $\|v_i\| \leq \gamma$ for all i , and let $\hat{v} = r^{-1} \sum_i v_i$ be their average. Then some i satisfies*

$$\|v_i - \hat{v}\|^2 \leq \|v_i\|^2 - \frac{\|\hat{v}\|^2}{2}.$$

Proof. The identity

$$\frac{1}{r} \sum_{i=1}^r \|v_i - \hat{v}\|^2 = \frac{1}{r} \sum_{i=1}^r \|v_i\|^2 - \|\hat{v}\|^2$$

is the usual bias-variance decomposition (expand the left side and use that $\sum_i (v_i - \hat{v}) = 0$). If the desired inequality failed for *every* i – that is, if $\|v_i - \hat{v}\|^2 > \|v_i\|^2 - \frac{1}{2}\|\hat{v}\|^2$ for every i – then averaging this strict inequality over i would give

$$\frac{1}{r} \sum_i \|v_i - \hat{v}\|^2 > \frac{1}{r} \sum_i \|v_i\|^2 - \frac{1}{2}\|\hat{v}\|^2,$$

which contradicts the identity above (it would force $-\|\hat{v}\|^2 > -\frac{1}{2}\|\hat{v}\|^2$, i.e. $\|\hat{v}\|^2 < 0$). So the inequality must hold for some i . $\square$

The point of this innocuous lemma, in context, will be to guarantee that either the average vector $\hat{v}$ is small, or else some individual vector v_{y_0} in our near-constant family sits substantially closer to $\hat{v}$ than its own norm would naively suggest – either way, we will be able to write $v_{y_0} = v' + v''$ with both pieces of norm strictly less than γ .

We can now state and prove the key local lemma.

The splitting mechanism below adapts [1, Lemma 2.4]. The sharper threshold comes from replacing the $O(\gamma^4)$ weighted-Littlestone estimate used there by the $O(\gamma^2)$ martingale estimate of Theorem 4.1.

Lemma 6.2 (Improved norm-decreasing column split). *Let $\gamma > 1/2$ and $0 < \epsilon < 1/6$, and let $A \in \mathbb{R}^{X \times Y}$ be a finite ϵ -almost integer-valued matrix equipped with a fixed γ -factorization $A(x, y) = \langle u_x, v_y \rangle$. Suppose that $A_{\mathbb{Z}}$ has no two equal columns, and that*

$$\max_x \sum_y |A_{\mathbb{Z}}(x, y)| \geq T(\gamma, \epsilon).$$

Then there exist $y_0 \in Y$ and vectors v', v'' with

$$v_{y_0} = v' + v'', \quad \max\{\|v'\|^2, \|v''\|^2\} \leq \gamma^2 - \frac{1}{16},$$

and, for every $x \in X$,

$$\text{dist}(\langle u_x, v' \rangle, \mathbb{Z}) \leq 3\epsilon, \quad \text{dist}(\langle u_x, v'' \rangle, \mathbb{Z}) \leq 3\epsilon.$$

Proof. The strategy is: find a row x_0 with enormous reduced mass; extract, from that row, a huge class S of columns that all take the same rounded value $b \neq 0$ in row x_0 ; shrink S , using Corollary 5.2, to a subset S' that is near-constant *in every row simultaneously*; then average the vectors v_y over S' , and apply Lemma 6.1. Because the rounded columns indexed by S' are pairwise distinct (as full columns, although they agree at x_0), their representing vectors are separated in Hilbert space; this average cannot be too large, forcing the desired norm decrease; and because S' is near-constant on every row, the average $\hat{v}$ is nearly as good as any single column of S' at reproducing near-integer values in every row – which is what lets us split v_{y_0} into two pieces that are *both* still almost-integer-valued. We now carry this out.

Write $M = \|A\|_{\max}$ and $D = \max_x \sum_y |A_{\mathbb{Z}}(x, y)|$ ($\geq T(\gamma, \epsilon) > 0$). Since $D > 0$, some rounded entry of A is nonzero, and an entry rounding to a nonzero integer must itself have absolute value at least $1 - \epsilon$ (else it would round to 0); so $M \geq 1 - \epsilon > 1/2$. On the other hand $M \leq \|A\|_{\gamma_2} \leq \gamma$ by the factorization.

Fix a row x_0 realizing D . Every nonzero rounded entry of row x_0 has absolute value at most $M + 1/2 \leq 2M$ (an entry rounding to integer k lies within $\epsilon < 1/2$ of k , and is itself bounded by M , so $|k| \leq M + 1/2$). Hence row x_0 has at least $D/(2M)$ nonzero entries. Since these nonzero rounded entries take at most $8M$ possible integer values (they lie in $[-2M, 2M] \setminus \{0\}$, and $2M \geq 1$), by pigeonhole some fixed nonzero integer b is attained on a set

$$S = \{y : A_{\mathbb{Z}}(x_0, y) = b\}, \quad |S| \geq \frac{D}{16M^2}.$$

Set $\eta = \epsilon/(4M)$, and apply Corollary 5.2 to the restriction of A to $X \times S$ (still of γ_2 -norm at most γ and max-norm at most M , since restricting a matrix cannot increase either norm). We obtain $S' \subseteq S$ and $g' : X \rightarrow [-M, M]$ with

$$|S'| \geq |S| \left(\frac{\epsilon}{CM(1+M)} \right)^{C\gamma^2} \geq \frac{D}{16M^2} \left(\frac{\epsilon}{CM(1+M)} \right)^{C\gamma^2},$$

and, for every x , at most an η -fraction of S' has $|A(x, y) - g'(x)| > 1/4$. Since $D > 0$ we already know $M \geq 1 - \epsilon > 5/6$, and hence $\gamma \geq 5/6$. Thus

$$M(1+M) \leq \gamma(1+\gamma) \leq 3\gamma^2.$$

Using $D \geq T(\gamma, \epsilon)$ in the preceding display, we obtain

$$|S'| \geq \frac{1}{8\gamma^2} \left(\frac{C_0\gamma^2}{\epsilon} \right)^{C_1\gamma^2} \left(\frac{\epsilon}{C'\gamma^2} \right)^{C\gamma^2}$$

for an absolute C' . Choosing first $C_1 > C$ and then C_0 sufficiently large makes this quantity at least 2 uniformly for $\gamma \geq 5/6$ and $0 < \epsilon < 1/6$. Hence $|S'| \geq 2$.

Let $g(x)$ be a nearest integer to $g'(x)$. If $|A(x, y) - g'(x)| \leq 1/4$, put $k = A_{\mathbb{Z}}(x, y)$. Since A is ϵ -almost integer-valued,

$$|g'(x) - k| \leq |g'(x) - A(x, y)| + |A(x, y) - k| \leq \frac{1}{4} + \epsilon < \frac{1}{2}.$$

Thus k is the unique nearest integer to $g'(x)$, so $k = g(x)$. Hence

$$\Pr_{y \in S'} [A_{\mathbb{Z}}(x, y) \neq g(x)] \leq \eta \quad \text{for every } x.$$

Now let $\hat{v} := |S'|^{-1} \sum_{y \in S'} v_y$. Because every $y \in S'$ lies in S , we have $A_{\mathbb{Z}}(x_0, y) = b \neq 0$ for all such y . Hence

$$|\langle u_{x_0}, \hat{v} \rangle - b| = \left| \frac{1}{|S'|} \sum_{y \in S'} (A(x_0, y) - b) \right| \leq \epsilon.$$

Since $|b| \geq 1$, it follows that

$$|\langle u_{x_0}, \hat{v} \rangle| \geq 1 - \epsilon > \frac{1}{2},$$

so $\|\hat{v}\| \geq |\langle u_{x_0}, \hat{v} \rangle| > 1/2$ (using $\|u_{x_0}\| \leq 1$ and Cauchy–Schwarz). By Lemma 6.1 applied to the family $\{v_y\}_{y \in S'}$, there is $y_0 \in S'$ with

$$\|v_{y_0} - \hat{v}\|^2 \leq \|v_{y_0}\|^2 - \frac{1}{2} \|\hat{v}\|^2 \leq \gamma^2 - \frac{1}{8}.$$

It remains to bound $\|\hat{v}\|^2$ itself away from γ^2 , and for this we use that the columns of S' are pairwise *distinct* rounded columns of a matrix with no duplicate rounded columns: any two $y, y' \in S' \subseteq Y$ differ in $A_{\mathbb{Z}}$ in some row, so $|A(x, y) - A(x, y')| \geq 1 - 2\epsilon > 1/2$ for that row. Since

$$A(x, y) - A(x, y') = \langle u_x, v_y - v_{y'} \rangle$$

and $\|u_x\| \leq 1$, Cauchy–Schwarz gives $\|v_y - v_{y'}\| \geq |A(x, y) - A(x, y')| > 1/2$. The pairwise form of the bias-variance identity,

$$\frac{1}{2|S'|^2} \sum_{y, y' \in S'} \|v_y - v_{y'}\|^2 = \frac{1}{|S'|} \sum_{y \in S'} \|v_y\|^2 - \|\hat{v}\|^2,$$

(obtained by expanding both sides and comparing) then gives, using $\|v_y - v_{y'}\|^2 > 1/4$ for the $|S'|(|S'| - 1)$ ordered pairs with $y \neq y'$ and $\|v_y\|^2 \leq \gamma^2$,

$$\|\hat{v}\|^2 \leq \gamma^2 - \frac{|S'| - 1}{8|S'|} \leq \gamma^2 - \frac{1}{16},$$

the last step using $|S'| \geq 2$.

Set $v'' := \hat{v}$ and $v' := v_{y_0} - \hat{v}$. We have just shown $\|v''\|^2 \leq \gamma^2 - 1/16$ and (from the previous paragraph) $\|v'\|^2 \leq \gamma^2 - 1/8 \leq \gamma^2 - 1/16$, so both norm bounds hold.

Finally we check almost-integrality of $\langle u_x, v' \rangle$ and $\langle u_x, v'' \rangle$ for every $x \in X$. On the $(1 - \eta)$ -fraction of S' where $A_{\mathbb{Z}}(x, y) = g(x)$, we have $|A(x, y) - g(x)| \leq \epsilon$; on the remaining η -fraction, we only know $|A(x, y) - g(x)| \leq |A(x, y)| + |g(x)| \leq M + (M + 1) \leq 2M + 1$ crudely. Averaging over S' ,

$$|\langle u_x, v'' \rangle - g(x)| = \left| \frac{1}{|S'|} \sum_{y \in S'} (A(x, y) - g(x)) \right| \leq (1 - \eta)\epsilon + \eta(2M + 1) \leq \epsilon + \eta(2M + 1).$$

Recalling $\eta = \epsilon/(4M)$, we get $\eta(2M + 1) \leq \epsilon(2M + 1)/(4M) \leq \epsilon$ (using $M \geq 1/2$, so $2M + 1 \leq 4M$ once $M \geq 1/2$: indeed $2M + 1 \leq 4M \iff 1 \leq 2M \iff M \geq 1/2$, which holds). Hence $|\langle u_x, v'' \rangle - g(x)| \leq 2\epsilon$, so $\text{dist}(\langle u_x, v'' \rangle, \mathbb{Z}) \leq 2\epsilon \leq 3\epsilon$.

For v' : since $v_{y_0} = v' + v''$,

$$\langle u_x, v' \rangle = A(x, y_0) - \langle u_x, v'' \rangle = (A(x, y_0) - g(x)) - (\langle u_x, v'' \rangle - g(x)).$$

The first bracket is within ϵ of the integer $A_{\mathbb{Z}}(x, y_0) - g(x)$, while the second bracket has absolute value at most 2ϵ . Therefore $\langle u_x, v' \rangle$ is within 3ϵ of the integer $A_{\mathbb{Z}}(x, y_0) - g(x)$, and hence $\text{dist}(\langle u_x, v' \rangle, \mathbb{Z}) \leq 3\epsilon$, completing the proof. $\square$

The two halves v', v'' produced by this lemma are, informally, “ v_{y_0} minus its near-constant average” and “the average itself”. Both improve the ambient squared column-norm bound by a universal amount, $1/16$, and, crucially, both are still nearly integer-valued when tested against *every* row x , not just the rows used to find the near-constant set. This is exactly the local move that will be iterated, and transported around group orbits, in the next section.

7 The equivariant recursion

We now assemble the pieces (Sections 2–6) into a proof of Theorem 1.4. As foreshadowed in Section 1.5, the only new idea needed is a way of performing the local split of Lemma 6.2 *compatibly with the group action*, and the mechanism for this is transport around a single free orbit, which we now describe.

7.1 Splitting one orbit

Lemma 7.1 (Splitting one orbit). *Let Γ act on X and freely on Y . Let $A : X \times Y \rightarrow \mathbb{R}$ be ϵ -almost integer-valued, $0 < \epsilon < 1/32$, with an equivariant γ -factorization. If $D^\sharp(A) > T(\gamma, \epsilon)$, then there is a Γ -orbit $\mathcal{O} \subseteq Y$ and equivariant column families $\{v_y^{(1)}\}_{y \in \mathcal{O}}, \{v_y^{(2)}\}_{y \in \mathcal{O}}$ such that*

$$v_y = v_y^{(1)} + v_y^{(2)} \quad (y \in \mathcal{O}), \quad \|v_y^{(j)}\|^2 \leq \gamma^2 - \frac{1}{16} \quad (j = 1, 2),$$

and $\text{dist}(\langle u_x, v_y^{(j)} \rangle, \mathbb{Z}) \leq 3\epsilon$ for every $x \in X$ and $y \in \mathcal{O}$.

Proof. Since $D^\sharp(A) > T(\gamma, \epsilon) > 0$, some rounded entry is nonzero. Hence $\gamma \geq \|A\|_{\max} \geq 1 - \epsilon > 1/2$, so the hypothesis $\gamma > 1/2$ needed in Lemma 6.2 is automatic. By definition of reduced row mass there is a row x_0 and a *finite* collection of distinct rounded-column classes whose masses in row x_0 already sum to more than $T(\gamma, \epsilon)$. Choose one representative column from each of these classes; call the resulting finite set $Y_0 \subseteq Y$. Because the classes were chosen distinct, the rounded columns restricted to Y_0 are pairwise different *somewhere*; so there is a finite set of rows $X_0 \ni x_0$ that separates every pair of columns in Y_0 (for each of the finitely many pairs, keep one witnessing row).

We would like to apply Lemma 6.2 directly to A restricted to $X \times Y_0$, but the lemma requires the ambient row set to be finite, whereas X itself may be infinite. We get around this by a compactness argument: for every *finite* superset $X' \supseteq X_0$, apply Lemma 6.2 to the finite matrix $A|_{X' \times Y_0}$ (which still has no duplicate rounded columns, since $X_0 \subseteq X'$ already separates them, and still has reduced row mass exceeding $T(\gamma, \epsilon)$ at x_0). This produces some $y_{X'} \in Y_0$ and vectors $w_{1,X'}, w_{2,X'}$ satisfying the norm-decrease and almost-integrality conclusions, but only verified so far on the rows in X' .

Order the finite supersets X' of X_0 by inclusion; this is a directed set. Since Y_0 is finite, we may pass to a subnet along which the chosen column $y_{X'}$ is eventually constant, equal to some $y_0 \in Y_0$. The vectors $w_{1,X'}$ and $w_{2,X'}$ all lie in the closed ball of radius $\sqrt{\gamma^2 - 1/16}$, which is weakly compact (closed bounded balls in Hilbert space are weakly compact); so, passing to a further subnet, we may assume $w_{1,X'} \rightharpoonup w_1$ and $w_{2,X'} \rightharpoonup w_2$ weakly.

We now pass the conclusions of Lemma 6.2 to this weak limit. The equation $v_{y_0} = w_{1,X'} + w_{2,X'}$, valid along the net, passes to the limit by weak continuity of addition, giving $v_{y_0} = w_1 + w_2$. The norm bound $\|w_{j,X'}\|^2 \leq \gamma^2 - 1/16$ passes to the limit by weak lower semicontinuity of the norm (a standard fact: if $w_{j,X'} \rightharpoonup w_j$ then $\|w_j\| \leq \liminf \|w_{j,X'}\|$), giving $\|w_j\|^2 \leq \gamma^2 - 1/16$. Finally, for each *fixed* $x \in X$, once $X' \supseteq \{x\}$ we have from the lemma that $\text{dist}(\langle u_x, w_{j,X'} \rangle, \mathbb{Z}) \leq 3\epsilon$; since $\langle u_x, \cdot \rangle$ is weakly continuous and $\text{dist}(\cdot, \mathbb{Z})$ is continuous, this bound also passes to the limit, and it holds for *every* $x \in X$ because x was an arbitrary fixed point (once in the directed set, it stays in every larger X').

We have therefore produced $y_0 \in Y$ and w_1, w_2 with $v_{y_0} = w_1 + w_2$, $\|w_j\|^2 \leq \gamma^2 - 1/16$, and $\text{dist}(\langle u_x, w_j \rangle, \mathbb{Z}) \leq 3\epsilon$ for every $x \in X$.

Finally, transport this single split around the whole orbit $\mathcal{O} = \Gamma y_0$. Because the action on Y is free, every $y \in \mathcal{O}$ has a *unique* representation $y = gy_0$; define $v_{gy_0}^{(j)} := R_g w_j$. This is well-defined (no ambiguity in g), and

$$v_{gy_0} = R_g v_{y_0} = R_g w_1 + R_g w_2 = v_{gy_0}^{(1)} + v_{gy_0}^{(2)},$$

using equivariance of the original factorization, and the norms are preserved since R_g is orthogonal: $\|R_g w_j\| = \|w_j\|$. For almost-integrality at a general point $x \in X$ and $y = gy_0 \in \mathcal{O}$,

$$\langle u_x, v_{gy_0}^{(j)} \rangle = \langle u_x, R_g w_j \rangle = \langle R_g^{-1} u_x, w_j \rangle = \langle u_{g^{-1}x}, w_j \rangle,$$

using that R_g is orthogonal (so $R_g^* = R_g^{-1}$) and equivariance $u_{g^{-1}x} = R_g^{-1}u_x$. Since $g^{-1}x$ ranges over X as x does, this is within 3ϵ of an integer by the bound already established for all of X . So the almost-integrality conclusion holds at every point of $X \times \mathcal{O}$, as claimed. $\square$

The role of the compactness step is worth emphasizing. It allows us to deduce the infinitary statement, for instance the group specialization when G is infinite, from Lemma 6.2, which is stated and proved only for finite matrices. We only need finitely many rows at a time, and weak compactness passes these finite conclusions to a global vector.

7.2 One step of the recursion

The following orbitwise decomposition is the equivariant counterpart of the decomposition step in [1, Lemma 2.5].

Lemma 7.2 (One equivariant decomposition step). *Let Γ act freely on X and Y , with Y/Γ finite. Let $A : X \times Y \rightarrow \mathbb{R}$ be ϵ -almost integer-valued, $0 < \epsilon < 1/32$, with an equivariant γ -factorization. Then*

$$A_{\mathbb{Z}} = K_1 + K_2 + F,$$

where each $K_j = (A_j)_{\mathbb{Z}}$ for an equivariant 3ϵ -almost integer-valued kernel A_j with an equivariant factorization satisfying $\sup_y \|v_y^{(j)}\|^2 \leq \gamma^2 - 1/16$, and

$$\text{eq}_{\Gamma}(F) \leq 2T(\gamma, \epsilon),$$

where $\text{eq}_{\Gamma}(F)$ denotes the least number of signed Γ -equivariant equality kernels needed to write F .

Proof. Think of every Γ -orbit of columns as either “processed” or “unprocessed”, starting with all orbits unprocessed. While the restriction of A to the union of the currently unprocessed orbits still has reduced row mass exceeding $T(\gamma, \epsilon)$, apply Lemma 7.1 to that restriction: it selects some unprocessed orbit and splits its column vectors into two equivariant families, which we append to two growing families $\{v_y^{(1)}\}$ and $\{v_y^{(2)}\}$ (defined so far only on processed orbits), and we mark that orbit as processed.

Because Y/Γ is finite, there are only finitely many orbits to process, so this loop terminates after finitely many iterations. When it terminates, either every orbit has been processed, or the restriction to the remaining unprocessed orbits has reduced row mass at most $T(\gamma, \epsilon)$.

On the union of the processed (split) orbits, define A_1, A_2 using the recorded families $v^{(1)}, v^{(2)}$; on the unprocessed orbits, set both to zero. Each A_j is then 3ϵ -almost integer-valued (by the conclusion of Lemma 7.1 on split orbits, and trivially on unprocessed orbits, where it is exactly 0), equivariant (each split was performed equivariantly, orbit by orbit), and satisfies the norm bound $\sup_y \|v_y^{(j)}\|^2 \leq \gamma^2 - 1/16$ (again by Lemma 7.1 on split orbits, trivially elsewhere).

On every split orbit we have the exact identity $A = A_1 + A_2$ (this is how the split was constructed); consequently, using the triangle inequality and the almost-integrality bounds $\|A - A_{\mathbb{Z}}\|_{\max} \leq \epsilon$ and $\|A_j - (A_j)_{\mathbb{Z}}\|_{\max} \leq 3\epsilon$,

$$|A_{\mathbb{Z}} - (A_1)_{\mathbb{Z}} - (A_2)_{\mathbb{Z}}| \leq |A_{\mathbb{Z}} - A| + |A - A_1 - A_2| + |A_1 - (A_1)_{\mathbb{Z}}| + |A_2 - (A_2)_{\mathbb{Z}}| \leq \epsilon + 0 + 3\epsilon + 3\epsilon < 1$$

pointwise on every split orbit. The left-hand side is an integer (a difference of integers), and an integer of absolute value less than 1 must be 0. So $A_{\mathbb{Z}} = K_1 + K_2$ exactly on every split orbit, where $K_j = (A_j)_{\mathbb{Z}}$.

On the unprocessed orbits, $K_1 = K_2 = 0$ by definition, so there $A_{\mathbb{Z}} = K_1 + K_2 + F$ with $F := A_{\mathbb{Z}}$ restricted to those orbits (and $F = 0$ on split orbits). By the termination condition, $D^{\#}(F) \leq T(\gamma, \epsilon)$ (reduced row mass can only decrease under restriction to a subset of columns, so the mass of F , supported on the unprocessed orbits, is controlled by the mass of A restricted to those same orbits). F is Γ -invariant (being a restriction, to a union of orbits, of the invariant kernel $A_{\mathbb{Z}}$) and Γ acts freely on X (inherited from the hypothesis on X), so Lemma 2.3 applies and gives $\text{eq}_{\Gamma}(F) \leq 2D^{\#}(F) \leq 2T(\gamma, \epsilon)$. $\square$

7.3 Proof of the master theorem

We are now ready to prove Theorem 1.4, by iterating Lemma 7.2: each application either terminates a branch (if the residual reduced row mass was already small) or splits it into two children, each of which has strictly smaller squared factorization norm. Since the norm cannot decrease indefinitely, the recursion tree has bounded depth, and we simply sum the (bounded) equality complexity picked up at every node.

Proof of Theorem 1.4. If $K = 0$ there is nothing to prove, so assume $K \neq 0$. Since K is integer-valued and nonzero, $\|K\|_{\max} \geq 1$, and therefore the initial column-norm parameter satisfies $\gamma \geq 1$. Set

$$\epsilon_0 := 2^{-32\gamma^2-5}.$$

It is convenient to keep track of what the recursive objects actually are. A node carries a pair (A, K_A) , where A has an equivariant factorization, is ϵ -almost integer-valued for some current error ϵ , and $K_A := A_{\mathbb{Z}}$ is the integer kernel to be reconstructed. At the root we take $A = K$, so $K_A = K$. Although the actual rounding error is 0, we use ϵ_0 as the bookkeeping error parameter in the first application of Lemma 7.2. If the current factorization has

$$\|u_x\| \leq 1, \quad \|v_y\| \leq \rho,$$

we call ρ the current column-norm parameter.

Apply Lemma 7.2 to A . It gives

$$K_A = (A_1)_{\mathbb{Z}} + (A_2)_{\mathbb{Z}} + F,$$

where F is already a signed sum of at most $2T(\rho, \epsilon)$ equivariant equality kernels, while each child with $(A_j)_{\mathbb{Z}} \neq 0$ is 3ϵ -almost integer-valued and has column-norm parameter ρ_j satisfying

$$\rho_j^2 \leq \rho^2 - \frac{1}{16}.$$

Thus the children are the pairs $(A_j, (A_j)_{\mathbb{Z}})$, not the rounded kernels viewed as though they themselves carried the factorization.

Along any branch, every genuine split lowers the squared column-norm parameter by at least $1/16$. Consequently a branch contains at most $16\gamma^2$ genuine splits. At depth $s \leq 16\gamma^2$, the error is at most

$$3^s \epsilon_0 \leq 3^{16\gamma^2} 2^{-32\gamma^2-5} < \frac{1}{32},$$

because $16 \log_2 3 < 26$. Hence every recursive use of Lemma 7.2 satisfies its error hypothesis. Moreover, whenever a node has nonzero rounded kernel, some entry of A is within $\epsilon < 1/32$ of a nonzero integer, so

$$\rho \geq \|A\|_{\max} \geq 1 - \epsilon > \frac{1}{2}.$$

Thus the local splitting lemma is applicable whenever a split is actually needed.

A binary tree with at most $16\gamma^2$ genuine splits on each branch has at most $2^{16\gamma^2+1}$ nodes (a node at which no split occurs is terminal). It remains to bound the equality complexity contributed at a node uniformly in terms of the root parameters. Along the recursion we have $\rho \leq \gamma$ and $\epsilon \geq \epsilon_0$. Since $C_0\gamma^2/\epsilon_0 > 1$,

$$T(\rho, \epsilon) \leq 2 \left(\frac{C_0\gamma^2}{\epsilon_0} \right)^{C_1\gamma^2}.$$

Therefore

$$\log_2 T(\rho, \epsilon) \leq O\left(\gamma^2 \log \gamma + \gamma^2 \log \frac{1}{\epsilon_0}\right) = O(\gamma^4).$$

So every node contributes at most $2^{O(\gamma^4)}$ equality kernels. Multiplying by the number of nodes gives

$$L \leq 2^{16\gamma^2+1} 2^{O(\gamma^4)} = 2^{O(\gamma^4)}.$$

Finally, the exact identity $K_A = (A_1)_{\mathbb{Z}} + (A_2)_{\mathbb{Z}} + F$ at every internal node implies, by induction from the leaves to the root, that the equality-kernel pieces produced throughout the tree sum exactly to the rounded kernel at that node. At the root the rounded kernel is K . This proves the theorem. $\square$

8 Matrix consequences

We record explicitly what Theorem 1.4 says when Γ is trivial, one of our two motivating applications. As explained in the introduction, for the trivial action freeness is automatic and the column-orbit finiteness assumption reduces to finiteness of Y ; in the present finite-matrix application X is finite as well. The trivial action turns Γ -equivariant equality kernels into ordinary equality kernels, which by Proposition 2.1 are exactly blocky matrices. So Theorem 1.4 is, in this case, precisely Corollary 1.5. Because $\|\cdot\|_{\gamma_2}$ coincides with the Schur-multiplier norm on finite matrices, we may restate this for Boolean matrices in the language of operator theory.

Corollary 8.1 (Idempotent Schur multipliers). *Every finite Boolean Schur multiplier A of multiplier norm at most γ is a signed sum of at most $2^{O(\gamma^4)}$ contractive idempotent Schur multipliers.*

Here we use the standard characterization of blocky Boolean matrices as contractive idempotent Schur multipliers; see, for example, [3, 1].

The improvement over Beke–Goh–Hatami–Jaffe–Naylor [1], whose bound is $2^{O(\gamma^6)}$, is entirely localized in one place: the sharper martingale bound of Theorem 4.1. Their argument, like ours, uses a near-constancy consequence of a bound on $\text{Ldim}_{1/8}(A)$; the weighted-Littlestone input underlying this step is $O(\gamma^4)$ in the earlier approach [3, 1], whereas ours is $O(\gamma^2)$ – precisely the saving discussed in the remark after Theorem 4.1. Tracing this through, the logarithm of the local splitting threshold is

$$\log T(\gamma, \epsilon) = O(\text{Ldim}_{1/8}(A) (\log(1 + \gamma) + \log(1/\epsilon))).$$

The recursion requires $\log(1/\epsilon_0) = O(\gamma^2)$. Hence the old bound $\text{Ldim}_{1/8}(A) = O(\gamma^4)$ gives $\log T = O(\gamma^6)$, while Theorem 4.1 gives $\log T = O(\gamma^4)$. The binary recursion tree itself contributes only an additional $O(\gamma^2)$ to the logarithm of the number of terms, because its depth is $O(\gamma^2)$. Thus the final exponent improves from six to four, and the saving is exactly the quadratic improvement in the weighted Littlestone bound.

9 The regular group action

We now carry out the other specialization: $X = Y = G$, with G a locally compact group acting on itself by left translation. The work here is not to reprove the recursion – that has already been done, once and for all, in Section 7 – but to identify correctly (i) what an equivariant factorization looks like for a function in $B(G)$, and (ii) what the terminal equality kernels look like for this particular action, namely that they are exactly indicator functions of cosets, with the acting group’s role explaining why the relevant subgroups come out *open*.

9.1 Coefficient functions give equivariant factorizations

Recall that the Fourier–Stieltjes algebra $B(G)$ of a locally compact group G consists of the coefficient functions $g \mapsto \langle \pi(g)\xi, \eta \rangle$ of strongly continuous unitary representations π of G on a complex Hilbert space, with norm $\|f\|_{B(G)}$ the infimum of $\|\xi\| \|\eta\|$ over all such representations of f [2].

Lemma 9.1 (Equivariant coefficient factorization). *Let $f \in B(G)$ be real-valued and let $\gamma > \|f\|_{B(G)}$. Then there are a real Hilbert space $\mathcal{H}_{\mathbb{R}}$, a strongly continuous orthogonal representation $R : G \rightarrow O(\mathcal{H}_{\mathbb{R}})$, and vectors $u_x, v_y \in \mathcal{H}_{\mathbb{R}}$ (for $x, y \in G$) such that*

$$f(x^{-1}y) = \langle u_x, v_y \rangle_{\mathbb{R}}, \quad \|u_x\| \leq 1, \quad \|v_y\| \leq \gamma,$$

and $u_{tx} = R_t u_x, v_{ty} = R_t v_y$ for every $t \in G$.

Proof. By the definition of the $B(G)$ norm, choose a coefficient realization $f(g) = \langle \pi(g)\xi, \eta \rangle_{\mathbb{C}}$ with $\|\xi\| \|\eta\| < \gamma$. If $f \neq 0$, rescale ξ, η so that $\|\eta\| = 1$ and $\|\xi\| < \gamma$; the case $f = 0$ is trivial. Define $u_x := \pi(x)\eta$ and $v_y := \pi(y)\xi$, vectors in the complex Hilbert space of π . Assuming, as is conventional, that the complex inner product is linear in its first argument, we compute

$$\langle v_y, u_x \rangle_{\mathbb{C}} = \langle \pi(y)\xi, \pi(x)\eta \rangle_{\mathbb{C}} = \langle \pi(x)^* \pi(y)\xi, \eta \rangle_{\mathbb{C}} = \langle \pi(x^{-1}y)\xi, \eta \rangle_{\mathbb{C}} = f(x^{-1}y),$$

using unitarity of π (so $\pi(x)^* = \pi(x)^{-1} = \pi(x^{-1})$) and the representation property. Since f is real-valued by hypothesis, the imaginary part of $\langle v_y, u_x \rangle_{\mathbb{C}}$ vanishes identically, and the real part of a complex inner product is exactly the real inner product obtained by viewing the same complex Hilbert space as a real Hilbert space $\mathcal{H}_{\mathbb{R}}$ (of twice the real dimension) in the usual way; write $\langle u_x, v_y \rangle_{\mathbb{R}}$ for this real part (note we have also swapped the order, which does not affect the real part since the real part of a complex number equals the real part of its conjugate). So $\langle u_x, v_y \rangle_{\mathbb{R}} = f(x^{-1}y)$.

The norms are unchanged by realification: $\|u_x\|_{\mathbb{R}} = \|\eta\|_{\mathbb{C}} = 1$ and $\|v_y\|_{\mathbb{R}} = \|\xi\|_{\mathbb{C}} < \gamma$. Equivariance is immediate from the representation property: $u_{tx} = \pi(tx)\eta = \pi(t)\pi(x)\eta = \pi(t)u_x$, and $\pi(t)$, restricted to act on the realification $\mathcal{H}_{\mathbb{R}}$, is exactly the orthogonal transformation R_t we want (a unitary map on a complex Hilbert space restricts to an orthogonal map on its realification, because it preserves the real inner product, being the real part of a preserved complex inner product); similarly for v_{ty} . Strong continuity of R is inherited from strong continuity of π . $\square$

This lemma allows us to apply Theorem 1.4 to $X = Y = G$ under the regular action: for every $\gamma > \|f\|_{B(G)}$, an integer-valued $f \in B(G)$ gives an equivariant γ -factorization of the kernel $K(x, y) := f(x^{-1}y)$, which is manifestly diagonally invariant under left translation (replacing x, y by tx, ty leaves $x^{-1}y$ unchanged) and, since there is only one column orbit on $Y = G$, satisfies the orbit-finiteness hypothesis trivially.

9.2 Why the terminal subgroups are open

The one piece of Theorem 1.4 that does not come for free, in this specialization, is that the subgroups H_i appearing in the final answer should be *open*: this is a genuinely topological fact, not visible in the purely algebraic/Hilbert-space argument of Section 7, and it is exactly where the hypothesis that $f \in B(G)$ (rather than an arbitrary function with a Hilbert factorization) is used.

For an integer-valued function $k : G \rightarrow \mathbb{Z}$, define its *right-period subgroup*

$$\text{Per}_R(k) := \{h \in G : k(zh) = k(z) \text{ for every } z \in G\}$$

(a routine check confirms this is indeed a subgroup: it contains the identity, and is closed under products and inverses).

Lemma 9.2 (Period subgroups are open). *Let $g : G \rightarrow \mathbb{R}$ be a coefficient of a strongly continuous orthogonal representation (equivalently, after complexification, of a strongly continuous unitary representation), and let $0 < \delta < 1/2$ be such that g is δ -almost integer-valued. Then $H := \text{Per}_R(g\mathbb{Z})$ is open.*

Proof. Coefficient functions of strongly continuous unitary representations are (right) uniformly continuous: given ξ, η with $g(x) = \langle \pi(x)\xi, \eta \rangle$, strong continuity of π at the identity gives a neighbourhood U of e such that $\|\pi(h)\xi - \xi\| < \varepsilon'$ for $h \in U$, and then

$$|g(xh) - g(x)| = |\langle \pi(x)(\pi(h)\xi - \xi), \eta \rangle| \leq \|\pi(h)\xi - \xi\| \|\eta\| < \varepsilon' \|\eta\|$$

uniformly in x (using that $\pi(x)$ is unitary, hence norm-preserving). Choosing ε' small enough that $\varepsilon'\|\eta\| < 1 - 2\delta$ (possible since $\delta < 1/2$), we obtain a neighbourhood U of the identity with

$$\sup_x |g(xh) - g(x)| < 1 - 2\delta \quad \text{for all } h \in U.$$

Now suppose $h \in U$ but $g_{\mathbb{Z}}(xh) \neq g_{\mathbb{Z}}(x)$ for some x : then $g_{\mathbb{Z}}(xh)$ and $g_{\mathbb{Z}}(x)$ are distinct integers, so differ by at least 1, while $g(xh)$ and $g(x)$ each lie within δ of their respective rounded values; the triangle inequality then forces

$$|g(xh) - g(x)| \geq 1 - 2\delta,$$

contradicting the bound above. So no such x exists, i.e. $U \subseteq H$; since U is a neighbourhood of the identity contained in the subgroup H , and subgroups containing a neighbourhood of the identity are open (translate the neighbourhood to cover every point of H), H is open. $\square$

9.3 Duplicate columns are exactly cosets

For $k : G \rightarrow \mathbb{Z}$, consider the associated convolution kernel $M_k(x, y) := k(x^{-1}y)$.

Lemma 9.3 (Duplicate columns are cosets). *Let $H = \text{Per}_R(k)$. Two columns $y, y' \in G$ of M_k are equal (as functions of x) if and only if $yH = y'H$.*

Proof. Write $y' = yh$. The columns indexed by y and y' agree, i.e. $k(x^{-1}y) = k(x^{-1}y')$ for every x , if and only if, substituting $z = x^{-1}y$ (which ranges over all of G as x does), $k(z) = k(zh)$ for every $z \in G$ – which is exactly the definition of $h \in H = \text{Per}_R(k)$. So the columns agree iff $h \in H$, i.e. iff $yH = y'H$. $\square$

Consequently, in the regular action, the equivalence classes $\mathcal{C}(M_k)$ of Section 2 are exactly the left cosets $\{yH : y \in G\}$ (the subgroup is written on the right), and the reduced row mass becomes

$$D^\sharp(M_k) = \sum_{aH \in G/H} |k(a)|.$$

When this is finite, k is trivially a finite integer combination of coset indicators, $k = \sum_{aH} k(a) \mathbf{1}_{aH}$, and expanding each integer coefficient into a sum of signs recovers a signed coset decomposition, exactly as Lemma 2.3 promises in this special case.

It remains to identify the equivariant equality kernels for the regular action with cosets in general (not just in the terminal, small- $D^\sharp$ case), confirming the picture sketched in the introduction.

Proposition 9.4. *A nonzero translation-invariant Boolean kernel $B(x, y) = \mathbf{1}_S(x^{-1}y)$, for $S \subseteq G$, is an equality kernel if and only if S is a left coset of some subgroup of G .*

Proof. ($\Leftarrow$) If $S = aH$ for a subgroup $H \leq G$, then

$$\mathbf{1}_{aH}(x^{-1}y) = 1 \iff x^{-1}y \in aH \iff (xa)H = yH,$$

so B is the equality kernel with labels in G/H , given by $\phi(x) = (xa)H$ and $\psi(y) = yH$.

($\Rightarrow$) Suppose $B(x, y) = \mathbf{1}_S(x^{-1}y)$ is a nonzero equality kernel. By Proposition 2.1, B is blocky: writing the row associated to x as $\{y : x^{-1}y \in S\} = xS$, blockiness says that the family of sets $\{xS : x \in G\}$ is pairwise either identical or disjoint. Fix any $a \in S$ (possible since $S \neq \emptyset$, as $B \neq 0$), and set

$$H := \{g \in G : gS = S\}.$$

This is a subgroup: $eS = S$; if $gS = S$ and $g'S = S$ then $(gg')S = g(g'S) = gS = S$; and if $gS = S$ then applying g^{-1} to both sides gives $S = g^{-1}S$.

Now take any $s \in S$. The translate $(sa^{-1})S$ contains $(sa^{-1})a = s$, and S also contains s ; so $(sa^{-1})S$ and S are two members of our pairwise-disjoint-or-equal family that share the point s , hence

they are equal: $(sa^{-1})S = S$. This says $sa^{-1} \in H$, i.e. $s \in Ha$. Since s was an arbitrary element of S , this shows $S \subseteq Ha$.

Conversely, if $h \in H$, then $a \in S$ and $hS = S$ imply $ha \in S$. Thus $Ha \subseteq S$. Combining the two inclusions, $S = Ha$, which is a right coset of H ; writing $Ha = a(a^{-1}Ha)$ exhibits it as a left coset of the conjugate subgroup $a^{-1}Ha$, as claimed. $\square$

9.4 Proof of the non-abelian idempotent theorem

Proof of Theorem 1.6. The case $f = 0$ is trivial. Put $M = \|f\|_{B(G)}$; then $M \geq 1$ because f is nonzero and integer-valued. Choose any γ with

$$M < \gamma \leq 2M.$$

Lemma 9.1 gives an equivariant γ -factorization of $K(x, y) = f(x^{-1}y)$ under the regular left action of G on itself. This action is free and has one orbit on each of $X = G$ and $Y = G$, so the recursion in the proof of Theorem 1.4 applies.

We verify that every equality piece produced by the recursion can be chosen to be the indicator of an *open* coset. The row vectors retain the form $u_x = R_x u_e$ throughout. If a split at a node is first constructed at a column y_0 , Lemma 7.1 transports a child vector as $v_{gy_0}^{(j)} = R_g w_j$. Writing $y = gy_0$ gives

$$v_y^{(j)} = R_{yy_0^{-1}} w_j = R_y (R_{y_0^{-1}} w_j),$$

so on every child the column vectors again have the form $v_y = R_y w$ for a fixed vector w . Hence every real kernel occurring at a node has the form

$$A(x, y) = \langle R_x u_e, R_y w \rangle = \langle u_e, R_{x^{-1}y} w \rangle = g(x^{-1}y),$$

where g is a coefficient of the same strongly continuous orthogonal representation (equivalently, after complexification, of a strongly continuous unitary representation), and A is δ -almost integer-valued with $\delta < 1/32$.

There is one further point specific to the regular action. At every node there is exactly one column orbit. Therefore, if $D^\sharp(A) > T(\rho, \delta)$, the orbit-processing loop in Lemma 7.2 processes that unique orbit and leaves *no* residual term F . If instead $D^\sharp(A) \leq T(\rho, \delta)$, no split occurs and the entire rounded kernel is terminal. Thus every nonzero residual contribution in the abstract recursion occurs here only at a terminal node.

At such a terminal node put $k = g\mathbb{Z}$ and $H = \text{Per}_R(k)$. Lemma 9.2 shows that H is open. By Lemma 9.3, the distinct columns of the rounded kernel are indexed by the cosets aH , and

$$D^\sharp(M_k) = \sum_{aH \in G/H} |k(a)|.$$

This quantity is at most the relevant threshold, so only finitely many cosets have nonzero value and

$$k = \sum_{aH \in G/H} k(a) \mathbf{1}_{aH}.$$

Expanding each nonzero integer coefficient into $|k(a)|$ copies with the appropriate sign gives an exact signed decomposition into indicators of cosets of the open subgroup H . Its number of terms is $D^\sharp(M_k)$, so it is no larger than the $2T(\rho, \delta)$ terminal cost used in the abstract master-theorem count. Consequently the same recursion bound gives at most $2^{C\gamma^4}$ open-coset terms. Since $\gamma \leq 2M$, this is at most $2^{C'M^4}$ after changing the absolute constant. Hence

$$f = \sum_{i=1}^L \sigma_i \mathbf{1}_{a_i H_i}, \quad L \leq 2^{C'M^4},$$

with every H_i open. $\square$

For finite groups G , the Fourier algebra $A(G)$ and Fourier–Stieltjes algebra $B(G)$ coincide isometrically (every function on a finite group is a coefficient function, and the two norms agree), giving the following clean statement, free of any topological caveats since every subgroup of a finite group is (trivially) open.

Corollary 9.5 (Finite groups). *Let G be finite and $f : G \rightarrow \mathbb{Z}$ satisfy $\|f\|_{A(G)} \leq \gamma$. Then*

$$f = \sum_{i=1}^L \sigma_i \mathbf{1}_{a_i H_i}, \quad L \leq 2^{O(\gamma^4)}.$$

10 Abelian consequences, and a comparison with earlier bounds

We now specialize once more, from an arbitrary locally compact group to a locally compact *abelian* group, to recover the classical Green–Sanders theorem, and take the opportunity to place our bound alongside other quantitative bounds available in the abelian literature.

Let K be a locally compact abelian group and $\mu \in M(K)$ an idempotent measure, meaning $\mu * \mu = \mu$. Taking Fourier transforms, this says $\hat{\mu}^2 = \hat{\mu}$ on the dual group $\hat{K}$, i.e. $\hat{\mu}$ is Boolean-valued. Under the standard isometric identification $B(\hat{K}) \cong M(K)$ (Fourier–Stieltjes transform), we have $\|\hat{\mu}\|_{B(\hat{K})} = \|\mu\|$. Applying Theorem 1.6 to the group $\hat{K}$ and the integer-valued (indeed Boolean-valued) function $\hat{\mu} \in B(\hat{K})$ gives Corollary 1.7 at once.

We next compare the quantitative bounds with the earlier results; the different settings are not directly comparable in every respect.

- Green and Sanders’s original theorem [4], valid for arbitrary locally compact abelian K , gave the qualitatively weaker bound $L \leq \exp \exp(C\|\mu\|^4)$: a tower of height two, rather than a single exponential. Theorem 1.6 removes one level of the tower, in a setting that is in fact more general at the group level, since Theorem 1.6 allows arbitrary locally compact groups.
- In the more restrictive setting of integer-valued functions on *finite abelian* groups, Sanders [8] summarizes his bound as $\exp(M^{4+o(1)})$. An $\exp(O(M^4))$ estimate removes the $M^{o(1)}$ loss in the exponent. On the specific model group $\mathbb{F}_2^n$, however, Sanders’s $\exp(M^{3+o(1)})$ bound [6] remains substantially sharper. Our main point is that the same argument works without abelian-specific input and gives a single-exponential bound for every locally compact group.
- The Green–Sanders argument also keeps track of the number of *distinct* subgroups appearing across the whole decomposition, a finer invariant than the total number L of coset terms (which allows repetitions). Our recursion, as it stands, only controls L ; different leaves of our recursion tree can and typically will produce unrelated period subgroups, and we have no mechanism at present for collapsing or coordinating them. We regard finding such a mechanism, in the non-abelian setting, as an interesting question left open by this paper.
- In the Boolean special case, Sanders’s quantitative non-abelian idempotent theorem [7] for finite groups gives a triply iterated tower bound in $O(\|f\|_{A(G)})$. Theorem 1.6 replaces this by a single exponential $\exp(O(\gamma^4))$, and, unlike [7], applies verbatim to arbitrary locally compact groups, not just finite ones.
- Sanders also obtained a related quantitative structural theorem for Boolean functions on finite groups in terms of *coset decision trees* [9]. If G is finite and $f : G \rightarrow \{0, 1\}$ satisfies $\|f\|_{A(G)} \leq M$, then f can be computed by a coset decision tree with at most $\exp(\exp(\exp(O(M^2))))$ leaves. Conversely, Sanders observes that a Boolean function computed by a coset decision tree with m leaves has Fourier-algebra norm at most $\exp(O(m))$. Thus coset decision trees provide another quantitative formulation of the principle that Boolean functions of small Fourier-algebra norm are governed by coset structure. This result controls a different structural parameter from the signed coset-decomposition length considered here, so the two bounds are not directly comparable.

- For the matrix theorem, Beke, Goh, Hatami, Jaffe and Naylor [1] proved that a Boolean matrix of γ_2 -norm at most γ is a signed sum of $2^{O(\gamma^6)}$ blocky matrices. Corollary 1.5 improves the exponent from six to four. The integer-valued formulation is included for convenience; the generalized almost-integer formulation in [1] already encompasses integer matrices as well.

Use of AI tools. AI-assisted tools were used solely to improve the wording, organization, and clarity of the exposition. They were not used to generate the mathematical arguments.

Acknowledgments

The author is grateful to Lianna Hambardzumyan for bringing this problem to her attention, to Daniel Naylor for helpful bibliographic suggestions, and to Imre Leader for his guidance and support. Part of this work was carried out during the Circuits, Communication, and Proofs program at ICTS. The author was supported by the CB European PhD Studentship, funded by Trinity College, Cambridge.

References

- [1] C. Beke, M. K. Goh, H. Hatami, S. Jaffe, and D. Naylor, *A characterization of idempotent Schur multipliers*, arXiv:2607.14316, version 2, 2026.
- [2] P. Eymard, *L’algèbre de Fourier d’un groupe localement compact*, Bull. Soc. Math. France **92** (1964), 181–236.
- [3] M. K. Goh and H. Hatami, *Block complexity and idempotent Schur multipliers*, Int. Math. Res. Not. IMRN (2025), rna356.
- [4] B. Green and T. Sanders, *A quantitative version of the idempotent theorem in harmonic analysis*, Ann. of Math. (2) **168** (2008), 1025–1054.
- [5] B. Host, *Le théorème des idempotents dans $B(G)$* , Bull. Soc. Math. France **114** (1986), 215–223.
- [6] T. Sanders, *Boolean functions with small spectral norm, revisited*, Math. Proc. Cambridge Philos. Soc. **167** (2019), 335–344.
- [7] T. Sanders, *A quantitative version of the non-abelian idempotent theorem*, Geom. Funct. Anal. **21** (2011), 141–221.
- [8] T. Sanders, *Bounds in Cohen’s idempotent theorem*, J. Fourier Anal. Appl. **26** (2020), Paper No. 25.
- [9] T. Sanders, *Coset decision trees and the Fourier algebra*, J. Anal. Math. **144** (2021), 227–259.