Approximate Representations and Approximate Homomorphisms

Cristopher Moore
Computer Science Department
University of New Mexico
and the Santa Fe Institute
moore@cs.unm.edu

Alexander Russell
Computer Science and Engineering
University of Connecticut
acr@cse.uconn.edu

October 1, 2010

Abstract

Approximate algebraic structures play a defining role in arithmetic combinatorics and have found remarkable applications to basic questions in number theory and pseudorandomness. Here we study *approximate representations* of finite groups: functions $\psi : G \to \mathsf{U}_d$ such that $\Pr[\psi(xy) = \psi(x)\,\psi(y)]$ is large, or more generally $\mathbb{E}_{x,y} \|\psi(xy) - \psi(x)\,\psi(y)\|_2^2$ is small, where x, y are uniformly random elements of the group G and U_d denotes the unitary group of degree d. We bound these quantities in terms of the ratio $d/d_{\min}$ where $d_{\min}$ is the dimension of the smallest nontrivial representation of G. As an application, we bound the extent to which a function $f : G \to H$ can be an approximate homomorphism where H is another finite group. We show that if H's representations are significantly smaller than G's, no such f can be much more homomorphic than a random function.

We interpret these results as showing that if G is quasirandom, that is, if $d_{\min}$ is large, then G cannot be embedded in a small number of dimensions, or in a less-quasirandom group, without significant distortion of G's multiplicative structure. We also prove that our bounds are tight by showing that minors of genuine representations and their polar decompositions are essentially optimal approximate representations.

In additive combinatorics and number theory, an *approximate subgroup* of a group G is a subset H which is roughly closed under multiplication: that is, such that $\Pr_{x,y}[xy \in H]$ is large where x, y are uniformly random elements of H. We focus on *approximate group representations*—functions ψ from G to U_d, the group of $d \times d$ unitary matrices, such that ψ acts roughly like a homomorphism. We then use our results to bound the existence of approximate homomorphisms from G to another finite group H.

Let G be a finite group and let $\psi : G \to \mathsf{U}_{d_\psi}$. If $\psi(xy) = \psi(x)\,\psi(y)$ for all $x, y \in G$, then we call ψ a *representation*. We are interested in understanding how close ψ can be to a representation if G does not in fact have any d_ψ-dimensional representations—in particular, in the case where G is *quasirandom* [1] in the sense that its smallest nontrivial representation has dimension $d_{\min} > d_\psi$.

We can measure the extent to which ψ fails to act as a representation by the expected ℓ_2 distance between $\psi(xy)$ and $\psi(x)\,\psi(y)$, where x and y are chosen uniformly from G. To control the trivial case where $\psi(x) = \mathbb{1}$ for all x, we assume that $\mathbb{E}_x\,\psi(x)$ is bounded in its operator norm. We also assume that the expected Frobenius norm squared of $\psi(x)$ is d_ψ, which holds, for example, if each $\psi(x)$ is unitary.

Our main theorem asserts that the expected ℓ_2 distance is bounded below by a function of the ratio $d_\psi / d_{\min}$. Roughly speaking, if we think of ψ as a low-dimensional embedding of G, we cannot avoid a certain amount of "distortion" of G's multiplicative structure. We let $\|A\|_{\mathrm{op}}$ denote the operator norm, and let $\|A\|_{\mathrm{F}}^2 = \operatorname{tr} A^\dagger A$ denote the Frobenius norm.

Theorem 1. *Let G be a group and let $d_{\min}$ denote the dimension of G's smallest nontrivial irrep. For any function $\psi : G \to \mathsf{U}_{d_\psi}$,*

$$\mathop{\mathbb{E}}_{x,y \in G} \|\psi(xy) - \psi(x)\,\psi(y)\|_{\mathrm{F}}^2 \geq 2d_\psi \left(1 - \|\mathop{\mathbb{E}}_{x} \psi(x)\|_{\mathrm{op}}^3 - \sqrt{\frac{d_\psi}{d_{\min}}} \right) . \tag{1}$$

If A and B are random unitary matrices of dimension d_ψ distributed according to Haar measure, then $\mathbb{E}_{A,B} \|A - B\|_{\mathrm{F}}^2 = 2d$. Thus Theorem 1 shows that when $d_\psi / d_{\min}$ and $\left\|\mathbb{E}_x \psi(x)\right\|_{\mathrm{op}}$ are small, ψ is little better than a random function from G to U_d as far as acting like a representation is concerned.

We comment that Theorem 1 holds in the more general setting where ψ is a function from G to the group GL_{d_ψ} of invertible d_ψ-dimensional matrices, as long as ψ is "unitary in expectation" in the sense that

$$\mathop{\mathbb{E}}_{x} \psi(x)^\dagger \psi(x) = \mathbb{1} . \tag{2}$$

In the regime where ψ is very close to a representation, our work is related to Babai, Friedl, and Lukács [2, 3]. They showed that if $\|\psi(xy) - \psi(x)\,\psi(y)\|_{\mathrm{F}}^2$ is sufficiently small, then there is a genuine representation ρ with $d_\rho = d_\psi$ such that ρ is close to ψ. Their definitions are slightly different; for instance, they consider uniform bounds on $\|\psi(xy) - \psi(x)\,\psi(y)\|_{\mathrm{F}}^2$ rather than its expectation over all pairs of elements $x, y \in G$, and also place a bound on $\|\psi(1) - \mathbb{1}\|_{\mathrm{F}}^2$. Nevertheless, the Fourier-analytic proof of Theorem 1 uses similar Fourier analytic techniques as in their work.

Theorem 1 yields the following corollary, bounding the probability that $\psi(xy) = \psi(x)\,\psi(y)$ for uniformly random x, y:

Corollary 1. *Let G, $d_{\min}$, and $\psi : G \to \mathsf{U}_{d_\psi}$ be as in Theorem 1. If $x, y \in G$ are uniformly random, then*

$$\Pr[\psi(xy) = \psi(x)\,\psi(y)] \leq \frac{1}{2} \left(1 + \|\mathop{\mathbb{E}}_{x} \psi(x)\|_{\mathrm{op}}^3 + \sqrt{\frac{d_\psi}{d_{\min}}} \right) .$$

When $d_\psi / d_{\min}$ is small, this is tight for a random function ψ that sends half the elements of G to $\mathbb{1}$ and the other half to $-\mathbb{1}$, where each half is chosen uniformly at random from all subsets of size $|G|/2$.

As an application of these results, we consider approximate homomorphisms $f : G \to H$ where H is another finite group, bounding the probability that $f(xy) = f(x)\,f(y)$ for uniformly random pairs $x, y \in G$. To avoid the trivial homomorphism $f(x) = 1$, we require that f's image is close to uniform. For each $y \in H$ define the probability

$$p_f(y) = \Pr_x[f(x) = y]$$

that a uniformly random $x \in G$ has image y. Then we bound the ℓ_2 distance between p_f and the uniform distribution $u(y) = 1/|H|$, requiring that

$$\|p_f - u\|_2^2 = \sum_{y \in H} \left| p_f(y) - \frac{1}{|H|} \right|^2 \le \frac{\epsilon}{|H|} . \tag{3}$$

For instance, this holds with $\epsilon = 1$ if $p_f(y)$ is uniform on a subgroup of H of index 2.

We will use the fact that if f is an approximate homomorphism then, for each irrep σ of H, the composition $\sigma \circ f$ is an approximate representation of G. Our first bound focuses on one σ at a time.

Theorem 2. *Let G and H be finite groups, and let $d_{\min}$ denote the dimension of G's smallest nontrivial irrep. Let $f : G \to H$ such that* (3) *holds. Then*

$$\Pr[f(xy) = f(x)\,f(y)] \le \frac{1}{2} \min_{\sigma \ne 1} \left(1 + \sqrt{\frac{\epsilon}{d_\sigma}} + \sqrt{\frac{d_\sigma}{d_{\min}}} \right) ,$$

where σ ranges over all of H's nontrivial irreps.

If p_f is perfectly uniform so that $\epsilon = 0$, this expression is minimized by H's smallest nontrivial irrep σ. In that case, f cannot act very homomorphically if H is much less quasirandom than G is.

Our second bound considers all of H's irreps, not just the smallest one. Recall that the *Plancherel measure* assigns each irrep $\sigma \in \widehat{H}$ the probability $P(\sigma) = d_\sigma^2/|H|$. If R_H denotes the expectation of $\sqrt{d_\sigma/d_{\min}}$ or 1, whichever is smaller,

$$R_H(d_{\min}) = \sum_{\sigma \in \widehat{H}} \frac{d_\sigma^2}{|H|} \min\left(\sqrt{\frac{d_\sigma}{d_{\min}}} , 1 \right) , \tag{4}$$

then we have the following.

Theorem 3. *Let G and H be finite groups, and let $d_{\min}$ denote the dimension of G's smallest nontrivial irrep. Let $f : G \to H$ such that* (3) *holds. Then*

$$\Pr[f(xy) = f(x)\,f(y)] \le \|p_f\|_2^2 + R_H(d_{\min}) = \frac{1+\epsilon}{|H|} + R_H(d_{\min}) .$$

If R_H and ϵ are small, i.e., if most of H's irreps are much smaller than $d_{\min}$ and f's image is close to uniform, Theorem 3 shows that f cannot act like a homomorphism much more often than a random function from G to H.

Finally, we give two results indicating that the bounds of Theorem 1 are essentially tight. First we show that they are achieved exactly if ψ is proportional to a minor of a genuine irreducible representation. While these minors are not unitary, we can scale them so that they are unitary in expectation in the sense of (2).

Theorem 4. *Let $\rho : G \to U(V)$ be an irreducible representation of G of dimension d_ρ and let $\Pi : V \to V$ be a projection operator onto a d_ψ-dimensional subspace $W \subseteq V$. If we define*

$$\psi(x) = \sqrt{\frac{d_\rho}{d_\psi}}\, \Pi \rho(x) \Pi ,$$

then $\mathbb{E}_x \psi(x) = 0$, $\mathbb{E}_x \psi(x)^\dagger \psi(x) = \Pi$, *and*

$$\mathop{\mathbb{E}}_{x,y \in G} \|\psi(xy) - \psi(x)\,\psi(y)\|_{\mathrm{F}}^2 = 2d_\psi \left(1 - \sqrt{\frac{d_\psi}{d_\rho}}\right) .$$

Note that this precisely matches our upper bound in Theorem 1 in the case $\mathbb{E}_x \psi(x) = 0$.

In our last result, we use the polar decomposition to make these approximate representations unitary. This comes at some cost to the expected Frobenius norm, but there is still a regime for d_ψ / d_ρ where we can achieve significantly stronger results than those of a random function. First recall that if A is a d-dimensional complex matrix of full rank, its *polar decomposition* expresses A as the product of a unitary matrix $\tilde{A}$ and a positive semidefinite matrix

$$A = \tilde{A} P ,$$

where

$$\tilde{A} = A(A^\dagger A)^{-1/2} \quad \text{and} \quad P = (A^\dagger A)^{1/2} .$$

That is, $\tilde{A} = AP^{-1}$ where P is the unique positive semidefinite matrix such that $P^2 = A^\dagger A$. It is a simple exercise to show that $\tilde{A}$ is unitary. More importantly, $\tilde{A}$ is the unitary matrix which is closest to A in ℓ_2 distance [7].

Then we have the following theorem. Note that unlike Theorem 4, we now assume that Π is chosen uniformly. Specifically, given a fixed projection operator Π' of rank d_ψ, we set $\Pi = U^\dagger \Pi' U$ where $U \in \mathsf{U}(V)$ is uniform according to the Haar measure.

Theorem 5. *Let $\rho : G \to U(V)$ be an irreducible representation of G of dimension d_ρ and let $\Pi : V \to V$ be a projection operator onto a subspace $W \subseteq V$ chosen uniformly from all projection operators of rank d_ψ. Let $\psi(x)$ be defined as in Theorem 4, and let $\tilde{\psi}(x)$ be the unitary part of its polar decomposition. Then*

$$\mathop{\mathbb{E}}_{\Pi} \mathop{\mathbb{E}}_{x,y \in G} \|\tilde{\psi}(xy) - \tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}^2 \le 2d_\psi \left(4\left(1 - \sqrt{\frac{d_\psi}{d_\rho}}\right) + 6\left(1 - \frac{d_\psi}{d_\rho}\right)\right) .$$

It follows that there exists a particular projection operator Π satisfying the bound above.

The difference between Theorems 4 and 5 is the cost of making $\psi(x)$ unitary—it comes from bounding the expected ℓ_2 distance between $\psi(x)$ and $\tilde{\psi}(x)$ and using the triangle inequality. While it is intuitive that this cost is nonzero, we have not attempted to optimize this bound. Nevertheless, even this relatively crude bound shows that there exist unitary approximate representations that perform noticeably better than random matrices—that is, for which

$$\frac{1}{2d_\psi} \mathop{\mathbb{E}}_{x,y \in G} \|\tilde{\psi}(xy) - \tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}^2 \le \alpha$$

for some $\alpha < 1$, whenever

$$d_\psi / d_\rho > \frac{31 - 2\sqrt{58}}{18} = 0.876\ldots$$

We conjecture that approximate representations exist with $\alpha < 1$ whenever $d_\psi / d_\rho > 0$.

Proofs are given in the following three sections.

1 Bounds on approximate representations

In this section we prove Theorem 1 and Corollary 1. In the process, we set our conventions for the nonabelian Fourier transform, and prove several inequalities that we will apply later on.

Proof of Theorem 1. For any A, B we have

$$\|A - B\|_{\mathrm{F}}^2 = \operatorname{tr}(A-B)^\dagger (A-B) = \|A\|_{\mathrm{F}}^2 + \|B\|_{\mathrm{F}}^2 - 2 \operatorname{Re} \operatorname{tr} A^\dagger B \, .$$

Since $z = xy$ is uniformly random whenever x and y are,

$$\mathop{\mathbb{E}}_{x,y} \|\psi(xy) - \psi(x)\,\psi(y)\|_{\mathrm{F}}^2 = \mathop{\mathbb{E}}_{z} \|\psi(z)\|_{\mathrm{F}}^2 + \mathop{\mathbb{E}}_{x,y} \|\psi(x)\,\psi(y)\|_{\mathrm{F}}^2 - 2 \operatorname{Re} \mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi(xy)^\dagger\, \psi(x)\,\psi(y) \, . \tag{5}$$

If ψ is unitary in expectation, (2) implies

$$\mathop{\mathbb{E}}_{z} \|\psi(z)\|_{\mathrm{F}}^2 = \operatorname{tr} \mathop{\mathbb{E}}_{z} \psi(z)^\dagger \psi(z) = \operatorname{tr} \mathbb{1} = d_\psi \, , \tag{6}$$

and of course this holds identically if ψ is unitary. Similarly,

$$\begin{aligned} \mathop{\mathbb{E}}_{x,y} \|\psi(x)\,\psi(y)\|_{\mathrm{F}}^2 &= \operatorname{tr} \mathop{\mathbb{E}}_{x,y} \psi(y)^\dagger \psi(x)^\dagger \psi(x) \psi(y) \\ &= \operatorname{tr} \left(\mathop{\mathbb{E}}_{x} \psi(x)^\dagger \psi(x) \right) \left(\mathop{\mathbb{E}}_{y} \psi(y) \psi(y)^\dagger \right) = \operatorname{tr} \mathbb{1} = d_\psi \, . \end{aligned}$$

Then (5) becomes

$$\mathop{\mathbb{E}}_{x,y} \|\psi(xy) - \psi(x)\,\psi(y)\|_{\mathrm{F}}^2 = 2 d_\psi \left(1 - \frac{1}{d_\psi} \operatorname{Re} \mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) \right) \, . \tag{7}$$

Thus we will focus on estimating the expected trace

$$\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) \, . \tag{8}$$

Note that if ψ is a genuine representation, $\psi^\dagger(xy)\, \psi(x)\, \psi(y) = \mathbb{1}$ and this trace is identically d_ψ.

We rely on nonabelian Fourier analysis, for which we refer the reader to [4]. In order to establish our notation and choice of normalizations, let $f : G \to \mathbb{C}$ and let $\rho : G \to \mathsf{U}_d$ be an irreducible unitary representation of G or "irrep" for short, and let $\widehat{G}$ denote the set of irreps of G. We adopt the Fourier transform

$$\widehat{f}(\rho) = \frac{1}{|G|} \sum_{x \in G} f(x)\, \rho^\dagger(x) = \mathop{\mathbb{E}}_{x} f(x)\, \rho^\dagger(x) \, ,$$

in which case we have the Fourier inversion formula

$$f(x) = \sum_{\rho \in \widehat{G}} d_\rho \operatorname{tr} \left(\widehat{f}(\rho)\, \rho(x) \right) \, .$$

The Fourier transform preserves inner products in the sense that

$$\langle f, g \rangle = \sum_x f(x)^* g(x) = |G| \sum_\rho d_\rho \operatorname{tr} \left(\widehat{f}(\rho)^\dagger\, \widehat{g}(\rho) \right) \, . \tag{9}$$

In particular, we have Plancherel's identity,

$$\|f\|_2^2 = \sum_x |f(x)|^2 = |G| \sum_\rho d_\rho \big\|\widehat{f}(\rho)\big\|_{\mathrm{F}}^2 . \tag{10}$$

Since ψ is a matrix-valued function, each entry $\psi(x)^i_j$ has its own Fourier transform. Therefore, we may treat the Fourier transform $\widehat{\psi}$ as a tensor with four indices,

$$\widehat{\psi}(\rho)^{ik}_{j\ell} = \mathop{\mathbb{E}}_x \psi(x)^i_j\, \rho^\dagger(x)^k_\ell \quad \text{or} \quad \widehat{\psi}(\rho) = \mathop{\mathbb{E}}_x \left[\psi(x) \otimes \rho^\dagger(x)\right] .$$

The Fourier inversion formula can then be expressed as a partial trace. We adopt the Einstein summation convention, where any index appearing twice is automatically summed over. For instance, $(AB)^i_j = A^i_k B^k_j$ and $\operatorname{tr} A = A^i_i$. Then

$$\psi(x)^i_j = \sum_\rho d_\rho\, \widehat{\psi}(\rho)^{ik}_{j\ell}\, \rho(x)^\ell_k .$$

Plancherel's identity becomes

$$\mathop{\mathbb{E}}_x \|\psi(x)\|_{\mathrm{F}}^2 = \sum_\rho d_\rho \,\big\|\widehat{\psi}(\rho)\big\|_{\mathrm{F}}^2 = \sum_\rho d_\rho \sum_{i,j,k,l} \left|\widehat{\psi}(\rho)^{ik}_{j\ell}\right|^2 . \tag{11}$$

We compute the trace (8) by evaluating it in the Fourier basis. First write

$$\begin{aligned}
\operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) &= \psi^\dagger(xy)^i_j\, \psi(x)^j_k\, \psi(y)^k_i \\
&= \sum_{\rho,\sigma,\tau \in \widehat{G}} d_\rho\, d_\sigma\, d_\tau\, \widehat{\psi^\dagger}(\rho)^{ia}_{jb}\, \widehat{\psi}(\sigma)^{jd}_{ke}\, \widehat{\psi}(\tau)^{kf}_{ig}\, \rho(xy)^b_a\, \sigma(x)^e_d\, \tau(y)^g_k \\
&= \sum_{\rho,\sigma,\tau \in \widehat{G}} d_\rho\, d_\sigma\, d_\tau\, \widehat{\psi^\dagger}(\rho)^{ia}_{jb}\, \widehat{\psi}(\sigma)^{jd}_{ke}\, \widehat{\psi}(\tau)^{kf}_{ig}\, \rho(x)^b_c\, \rho(y)^c_a\, \sigma(x)^e_d\, \tau(y)^g_f .
\end{aligned} \tag{12}$$

Schur's lemma implies

$$\mathop{\mathbb{E}}_x \left[\rho(x)^b_c\, \sigma(x)^e_d\right] = \frac{1}{d_\rho} \begin{cases} \delta^{be}\, \delta_{cd} & \sigma = \rho^* \\ 0 & \sigma \neq \rho^* . \end{cases} \tag{13}$$

Thus taking the expectation over x and y turns (12) into

$$\begin{aligned}
\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) &= \sum_\rho d_\rho\, \widehat{\psi^\dagger}(\rho)^{ia}_{jb}\, \widehat{\psi}(\rho^*)^{jd}_{ke}\, \widehat{\psi}(\rho^*)^{kf}_{ig}\, \delta^{be}\, \delta_{cd}\, \delta^{cg}\, \delta_{af} \\
&= \sum_\rho d_\rho\, \widehat{\psi^\dagger}(\rho)^{ia}_{jb}\, \widehat{\psi}(\rho^*)^{jd}_{kb}\, \widehat{\psi}(\rho^*)^{ka}_{id} .
\end{aligned}$$

We rearrange this slightly, writing $\widehat{\psi}(\rho^\dagger)^{ik}_{j\ell}$ for the partial transpose $\widehat{\psi}(\rho^*)^{i\ell}_{jk}$. Then

$$\widehat{\psi}(\rho^\dagger) = \mathop{\mathbb{E}}_x \left[\psi(x) \otimes \rho(x)\right] , \tag{14}$$

and

$$\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) = \sum_\rho d_\rho\, \widehat{\psi^\dagger}(\rho)^{ia}_{jb}\, \widehat{\psi}(\rho^\dagger)^{jb}_{kd}\, \widehat{\psi}(\rho^\dagger)^{kd}_{ia} . \tag{15}$$

If we view $\widehat{\psi}(\rho^\dagger)$ as a linear operator on $\mathbb{C}^d \otimes \mathbb{C}^{d_\rho}$, then

$$\widehat{\psi^\dagger}(\rho) = \left(\widehat{\psi}(\rho^\dagger)\right)^\dagger = \mathop{\mathbb{E}}_x \left[\psi^\dagger(x) \otimes \rho^\dagger(x)\right] ,$$

and we can write

$$\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) = \sum_\rho d_\rho \operatorname{tr}\, \widehat{\psi}(\rho^\dagger)\, \left(\widehat{\psi}(\rho^\dagger)\right)^\dagger\, \widehat{\psi}(\rho^\dagger)\,. \tag{16}$$

We remark that in the case where ψ is an irrep the only irrep contributing to the sum (16) is ψ^* since, as in (13), we have

$$\widehat{\psi}(\rho^\dagger)^{ik}_{j\ell} = \mathop{\mathbb{E}}_x \left[\psi(x)^i_j\, \rho(x)^k_\ell\right] = \frac{1}{d_\psi} \begin{cases} \delta^{ik}\, \delta_{j\ell} & \rho = \psi^*\,, \\ 0 & \rho \neq \psi^*\,. \end{cases}$$

Let Π denote the operator $(1/d_\psi)\, \delta^{ik}\, \delta_{j\ell}$. Diagrammatically, Π is proportional to the "cupcap." It is a one-dimensional projection operator, equal to the outer product of the vector

$$(1/\sqrt{d_\psi}) \sum_i e_i \otimes e_i$$

with itself, where e_i denotes the ith basis vector. Since Π is Hermitian, (16) implies that

$$\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) = d_\psi \operatorname{tr} \Pi^3 = d_\psi \operatorname{tr} \Pi = d_\psi$$

which holds since $\psi^\dagger(xy)\, \psi(x)\, \psi(y) = \mathbb{1}$.

Returning to (16), since $A^\dagger A$ is positive for any A we have

$$\operatorname{tr} A A^\dagger A \leq \|A\|_{\mathrm{op}}\, \|A\|_{\mathrm{F}}^2\,.$$

This gives

$$\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) \leq \sum_\rho d_\rho\, \|\widehat{\psi}(\rho^\dagger)\|_{\mathrm{op}}\, \|\widehat{\psi}(\rho^\dagger)\|_{\mathrm{F}}^2\,. \tag{17}$$

We separate out the term corresponding to the trivial representation $\rho = 1$, for which $\widehat{\psi}(1) = \mathbb{E}_x\, \psi(x)$. Since $\|A\|_{\mathrm{F}}^2 \leq d\, \|A\|_{\mathrm{op}}^2$ for any d-dimensional matrix A, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) &\leq \|\mathop{\mathbb{E}}_x \psi(x)\|_{\mathrm{op}}\, \|\mathop{\mathbb{E}}_x \psi(x))\|_{\mathrm{F}}^2 + \sum_{\rho \neq 1} d_\rho\, \|\widehat{\psi}(\rho^\dagger)\|_{\mathrm{op}}\, \|\widehat{\psi}(\rho^\dagger)\|_{\mathrm{F}}^2 && (18) \\
&\leq d_\psi \|\mathop{\mathbb{E}}_x \psi(x)\|_{\mathrm{op}}^3 + \left(\max_{\rho \neq 1} \|\widehat{\psi}(\rho^\dagger)\|_{\mathrm{op}}\right) \sum_\rho d_\rho\, \|\widehat{\psi}(\rho^\dagger)\|_{\mathrm{F}}^2 \\
&= d_\psi \|\mathop{\mathbb{E}}_x \psi(x)\|_{\mathrm{op}}^3 + \left(\max_{\rho \neq 1} \|\widehat{\psi}(\rho^\dagger)\|_{\mathrm{op}}\right) \mathop{\mathbb{E}}_x \|\psi(x)\|_{\mathrm{F}}^2 \\
&= d_\psi \left(\|\mathop{\mathbb{E}}_x \psi(x)\|_{\mathrm{op}}^3 + \max_{\rho \neq 1} \|\widehat{\psi}(\rho^\dagger)\|_{\mathrm{op}}\right), && (19)
\end{aligned}$$

where we used Plancherel's identity (11) in the third line and the fact that $\mathbb{E}_x \big\| \psi(x) \big\|_{\mathrm{F}}^2 = d_\psi$ is unitary, or unitary in expectation, in the fourth. This is analogous to the Fourier-analytic treatment of the Blum-Luby-Rubinfeld linearity test [5, 6].

Our next goal is to bound the operator norm of $\widehat{\psi}(\rho^\dagger)$. Let V and W denote the spaces on which ψ and ρ act, respectively. Then $\big\| \widehat{\psi}(\rho^\dagger) \big\|_{\mathrm{op}}$ is the maximum, taken over all vectors $u \in V \otimes W$ of norm 1, of $\langle u, \widehat{\psi}(\rho^\dagger)\, u \rangle$. Using the Schmidt decomposition we can write

$$u = \sum_i \alpha_i v_i \otimes w_i$$

where $\{v_i\}$ and $\{w_i\}$ are orthogonal bases for V and a d_ψ-dimensional subspace of W respectively, and where $\sum_i |\alpha_i|^2 = 1$. Then separating the tensor product and using Cauchy-Schwarz gives

$$\begin{aligned}
\Big\langle u, \widehat{\psi}(\rho^\dagger)\, u \Big\rangle &= \sum_{i,j} \alpha_i^* \alpha_j \Big\langle v_i \otimes w_i, \widehat{\psi}(\rho^\dagger)\, v_j \otimes w_j \Big\rangle \\
&= \sum_{i,j} \alpha_i^* \alpha_j \Big\langle v_i \otimes w_i, \Big(\mathop{\mathbb{E}}_x \left[\psi(x) \otimes \rho(x) \right] \Big)\, v_j \otimes w_j \Big\rangle \\
&= \sum_{i,j} \alpha_i^* \alpha_j \mathop{\mathbb{E}}_x \left[\langle v_i, \psi(x)\, v_j \rangle \langle w_i, \rho(x)\, w_j \rangle \right] \\
&\le \sum_{i,j} \alpha_i^* \alpha_j \sqrt{\Big(\mathop{\mathbb{E}}_x \langle v_i, \psi(x)\, v_j \rangle^2 \Big) \Big(\mathop{\mathbb{E}}_x \langle w_j, \rho(x)\, w_j \rangle^2 \Big)}\,.
\end{aligned}$$

By Schur's lemma we have $\mathbb{E}_x \langle w_i, \rho(x)\, w_j \rangle^2 = 1/d_\rho$, giving

$$\Big\langle u, \widehat{\psi}(\rho^\dagger)\, u \Big\rangle \le \frac{1}{\sqrt{d_\rho}} \sum_{i,j} \alpha_i^* \alpha_j \sqrt{\mathop{\mathbb{E}}_x \langle v_i, \psi(x)\, v_j \rangle^2}\,.$$

Another application of Cauchy-Schwarz and the fact that $\| \psi(x) \|_{\mathrm{F}}^2 = d_\psi$ gives

$$\begin{aligned}
\Big\langle u, \widehat{\psi}(\rho^\dagger)\, u \Big\rangle &\le \frac{1}{\sqrt{d_\rho}} \sqrt{\Big(\sum_{i,j} |\alpha_i|^2 |\alpha_j|^2 \Big) \sum_{ij} \mathop{\mathbb{E}}_x \langle v_i, \psi(x)\, v_j \rangle^2} \\
&= \frac{1}{\sqrt{d_\rho}} \sqrt{\Big(\sum_i |\alpha_i|^2 \Big)^2 \mathop{\mathbb{E}}_x \Big[\sum_{i,j} \langle v_i, \psi(x)\, v_j \rangle^2 \Big]} \\
&= \frac{1}{\sqrt{d_\rho}} \sqrt{\mathop{\mathbb{E}}_x \| \psi(x) \|_{\mathrm{F}}^2} \\
&= \sqrt{\frac{d_\psi}{d_\rho}}\,,
\end{aligned} \tag{20}$$

where we again used the fact that $\psi(x)$ is unitarity, or unitary in expectation, in the fourth line.

Combining (19) with (20) and using our hypothesis that $\min_{\rho \neq 1} d_\rho = d_{\min}$ then gives

$$\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\, \psi(x)\, \psi(y) \le d_\psi \left(\big\| \mathop{\mathbb{E}}_x \psi(x) \big\|_{\mathrm{op}}^3 + \sqrt{\frac{d_\psi}{d_{\min}}} \right)\,. \tag{21}$$

Finally, combining this with (7) completes the proof. □

Proof of Corollary 1. For any $A, B \in \mathsf{U}_d$ with $A \neq B$ we have $\|A - B\|_{\mathrm{F}}^2 \leq 4d$. Thus

$$\mathop{\mathbb{E}}_{x,y} \|\psi(xy) - \psi(x)\, \psi(y)\|_{\mathrm{F}}^2 \leq 4d_\psi \Pr[\psi(xy) \neq \psi(x)\, \psi(y)]\,,$$

and so

$$\Pr[\psi(xy) = \psi(x)\, \psi(y)] \leq 1 - \frac{1}{4d_\psi} \mathop{\mathbb{E}}_{x,y} \|\psi(xy) - \psi(x)\, \psi(y)\|_{\mathrm{F}}^2\,.$$

Combining this with the bound (1) completes the proof. □

2 Approximate homomorphisms

In this section we prove Theorems 2 and 3, bounding the extent to which a function from one finite group to another can act like a homomorphism.

Proof of Theorem 2. Let σ be an irreducible representation of H. We treat $\psi_\sigma = \sigma \circ f$ as an approximate representation of G of dimension d_σ. To bound $\left\|\mathbb{E}_x\, \psi(x)\right\|_{\mathrm{op}}$, note that

$$\mathop{\mathbb{E}}_{x} \psi_\sigma(x) = \sum_{y \in H} p_f(y)\, \sigma(y) = \sum_{y \in H} (p_f - u)(y)\, \sigma(y) = |H|\, \widehat{(p_f - u)}(\sigma)\,.$$

where we used the fact that $\mathbb{E}_y\, \sigma(y) = 0$. Then we have

$$\left\|\mathop{\mathbb{E}}_{x} \psi_\sigma(x)\right\|_{\mathrm{op}}^2 \leq \left\|\mathop{\mathbb{E}}_{x} \psi_\sigma(x)\right\|_{\mathrm{F}}^2 \leq |H|^2 \left\|\widehat{(p_f - u)}(\sigma)\right\|_{\mathrm{F}}^2 \leq \frac{|H|}{d_\sigma} \left\|p_f - u\right\|_2^2 \leq \frac{\epsilon}{d_\sigma}\,,$$

where we used Plancherel's identity (10) in the third inequality. Thus

$$\left\|\mathop{\mathbb{E}}_{x} \psi_\sigma(x)\right\|_{\mathrm{op}} \leq \sqrt{\frac{\epsilon}{d_\sigma}}\,, \tag{22}$$

and applying Corollary 1 completes the proof. □

Proof of Theorem 3. We let $R = \sum_{\sigma \in \hat{H}} d_\sigma \chi_\sigma$ denote the regular representation of H. As above, for an irreducible representation σ of H we define $\psi_\sigma = \sigma \circ f$. Observe that

$$\begin{aligned}
\Pr_{x,y}[f(xy) = f(x)\, f(y)] &= \frac{1}{|H|} \mathop{\mathbb{E}}_{x,y}[R(f(xy)^{-1} f(x) f(y))] = \sum_{\sigma \in \widehat{H}} \frac{d_\sigma}{|H|} \mathop{\mathbb{E}}_{x,y}\left[\chi_\sigma\big(f(xy)^{-1}\, f(x)\, f(y)\big)\right] \\
&= \sum_{\sigma \in \widehat{H}} \frac{d_\sigma}{|H|} \mathop{\mathbb{E}}_{x,y}\left[\operatorname{tr} \psi_\sigma^\dagger(xy)\, \psi_\sigma(x)\, \psi_\sigma(y)\big)\right] \\
&\leq \frac{1}{|H|} + \underbrace{\sum_{\sigma \neq 1} \frac{d_\sigma}{|H|} \left\|\mathop{\mathbb{E}}_{x} \psi_\sigma(x)\right\|_{\mathrm{op}} \left\|\mathop{\mathbb{E}}_{x} \psi_\sigma(x)\right\|_{\mathrm{F}}^2}_{(*)} + \sum_{\sigma \neq 1} \frac{d_\sigma^2}{|H|} \min\left(\sqrt{\frac{d_\sigma}{d_{\min}}}\,,\, 1\right)\,,
\end{aligned}$$

the last inequality following from equation (18). Focusing on the term $(*)$ above,

$$\begin{aligned}\sum_{\sigma\neq 1} \frac{d_\sigma}{|H|} \left\| \mathop{\mathbb{E}}_x \psi_\sigma(x) \right\|_{\mathrm{op}} \left\| \mathop{\mathbb{E}}_x \psi_\sigma(x) \right\|_{\mathrm{F}}^2 &\leq \sum_{\sigma\neq 1} \frac{d_\sigma}{|H|} \left\| \mathop{\mathbb{E}}_x \psi_\sigma(x) \right\|_{\mathrm{F}}^2 \\ &= \frac{\left\| \mathbb{E}_x R(f(x)) \right\|_{\mathrm{F}}^2 - 1}{|H|} \\ &= \frac{\mathrm{tr}\left(\mathbb{E}_x \mathbb{E}_y R(f(x)^{-1}) R(f(y)) \right)}{|H|} - \frac{1}{|H|} \\ &= \Pr_{x,y}[f(x) = f(y)] - \frac{1}{|H|}\,.\end{aligned}$$

Hence

$$\begin{aligned}\Pr_{x,y}[f(xy) = f(x)\,f(y)] &\leq \Pr_{x,y}[f(x) = f(y)] + \sum_{\sigma\neq 1} \frac{d_\sigma^2}{|H|} \min\left(\sqrt{\frac{d_\sigma}{d_{\min}}}, 1 \right) \\ &\leq \Pr_{x,y}[f(x) = f(y)] + R_H\,.\end{aligned}$$

Finally, since

$$\Pr[f(x) = f(y)] = \sum_{x\in H} p_f(x)^2 = \left\| p_f \right\|_2^2 = \|u\|_2^2 + \left\| p_f - u \right\|_2^2 \leq \frac{1+\epsilon}{|H|}\,,$$

the statement of the theorem follows. □

3 Minors of representations and their polar decompositions

In this section we prove Theorems 4 and 5.

Proof of Theorem 4. Let $\rho : G \to \mathsf{U}(V)$ be an irreducible representation of G of dimension d_ρ and let $\Pi : V \to V$ be a projection operator of rank d_ψ. Treating the image of Π as a subspace W, we consider the function $\psi : G \to \mathrm{End}(W)$ given by

$$\psi(x) = \sqrt{\frac{d_\rho}{d_\psi}}\, \Pi \rho(x) \Pi\,.$$

Then

$$\mathop{\mathbb{E}}_x \psi(x) = \sqrt{\frac{d_\psi}{d_\rho}}\, \Pi \left(\mathop{\mathbb{E}}_x \rho(x) \right) \Pi = 0\,.$$

Moreover, since ρ is irreducible, Schur's lemma gives

$$\mathop{\mathbb{E}}_x \rho(x)^\dagger \Pi \rho(x) = \frac{\mathbf{rk}\,\Pi}{d_\rho} \mathbb{1} = \frac{d_\psi}{d_\rho} \mathbb{1}\,. \tag{23}$$

Thus

$$\mathop{\mathbb{E}}_x \psi(x)^\dagger \psi(x) = \frac{d_\rho}{d_\psi} \mathop{\mathbb{E}}_x \Pi \rho(x)^\dagger \Pi \rho(x) \Pi = \frac{d_\rho}{d_\psi} \Pi \left(\mathop{\mathbb{E}}_x \rho(x)^\dagger \Pi \rho(x) \right) \Pi = \Pi\,.$$

Since Π is the identity on the subspace W, $\psi(x)$ is unitary in expectation.

As in the proof of Theorem 1, we then have

$$\mathop{\mathbb{E}}_{x,y} \|\psi(xy) - \psi(x)\,\psi(y)\|_{\mathrm{F}}^2 = 2d_\psi \left(1 - \frac{1}{d_\psi} \operatorname{Re} \mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\,\psi(x)\,\psi(y)\right) . \tag{24}$$

Since ρ is a genuine representation, $\rho(xy) = \rho(x)\,\rho(y)$ and

$$\begin{aligned}
\mathop{\mathbb{E}}_{x,y} \operatorname{tr} \psi^\dagger(xy)\,\psi(x)\,\psi(y) &= \left(\frac{d_\rho}{d_\psi}\right)^{3/2} \mathop{\mathbb{E}}_{x,y} \operatorname{tr} \Pi\rho^\dagger(xy)\Pi\rho(x)\Pi\rho(y)\Pi \\
&= \left(\frac{d_\rho}{d_\psi}\right)^{3/2} \mathop{\mathbb{E}}_{x,y} \operatorname{tr} \Pi\rho(y)^\dagger\rho(x)^\dagger\Pi\rho(x)\Pi\rho(y)\Pi \\
&= \left(\frac{d_\rho}{d_\psi}\right)^{3/2} \operatorname{tr} \left[\left(\mathop{\mathbb{E}}_{y} \rho(y)\Pi\rho(y)^\dagger \right) \mathop{\mathbb{E}}_{x} \left(\rho(x)^\dagger\Pi\rho(x) \right) \Pi \right] \\
&= \sqrt{\frac{d_\psi}{d_\rho}} \operatorname{tr} \Pi = d_\psi \sqrt{\frac{d_\psi}{d_\rho}} \,.
\end{aligned}$$

Combining this with (24) completes the proof. □

Proof of Theorem 5. The squared ℓ_2 distance between a matrix A and the unitary part of its polar decomposition, $\tilde{A} = A(A^\dagger A)^{-1/2}$, is

$$\|A - \tilde{A}\|_{\mathrm{F}}^2 = \left\|A - A(A^\dagger A)^{-1/2}\right\|_{\mathrm{F}}^2 = \operatorname{tr} A^\dagger A - 2\operatorname{tr}(A^\dagger A)^{1/2} + d = \sum_\lambda (\lambda - 1)^2 \,.$$

Here λ ranges over the singular values of A, i.e., the square roots of the eigenvalues of $A^\dagger A$. For any $\lambda \geq 0$ we have

$$(\lambda - 1)^2 \leq (\lambda - 1)^2(\lambda + 1)^2 = (\lambda^2 - 1)^2 \,. \tag{25}$$

Thus the distance between A and $\tilde{A}$ is at most the distance between $A^\dagger A$ and the identity,

$$\|A - \tilde{A}\|_{\mathrm{F}}^2 \leq \sum_\lambda (\lambda^2 - 1)^2 = \left\|A^\dagger A - \mathbb{1}\right\|_{\mathrm{F}}^2 \,. \tag{26}$$

Let $\psi(x) = \sqrt{d_\rho/d_\psi}\,\Pi\rho(x)\Pi$. Since Π is the identity on the subspace W, the rest of our proof consists of bounding

$$\begin{aligned}
\mathop{\mathbb{E}}_{\Pi,x} \|\psi(x) - \tilde{\psi}(x)\|_{\mathrm{F}}^2 &\leq \mathop{\mathbb{E}}_{\Pi,x} \left\|\psi(x)^\dagger\psi(x) - \Pi\right\|_{\mathrm{F}}^2 \\
&= \mathop{\mathbb{E}}_{\Pi,x} \left\|\psi(x)^\dagger\psi(x)\right\|_{\mathrm{F}}^2 - 2\operatorname{tr} \mathop{\mathbb{E}}_{x} \psi(x)^\dagger\psi(x) + d_\psi \\
&= \mathop{\mathbb{E}}_{\Pi,x} \left\|\psi(x)^\dagger\psi(x)\right\|_{\mathrm{F}}^2 - d_\psi \,,
\end{aligned} \tag{27}$$

where in the last line we used the fact, proved in Theorem 4, that ψ is unitary in expectation. We will then use the triangle inequality to bound

$$\mathop{\mathbb{E}}_{\Pi} \mathop{\mathbb{E}}_{x,y} \|\tilde{\psi}(xy) - \tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}^2 \,.$$

We write

$$\mathop{\mathbb{E}}_{\Pi,x}\left\|\psi(x)^\dagger\psi(x)\right\|_{\mathrm{F}}^2=\left(\frac{d_\rho}{d_\psi}\right)^2\operatorname{tr}\mathop{\mathbb{E}}_{\Pi,x}\left[\Pi\rho^\dagger(x)\Pi\rho(x)\Pi\rho^\dagger(x)\Pi\rho(x)\right]\,. \tag{28}$$

We can view this trace as a contraction of two tensors. One is $\rho\otimes\rho^\dagger\otimes\rho\otimes\rho^\dagger$. Since we can take the expectation over all Π by conjugating a particular Π by a random unitary $U\in\mathsf{U}_{d_\rho}$, the other is the "twirl" of $\Pi^{\otimes 4}$, namely

$$\Upsilon=\mathop{\mathbb{E}}_{U}(U^\dagger\Pi U)^{\otimes 4}\,.$$

Since Υ commutes with the diagonal action of $U(d_\rho)$, it is a member of the commutant, and hence an element of the group algebra $\mathbb{C}[S_4]$. Thus we can write

$$\Upsilon=\sum_{\pi\in S_4}v(\pi)\cdot\pi\,,$$

where we identify each $\pi\in S_4$ with its action on $V^{\otimes 4}$. Moreover, since Υ commutes with any $\pi\in S_4$ the coefficients $v(\pi)$ form a *class function*: v is constant on each conjugacy class and lies in the linear span of the characters of S_4.

We can compute the coefficients $v(\pi)$ as follows. For any permutation $\sigma\in S_4$, we have

$$\operatorname{tr}T\sigma=\operatorname{tr}\Pi^{\otimes 4}\sigma=d_\psi^{c(\sigma)}\,,$$

where $c(\sigma)$ is the number of cycles in σ. For any $\lambda\in\widehat{S}_4$, the inner product of the character χ_λ with the function $d^{c(\cdot)}$ is

$$\left\langle\chi_\lambda,d^{c(\cdot)}\right\rangle=T(\lambda,d)\,,$$

where $T(\lambda,d)$ denotes the number of semistandard tableaux of shape λ and content in $\{1,\dots,d\}$. The multiplicity of λ in $(\mathbb{C}^d)^{\otimes 4}$ is also $T(\lambda,d)$, so taking traces gives

$$v(\pi)=\sum_{\lambda\in\widehat{S}_4}d_\lambda\chi_\lambda(\pi)\frac{T(\lambda,d_\psi)}{T(\lambda,d_\rho)}\,.$$

A somewhat lengthy calculation gives the following coefficients for each of the five conjugacy classes in S_4:

$$\begin{aligned}
v(1)&=\left(\frac{d_\psi}{d_\rho}\right)^4+O(d_\rho^{-2})\\
v((12))&=\frac{1}{d_\rho}\left(\frac{d_\psi}{d_\rho}\right)^3\left(1-\frac{d_\psi}{d_\rho}\right)+O(d_\rho^{-3})\\
v((123))&=\frac{1}{d_\rho^2}\left(\frac{d_\psi}{d_\rho}\right)^2\left(2\,\frac{d_\psi}{d_\rho}-1\right)\left(1-\frac{d_\psi}{d_\rho}\right)+O(d_\rho^{-4})\\
v((12)(34))&=\frac{1}{d_\rho^2}\left(\frac{d_\psi}{d_\rho}\right)^2\left(1-\frac{d_\psi}{d_\rho}\right)^2+O(d_\rho^{-4})\\
v((1234))&=\frac{1}{d_\rho^3}\left(5\left(\frac{d_\psi}{d_\rho}\right)^2-5\,\frac{d_\psi}{d_\rho}+1\right)\left(1-\frac{d_\psi}{d_\rho}\right)\frac{d_\psi}{d_\rho}+O(d_\rho^{-5})
\end{aligned}$$

Note that $v(\pi)$ scales as $d_\rho^{-t(\pi)}$ where $t(\pi) = 4 - c(\pi)$ is the transposition distance, i.e., the minimum number of transpositions whose product gives x.

When π is the identity, one of the pairs of transpositions (12)(34), and four of the 3-cycles (123), we get $\rho(x)\,\rho^\dagger(x)\,\rho(x)\,\rho^\dagger(x) = \mathbb{1}$, contributing d_ρ to the trace. Two of the six transpositions (12) contract with $\rho \otimes \rho^\dagger \otimes \rho \otimes \rho^\dagger$ to give $\rho(x)\,\rho^\dagger(x) \otimes \rho(x)\,\rho^\dagger(x) = \mathbb{1} \otimes \mathbb{1}$, which has trace d_ρ^2. These are the leading terms, and we get

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\Pi,x} \operatorname{tr}\left[\Pi\rho^\dagger(x)\Pi\rho(x)\Pi\rho^\dagger(x)\Pi\rho(x)\right] \\
&\qquad = \big(v(1) + v((12)(34)) + 4v((123))\big)\, d_\rho + 2v((12))\, d_\rho^2 + O(d_\rho^{-1}) + \text{other terms} \\
&\qquad = d_\rho \left(\frac{d_\psi}{d_\rho}\right)^3 \left(2 - \frac{d_\psi}{d_\rho}\right) + O(d_\rho^{-1}) + \text{other terms}\,,
\end{aligned}$$

where here and in the sequel $O(\cdot)$ refers to the limit $d_\rho \to \infty$ while d_ψ / d_ρ stays constant.

To bound the other terms, let m denote the total multiplicity of irreducible representations appearing in the decomposition of $\rho \otimes \rho^*$. Then

$$\mathop{\mathbb{E}}_{x\in G} \left|\chi_\rho(x)\right|^4 = m \quad \text{and} \quad \mathop{\mathbb{E}}_{x\in G} \left|\chi_\rho(x^2)\right|^2 \le m\,.$$

The first of these follows from Schur's lemma, since for any irrep τ we have $\mathbb{E}_{x\in G} \left|\chi_\tau(x)\right|^2 = 1$. The second follows from the *Frobenius-Schur indicator*, which for any irrep τ is

$$\mathop{\mathbb{E}}_{x\in G} \chi_\tau(x^2) = \begin{cases} +1 & \text{if } \tau \text{ is real,} \\ 0 & \text{if } \tau \text{ is complex,} \\ -1 & \text{if } \tau \text{ is quaternionic.} \end{cases}$$

The other terms include contractions such as $(\rho\rho^\dagger\rho) \otimes \rho^\dagger = \rho \otimes \rho^\dagger$, $(\rho\rho^\dagger) \otimes \rho \otimes \rho^\dagger = \mathbb{1} \otimes \rho \otimes \rho^\dagger$, and so on. These terms scale as

$$\begin{gathered}
\big(4v((12)) + 4v((1234))\big) \mathop{\mathbb{E}}_{x} \left|\chi_\rho(x)\right|^2 = O(d_\rho^{-1}) \\
4v((123))\, d_\rho \mathop{\mathbb{E}}_{x} \left|\chi_\rho(x)\right|^2 = O(d_\rho^{-1}) \\
v((1234)) \mathop{\mathbb{E}}_{x} \left|\chi_\rho(x)\right|^4 = O(d_\rho^{-3} m) \\
v((1234)) \mathop{\mathbb{E}}_{x} \left|\chi_\rho(x^2)\right|^2 = O(d_\rho^{-3} m) \\
2v((12)(34)) \mathop{\mathbb{E}}_{x} \chi_\rho(x^2)^* \chi_\rho(x)^2 = O(d_\rho^{-2}) \sqrt{\mathop{\mathbb{E}}_{x} \left|\chi_\rho(x^2)\right|^2 \mathop{\mathbb{E}}_{x} \left|\chi_\rho(x)\right|^4} = O(d_\rho^{-2} m)\,.
\end{gathered} \tag{29}$$

If G is quasirandom, with $d_{\min}$ the dimension of its smallest nontrivial irrep, then

$$m \le 1 + \frac{d_\rho^2 - 1}{d_{\min}} \le d_\rho^2$$

since $\rho \otimes \rho^*$ contains exactly one copy of the trivial irrep. Even if we content ourselves with the generous bound $m \le d_\rho^2$, the largest error term in (29) is $O(1)$. Thus

$$\mathop{\mathbb{E}}_{\Pi,x} \operatorname{tr}\left[\Pi\rho^\dagger(x)\Pi\rho(x)\Pi\rho^\dagger(x)\Pi\rho(x)\right] = d_\rho \left(\frac{d_\psi}{d_\rho}\right)^3 \left(2 - \frac{d_\psi}{d_\rho}\right) + O(1)\,.$$

Combining this with (28) gives

$$\left\|\psi(x)^\dagger\psi(x)\right\|_{\mathrm{F}}^2 = d_\psi\left(2-\frac{d_\psi}{d_\rho}\right)+O(1)\,,$$

and so (27) gives

$$\mathop{\mathbb{E}}_{\Pi,x}\|\psi(x)-\tilde{\psi}(x)\|_{\mathrm{F}}^2 \le d_\psi\left(1-\frac{d_\psi}{d_\rho}\right)+O(1)\,. \tag{30}$$

Finally, we return to our task of bounding

$$\mathbb{E}\,\|\tilde{\psi}(xy)-\tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}^2\,.$$

For this purpose, we use the triangle inequality and write

$$\begin{aligned}\|\tilde{\psi}(xy)-\tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}} \le{}& \|\tilde{\psi}(xy)-\psi(xy)\|_{\mathrm{F}} && (31)\\ &+\|\psi(xy)-\psi(x)\,\psi(y)\|_{\mathrm{F}} && (32)\\ &+\|\psi(x)\,\psi(y)-\tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}\,. && (33)\end{aligned}$$

In expectation, the squares of the terms appearing in (31) and (32) are precisely the topic of (30) and Theorem 4, respectively. As for the quantity (33), we may further expand it as

$$\begin{aligned}\|\psi(x)\,\psi(y)-\tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}} &\le \|\psi(x)\,\psi(y)-\tilde{\psi}(x)\,\psi(y)\|_{\mathrm{F}}+\|\tilde{\psi}(x)\,\psi(y)-\tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}\\ &= \left\|\big(\psi(x)-\tilde{\psi}(x)\big)\psi(y)\right\|_{\mathrm{F}}+\left\|\tilde{\psi}(x)\big(\psi(y)-\tilde{\psi}(y)\big)\right\|_{\mathrm{F}}\\ &= \left\|\big(\psi(x)-\tilde{\psi}(x)\big)\psi(y)\right\|_{\mathrm{F}}+\|\psi(y)-\tilde{\psi}(y)\|_{\mathrm{F}}\,,\end{aligned}$$

where in the last line we used the unitarity of $\tilde{\psi}(x)$. Squaring both sides and using the inequality $(a+b+c+d)^2 \le 4(a^2+b^2+c^2+d^2)$ gives

$$\begin{aligned}\|\tilde{\psi}(xy)-\tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}^2 \le 4\Big(&\|\tilde{\psi}(xy)-\psi(xy)\|_{\mathrm{F}}^2\\ &+\|\psi(xy)-\psi(x)\,\psi(y)\|_{\mathrm{F}}^2\\ &+\left\|\big(\psi(x)-\tilde{\psi}(x)\big)\psi(y)\right\|_{\mathrm{F}}^2\\ &+\|\psi(y)-\tilde{\psi}(y)\|_{\mathrm{F}}^2\Big)\,.\end{aligned}$$

When we take the expectation over y, the third term simplifies since $\psi(y)$ is unitary in expectation:

$$\begin{aligned}\mathop{\mathbb{E}}_{y}\left\|\big(\psi(x)-\tilde{\psi}(x)\big)\psi(y)\right\|_{\mathrm{F}}^2 &= \operatorname{tr}\left[\big(\psi(x)-\tilde{\psi}(x)\big)^\dagger\left(\mathop{\mathbb{E}}_{y}\psi(y)^\dagger\psi(y)\right)\big(\psi(x)-\tilde{\psi}(x)\big)\right]\\ &= \operatorname{tr}\left[\big(\psi(x)-\tilde{\psi}(x)\big)^\dagger\big(\psi(x)-\tilde{\psi}(x)\big)\right]\\ &= \|\psi(x)-\tilde{\psi}(x)\|_{\mathrm{F}}^2\,.\end{aligned}$$

Putting this together, taking expectations over Π, x, and y, using the fact that xy is uniformly random, and applying (30) and Theorem 4 gives

$$\begin{aligned}\mathop{\mathbb{E}}_{\Pi,x,y}\|\tilde{\psi}(xy)-\tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}^2 &\le 4\left(\mathop{\mathbb{E}}_{x,y}\|\psi(xy)-\psi(x)\,\psi(y)\|_{\mathrm{F}}^2+3\mathop{\mathbb{E}}_{\Pi,x}\|\psi(x)-\tilde{\psi}(x)\|_{\mathrm{F}}^2\right)\\ &\le 2d_\psi\left(4\left(1-\sqrt{\frac{d_\psi}{d_\rho}}\right)+6\left(1-\frac{d_\psi}{d_\rho}\right)\right)\,,\end{aligned}$$

completing the proof. □

There are a number of ways one might improve Theorem 5. The bound (25), and therefore (26), is off by a factor of $(\lambda+1)^2 \approx 4$ when λ is close to 1, i.e., when $\psi(x)$ is close to unitary. Using the triangle inequality is also rather crude. With more thought one should be able to bound $(1/d_\psi)\,\mathbb{E}_{\Pi,x,y}\,\|\tilde{\psi}(xy)-\tilde{\psi}(x)\,\tilde{\psi}(y)\|_{\mathrm{F}}^2$ with a smaller function of the ratio d_ψ/d_ρ, and thus achieve good approximate representations in lower dimensions.

Acknowledgments We benefited from the lectures of Avi Wigderson and Ben Green at the Bellairs Research Institute of McGill University, and from conversations with Jon Yard about twirled operators. This work was supported by the NSF under grants CCF-0829931, 0835735, and 0829917, and by the NSA under contract W911NF-04-R-0009.

References

[1] W. T. Gowers. Quasirandom groups. Combinatorics, Probability and Computing 17(3):363–387, 2008.

[2] László Babai and Katalin Friedl. Approximate Representation Theory of Finite Groups. *Proc. 32nd Symp. Foundations of Computer Science*, 733–742, 1991.

[3] László Babai, Katalin Friedl and András Lukács. Near representations of finite groups, Manuscript, 2003.

[4] Jean-Pierre Serre. *Linear Representations of Finite Groups*. Number 42 in Graduate Texts in Mathematics. Springer-Verlag, 1977.

[5] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.

[6] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos A. Kiwi, and Madhu Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.

[7] K. Fan and A. J. Hoffman, Some metric inequalities in the space of matrices. *Proc. Amer. Math. Soc.* 6 (1955) 111–116.