Introduction to and applications of harmonic analysis and representation theory of finite groups[1]

A course instructed by Amir Yehudayoff, Department of Mathematics, Technion-IIT

[1] An apology: this text may contain errors.

Contents

Chapter 1

Harmonic analysis and linear algebra

We start with an abstract study of abelian groups. All groups we consider will be finite. Let G be an abelian group, that is, $gg' = g'g$ for all g, g' in G. There are many examples of such groups in the theory of computing, but perhaps the most notable one is the discrete cube

$$G = \{0,1\}^n$$

with the group operation being coordinate-wise addition modulo two. It will be a good example to keep in mind.

1.1 Diagonalizing the space of functions on G

We wish to understand properties of G. The main tool we shall use will be linear algebra. To do so, we need to find a useful vector space. As is typically done, we consider

$$\mathbb{C}[G] = \{f : G \to \mathbb{C}\}.$$

This vector space has a natural inner product defined over it

$$\langle f, f' \rangle = \frac{1}{|G|} \sum_{g \in G} f(g)\overline{f'(g)} := \mathbb{E} f \overline{f'},$$

where $\overline{c}$ is the conjugate of $c \in \mathbb{C}$.

So we have a vector space to play with. We shall try to study G by its action on $\mathbb{C}[G]$. A group G acts on a set X if every $g \in G$ defines a map[1] $g(x)$ from X to itself

[1] We slightly abuse notation.

that is a group homeomorphism, that is,

$$(gg')(x) = g(g'(x)),$$

for all $g, g' \in G$ and $x \in X$. Every map $g : X \to X$ is one-to-one and onto since it has an inverse g^{-1}. Here are some well-known examples of group actions:

- A group G can act on itself $X = G$ by left multiplication $g(h) = gh$.
- The two-dimensional affine group G over a finite field $X = \mathbb{F}$ acts on the line $\mathbb{F}$: Every element of G is of form $g = (a, b)$ where $a, b \in \mathbb{F}$ and $a \neq 0$, and

  $$g(x) = ax + b,$$

  for every $x \in X$.

When G acts on X, it also acts on $\mathbb{C}[X] = \{f : X \to \mathbb{C}\}$ by [2]

$$(g(f))(x) = f(g^{-1}(x)).$$

So the group G also acts on $\mathbb{C}[G]$. The map $f \mapsto g(f)$ is clearly a linear map on $\mathbb{C}[G]$ and we may try to study the structure of these linear maps and from them learn something about G. Choosing our basis to be the standard basis, we may represent this map by a permutation matrix $M_g \in \mathbb{C}^{G \times G}$ defined by

$$M_g(x, y) = \mathbf{1}_{gx=y} = \mathbf{1}_{g=x^{-1}y}.$$

We thus defined a map

$$g \mapsto M_g.$$

This map is called the regular (matrix) representation of G and we shall discuss it in more detail later on.

It is straightforward to observe that every two such matrices commute

$$M_g M_h = M_h M_g,$$

because G is abelian and so $(gh)(f) = (hg)(f)$. We can now start using known properties from linear algebra.

Lemma 1. *If two diagonalisable matrices A, B commute then they can be diagonalised together.*

[2] The inverse of g is there to guarantee that it is an action since $(gh)^{-1} = h^{-1}g^{-1}$.

Proof when all eigenvalues are distinct. Assume A, B are n by n. Assume that there is a list $v_1, \dots, v_n$ of eigenvalues of A so that $Av_i = \lambda_i v_i$ for eigenvalues λ_i, and that $\lambda_i \neq \lambda_j$ for all $i \neq j$. Since A, B commute, for all i,

$$ABv_i = BAv_i = \lambda_i Bv_i,$$

so Bv_i is also an eigenvector of A with eigenvalue λ_i which means that Bv_i is in the span of v_i. In other words, v_i is also an eigenvector of B.

We leave the proof in the general case as an exercise. □

1.2 Eigenvectors are characters

With the lemma in mind, we can deduce that if $|G| = n$ then there are $v_1, \dots, v_n$ in $\mathbb{C}[G]$ so that for all $g \in G$, we have

$$M_g v_i = \lambda_{i,g} v_i.$$

Fix $v = v_i$ and $\lambda_{i,g} = \lambda_g$. By definition, for all $g \in G$,

$$v(g) = (g(v))(1) = (\lambda_g v)(1) = \lambda_g v(1),$$

so all entries of v are determined by $v(1)$ which we may take to be $v(1) = 1$ and then $\lambda_g = v(g)$. In this case,

$$v(gh) = (gh)(v) = g(h(v)) = g(v(h) \cdot v) = v(h) \cdot v(g).$$

Since M_g is a permutation matrix, the eigenvalues $v(g)$ of M_g are complex roots of unity. Denote $D = \{c \in \mathbb{C} : |c| = 1\}$, the complex unit circle. We see that every eigenvector v defines a homeomorphism from G to D. These homeomorphisms are called characters.

Definition 2. *A map $\chi : G \to D$ is a character of G if*

$$\chi(gh) = \chi(g)\chi(h)$$

for all $g, h \in G$.

Every eigenvalues thus yields a character. The implication in the other directions holds as well: For every $g, h \in G$ and character χ of G,

$$(M_g \chi)(h) = \chi(gh) = \chi(g)\chi(h).$$

The character χ is an eigenvector of eigenvalue $\chi(g)$ of M_g.

There are therefore $|G|$ different characters of G. The norm of each character is clearly one:

$$\|\chi\|_2^2 = \langle \chi, \chi \rangle = \frac{1}{|G|} \sum_{g \in G} \chi(g)\overline{\chi(g)} = 1.$$

They are also orthogonal: A useful observation is that

$$\overline{\chi(g)} = \chi(g^{-1}).$$

If $\chi \neq \psi$ are two characters of G then

$$\begin{aligned} |\langle \chi, \psi \rangle|^2 &= \frac{1}{|G|^2} \sum_{g,g' \in G} \chi(g)\psi(g^{-1})\chi(g'^{-1})\psi(g') \\ &= \frac{1}{|G|^2} \sum_{g,g' \in G} \chi(gg'^{-1})\psi(g^{-1}g') \\ &= \frac{1}{|G|^2} \sum_{h \in G} \sum_{g,g' : gg'^{-1} = h} \chi(h)\psi(h^{-1}) \\ &= \frac{1}{|G|} \sum_{h \in G} \chi(h)\psi(h^{-1}) = \langle \chi, \psi \rangle. \end{aligned}$$

Thus, $\langle \chi, \psi \rangle \in \mathbb{R}$ and moreover

$$\langle \chi, \psi \rangle (1 - \langle \chi, \psi \rangle) = 0.$$

Since $\chi(h) \neq \psi(h)$ for some $h \in G$, we know that $\chi(h)\psi(h^{-1}) \neq 1$ and so

$$|\langle \chi, \psi \rangle| \leq 1 - \frac{1}{|G|} + \frac{|\chi(h)\psi(h^{-1})|}{|G|} < 1,$$

which means

$$\langle \chi, \psi \rangle = 0.$$

This concludes the first part of the discussion. When studying an abelian group, we looked on the vectors space $\mathbb{C}[G]$, and found the homeomorphisms from G to $D \subset \mathbb{C}$ form a different orthonormal basis to $\mathbb{C}[G]$. This basis will be sometimes much easier to work with.

1.3 Examples and basic properties

Here are four examples:

- The group $\mathbb{Z}_2$ has two characters χ_0, χ_1 defined by

$$\begin{array}{ccc} & 0 & 1 \\ \chi_0 & 1 & 1 \\ \chi_1 & 1 & -1 \end{array}$$

  The all one vector χ_0 is called the trivial character.

- The cyclic group $\mathbb{Z}_n$ has n characters $\chi_0, \ldots, \chi_{n-1}$: If we denote by $\omega = e^{2\pi i/n}$ the n'th root of unity, then
$$\chi_i(j) = \omega^{ij}.$$

- If a group G has characters $\chi_0, \ldots, \chi_n$ and G' has characters $\chi'_0, \ldots, \chi'_{n'}$ then the group $G \times G'$ has characters
$$\chi_{(i,i')}(g, g') = \chi_i(g)\chi_{i'}(g').$$

- The discrete cube $\mathbb{Z}_2^n$ has 2^n characters indexed by $y \in \{0,1\}^n$:
$$\chi_y(x) = (-1)^{\sum_{i \in [n]} x_i y_i}.$$

  We shall sometimes identify y with the set $\{i \in [n] : y_i = 1\}$.

The second and third bullets above describe how to build the characters of all finite abelian groups, due to the known structure of such groups as a product of cyclic groups.

The set of characters $\hat{G}$ of G forms a group by itself: The group operation is defined by
$$(\chi\psi)(g) = \chi(g)\psi(g).$$
It can be verified that G and $\hat{G}$ are isomorphic as groups.

Given a function $f \in \mathbb{C}[G]$, we may therefore write it uniquely as
$$f = \sum_{\chi \in \hat{G}} \hat{f}(\chi)\chi,$$
where
$$\hat{f}(\chi) = \langle f, \chi \rangle \in \mathbb{C}.$$
We can therefore think of $\hat{f}$ as a function from $\hat{G}$ to $\mathbb{C}$. This function is called to Fourier or Walsh-Fourier transform of f. Sometimes there will be a clear identification between G and $\hat{G}$ and then we shall think of $\hat{f}$ as a function on G. For example, for a prime cyclic group $G = \mathbb{Z}_p$, there is a natural identification between $g \in G$ and the character $\chi_g(x) = \omega_g^x$ where $\omega_g = e^{2\pi i g/p}$.

Since $\hat{G}$ is a set of orthonormal vectors, it preserves inner products

$$\langle f, h\rangle = \sum_{\chi,\psi\in\hat{G}} \hat{f}(\chi)\overline{\hat{h}(\psi)}\langle\chi,\psi\rangle = \sum_{\chi\in\hat{G}} \hat{f}(\chi)\overline{\hat{h}(\chi)}.$$

Parseval's identity follows

$$\|f\|_2^2 = \frac{1}{|G|}\sum_{g\in G}|f(g)|^2 = \sum_{\chi\in\hat{G}}|\hat{f}(\chi)|^2.$$

Chapter 2

Applications

2.1 Pseudorandom sets

As a first applications, we consider pseudorandom sets. The basic notions we wish to define and study is the amount of randomness in of subsets of G or more generally functions on it.

2.1.1 Measuring amount of randomness

Let p be a probability distribution on G. How do we measure the amount of randomness in p? The "most random" distribution on G is the uniform distribution u on it. The amount of randomness in p can be therefore measured by its distance from u. There are several metrics we can use:

The first we mention is called statistical distance or L_1 distance. The statistical distance (a.k.a. total variation distance) between two distributions p, q on G is defined as

$$\text{stat-dist}(p, q) = \max\{|p(S) - q(S)| : S \subset G\}.$$

It measures the difference between the p-measure and the q-measure of every event in the space. It is equivalent to L_1 distance

$$\|p - q\|_1 = \sum_{g \in G} |p(g) - q(g)|,$$

that is,

$$\|p - q\|_1 = 2 \cdot \text{stat-dist}(p, q)$$

(an exercise: draw histograms).

The second metric is the L_2 metric, which is most natural in the context of harmonic analysis:

$$\|p - q\|_2^2 = \langle p - q, p - q \rangle.$$

Observe that by Cauchy-Schwartz

$$\|p - q\|_1 = \sum_g |p(g) - q(g)| \le |G| \cdot \|p(g) - q(g)\|_2.$$

Specifically, when measuring the distance from u, using the linearity of the Fourier transform and Parseval,

$$\|p - u\|_2^2 = \sum_\chi |\hat{p}(\chi) - \hat{u}(\chi)|^2.$$

What is the Fourier transform of u? Denote by χ_0 the trivial character of G. Thus,

$$\hat{u}(\chi) = \frac{1}{|G|}\langle \chi, \chi_0 \rangle = \frac{1}{|G|}\mathbf{1}_{\chi=\chi_0}.$$

Observe also that

$$\hat{p}(\chi_0) = \frac{1}{|G|}.$$

Therefore,

$$\|p - u\|_2^2 = \sum_{\chi \neq \chi_0} |\hat{p}(\chi)|^2.$$

The L_2 distance of p from u is thus neatly expressible in the language of harmonic analysis.

As a comment we mention another useful way to measure "distance" between distribution. The *divergence* between p, q is defined as

$$D(p||q) = \sum_g p(g) \log_2 \frac{p(g)}{q(g)}.$$

It is not a metric but it satisfies similar properties. When measuring $D(p||u)$ we get (up to translation and a minus sign) the entropy of p, which is a well known measure of the amount of randomness in p.

2.1.2 Small biased sets

From the discussion above, we see that one way to measure the amount of randomness in p has a natural interpretation in term of harmonic analysis ($\|p - u\|_2$). This is one of the motivations for the following definition.

Definition 3. *A distribution p on a finite abelian group G is called ϵ-biased if for every nontrivial character $\chi \in \hat{G}$,*

$$|\widehat{p}(\chi)| \le \frac{\epsilon}{|G|}.$$

Every distribution is 1-biased, using the triangle inequality. The uniform distribution u is 0-biased, and this is the only 0-biased set. An ϵ-biased set S is close to the uniform distribution in the sense that

$$\sum_{g \in G} (p(g) - u(g))^2 \le \epsilon.$$

Motivation. Motivation for studying ϵ-biased sets over say $G = \mathbb{Z}_2^n$ comes from pseudorandomness. The general aim of this area is constructing objects that are not truly random but are close enough (in a useful way) to being random.

Fooling statistical test. A useful notion in the theory of computing is a fooling set against a given family of statistical tests. A distribution p on G is said to ϵ-fool the set $F \subset \mathbb{C}[G]$ if for every $f \in F$,

$$|\mathbb{E}_{g \sim p} f(g) - \mathbb{E}_{g \sim u} f(g)| \le \epsilon,$$

where $g \sim p$ means an element drawn from p_T. When $F = \{\mathbf{1}_S : S \subseteq G\}$, for example, p ϵ-fools F iff p is ϵ-close to u in statistical distance. When $F = \{f_R : R \subseteq [n]\}$ is the set of linear functions ($f_R(x) = \sum_{i \in R} x_i$ from $G = \mathbb{Z}_2^n$ to $\mathbb{Z}_2$), ϵ-fooling F means exactly ϵ-biased. So, ϵ-biased is a weaker notion than being close to uniform in statistical distance. It is nevertheless sufficient in many applications.

Codes. Another motivation to study ϵ-biased sets over $G = \mathbb{Z}_2^n$ is the connection to coding theory [Azar, Motwani, and Naor]. Let $S \subset \mathbb{Z}_2^n$ be an ϵ-biased set. For every $R \subseteq [n]$, define $v_R \in \mathbb{Z}_2^S$ by $v_R(s) = f_R(s)$. Two key observations are

- The fraction of ones in v_R is between $1/2 - \epsilon$ and $1/2 + \epsilon$.
- The set $V = \{v_R\}$ is closed under addition in $\mathbb{Z}_2^S$.

This means that V is a linear subspace of $\mathbb{Z}_2^S$, of dimension n so that the hamming distance (i.e. number of entries of difference) between every $v \neq v'$ in V is at least $|S|(1/2 - \epsilon)$. An ϵ-biased set thus corresponds to a linear error correcting code. We shall discuss codes in more detail later on.

Background. We see that finding as small as possible ϵ-biased sets is of interest in coding theory as well as pseudorandomness. It is known [McEliece, Rodemich, Rumsey and Welch] that there are limitation to size of such sets. For $G = \mathbb{Z}_2^n$, their size must be at least order $\frac{n}{\epsilon^2 \log(1/\epsilon)}$, when ϵ is not too small. It can also be shown that a random subset of $G = \mathbb{Z}_2^n$ of size order n/ϵ^2 is w.h.p. ϵ-biased.

Constructions. Explicit constructions of such small ϵ-biased sets are not known, but constructions that are almost as good are known. We shall now explain one simple construction [Alon-Mansour, Alon-Goldreich-Hastad-Peralta] that gives an ϵ-biased distribution with support of size order $(n/\epsilon)^2$. The analysis uses harmonic analysis.

For simplicity assume that n/ϵ is a power of two, so log base two of it is k. Consider the field $\mathbb{F} = \mathbb{F}_{2^k}$ of size n/ϵ. For every $x, z \in \mathbb{F}$, define $s = s(x, z)$ in G by

$$s_i = (xz^i)_1,$$

for every $i \in [n]$, i.e., the first bit in the field element xz^i (out of k possible bits). Define a distribution p on G as the distribution on $s = s(x, z)$ when x, z are chosen independently at random from $\mathbb{F}$. The support of p is indeed of size at most $(n/\epsilon)^2$. Bound $|G| \cdot |\hat{p}(\chi_y)|$ for $y \neq \emptyset$ as follows:

$$\begin{aligned} |G| \cdot |\hat{p}(\chi_y)| &= \sum_g p(g)\chi_y(g) \\ &= \sum_g \sum_{x,z:s(x,z)=g} \frac{1}{|\mathbb{F}|^2} \chi_y(g) \\ &= \frac{1}{|\mathbb{F}|^2} \sum_{x,z} (-1)^{\sum_{i\in y}(xz^i)_1} \\ &= \frac{1}{|\mathbb{F}|^2} \sum_{x,z} \prod_{i \in y} (-1)^{(xz^i)_1}. \end{aligned}$$

The map $t \mapsto (-1)^{t_1}$ is a character of the field $\mathbb{F}$. Denote it by ψ. Therefore,

$$\begin{aligned} |G| \cdot |\hat{p}(\chi_y)| &= \frac{1}{|\mathbb{F}|^2} \sum_{x,z} \prod_{i \in y} \psi(xz^i) \\ &= \frac{1}{|\mathbb{F}|^2} \sum_{x,z} \psi(x \sum_{i \in y} z^i). \end{aligned}$$

Since $y \neq \emptyset$, the polynomial $h(\xi) = \sum_{i \in y} \xi^i$ is nonzero of degree at most n, which means that

$$\Pr_z[h(z) = 0] \leq \frac{n}{|\mathbb{F}|} \leq \epsilon.$$

On the other hand, for all z so that $g(z) \neq 0$, we get

$$\frac{1}{|\mathbb{F}|} \sum_x \psi(xh(z)) = 0,$$

since it is an inner product of a nontrivial character of $\mathbb{F}$ with the trivial one. Overall,

$$|G| \cdot |\hat{p}(\chi_y)| \leq \epsilon \frac{1}{|\mathbb{F}|} \sum_x \psi(x \cdot 0) = \epsilon.$$

One may observe that in the above argument, we did not really require x to be every element of $\mathbb{F}$, but we could have chosen it from an ϵ-biased set in $\mathbb{F}$. The bound obtained would be 2ϵ instead of ϵ. This gives a construction [Alon et al.] with S of size at most

$$|\mathbb{F}| \cdot \frac{\log^2(|\mathbb{F}|)}{\epsilon^2} \approx \frac{n}{\epsilon^3} \log^2(n/\epsilon),$$

which is smaller than $(n/\epsilon)^2$ for some choice of parameters.

2.2 Basic number theory

We now discuss Gauss sums which are a basic object and tool in number theory. Let $\mathbb{F}$ be a finite prime field of size $p > 2$. The field consists of two groups, and additive one $(\mathbb{F}, +)$ and a multiplicative one $(\mathbb{F}^*, \cdot)$. Each of these groups has its own characters.

The additive group is of size p is cyclic and we know its characters. We shall use χ to denote additive characters. The multiplicative group is cyclic as well, but not of prime order so we do not have nice formulas for it characters. We shall use ψ to denote multiplicative characters.

An example to a multiplicative character ψ is quadratic residuosity: $\psi(g) = 1$ if $g = x^2$ for some $x \in \mathbb{F}^*$ and $\psi(g) = -1$ otherwise (if extended to $\mathbb{F}$ by $\psi(0) = 0$, it is called the Legendre symbol we consider below). It is a multiplicative character for the following reason. Observe that the map $A(g) = g^{(p-1)/2}$ maps every x to $\{1, -1\}$ (the only elements whose squares are one). The claim is that A and ψ agree (although take values in different fields). The map $B(g) = g^2$ is onto the quadratic residues and is two-to-one ($B(g) = B(-g)$ and $g \neq -g$ since $p > 2$). So there are exactly $(p-1)/2$ quadratic residues. If $\psi(g) = 1$ then clearly $A(g) = 1$. On the other hand, $A(g) = 1$ is a polynomial equation of degree $(p-1)/2$. So, $|A^{-1}(1)| = (p-1)/2$ and so if $\psi(g) = -1$ then $A(g) \neq 1$ which means $A(g) = -1$.

This is already non-trivial as it tells us that half of $\mathbb{F}^*$ are quadratic residues and that if e.g. g is a quadratic residue and g' is not, then gg' is not as well.

A Gauss sum is a sum of the form

$$\sum_{x \in \mathbb{F}^*} \psi(x)\chi(x).$$

It measures the correlation between additive and multiplicative characters. In other words, it is an additive Fourier coefficient of the set of the quadratic residues. Trivially, the maximum value it may attain is n. A general formula for Gauss sums is not known, but its modulus is known.

$$\begin{aligned}
\left|\sum_{x\in\mathbb{F}^*}\psi(x)\chi(x)\right|^2 &= \sum_{x,y\in\mathbb{F}^*}\psi(xy^{-1})\chi(x-y)\\
&= \sum_{z,y\in\mathbb{F}^*}\psi(z)\chi(y(z-1))\\
&= \sum_{z\in\mathbb{F}^*}\psi(z)\sum_{y\in\mathbb{F}^*}\chi(y(z-1))\\
&= \sum_{z\in\mathbb{F}^*}\psi(z)(\mathbf{1}_{z=1}(p-1)+\mathbf{1}_{z\neq1}(-1))\\
&= p\psi(1)-\sum_{z\in\mathbb{F}^*}\psi(z)=p.
\end{aligned}$$

This is a sum of $(p-1)$ root of unity which always has the same modulus $\sqrt{p}$. The proof is quite simple but shows the strength of rearranging sums, which is a basic and very useful idea.

One application of Gauss sums is the following theorem due to Gauss. The Legendre symbol of $x\in\mathbb{F}=\mathbb{F}_p$ modulo p is the map

$$\left(\frac{x}{p}\right)=\begin{cases}1 & \text{if there is } y\in\mathbb{F}^* \text{ so that } x=y^2,\\ -1 & \text{if for all } y\in\mathbb{F}^* \text{ we have } x\neq y^2,\\ 0 & x=0.\end{cases}$$

Theorem 4 (Gauss)**.** *For every two odd primes $p\neq q$,*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

We shall not prove the theorem here, but some proofs of it use the above estimate of gauss sums. The theorem tells us several nontrivial properties of quadratic residues and equations. For example, it shows that if $p=1 \mod 4$ then there is a solution to

$$x^2=p \mod q$$

iff there is a solution to

$$x^2=q \mod p.$$

2.3 Random walks

Random walks model many natural phenomena and have been studied extensively in various settings. We focus on random walks on finite abelian groups, like the discrete cube.

Let G be a finite abelian group, and let $S \subset G$ be a symmetric set (that is, $g \in S$ iff $g^{-1} \in S$) that generates G (that is, every g in G can be expressed as word in S). The *Cayley* graph $Cay(G,S)$ defined by S on G has vertex set G and its edges are of the form $\{g, sg\}$ for $s \in S$. It is an undirected graphs since S is symmetric.

A random walk on $Cay(G,S)$ with initial distribution p is a sequence $X_0, X_1, X_2, \ldots$ of random elements of G so that $X_0 \sim p$ and X_{t+1} is a random neighbour of X_t, chosen independently of previous choices. If X_0 is chosen at random according to a distribution p_0 on G then the distribution p_1 of X_1 can be written as

$$p_1(g) = \sum_{h \sim g} \frac{1}{|S|} \cdot p_0(h),$$

where here $h \sim g$ denotes[1] that $\{h, g\}$ is an edge in $Cay(G,S)$. There is therefore a stochastic (symmetric) matrix

$$M = \sum_{s \in S} \frac{1}{|S|} M_s,$$

where M_s is the permutation matrix s defines that we already saw, so that

$$p_1 = M p_0.$$

In general, for all $t \geq 0$, the distribution of X_t is

$$p_t = M^t p_0.$$

We can use harmonic analysis to understand properties of this random walk. Write

$$p = \sum_{\chi \in \hat{G}} \hat{p}(\chi) \chi.$$

Thus,

$$p_t = M^t \left(\sum_{\chi \in \hat{G}} \hat{p}(\chi) \chi \right) = \sum_{\chi \in \hat{G}} \lambda_\chi^t \hat{p}(\chi) \chi,$$

[1] A slight abuse of notation.

where

$$\lambda_\chi = \sum_{s \in S} \frac{1}{|S|} \chi(s) \in \mathbb{R}$$

is the eigenvalue of M that corresponds to χ.

This implies e.g. that $u = \chi_0/|G|$ is the *stationary* distribution of the Markov chain, i.e.,

$$Mu = u$$

since $\chi_0(s) = 1$ for all $s \in S$.

Convergence. To show that $p_t \to u$, we need to show $|\lambda_\chi| < 1$ for every $\chi \neq \chi_0$. Assuming that $\lambda_\chi = 1$ implies that $\chi(s) = b \in \{1, -1\}$ for all $s \in S$. Since S generates G, it follows that $\chi(g) \in \{1, -1\}$ for all $g \in G$. If $Cay(G, S)$ has a cycle of odd length d then

$$1 = \chi(1) = \chi(s_1 s_2 \cdots s_d) = b^d,$$

which means that $b \neq 1$, a contradiction. Contra-positively, if the graph has a cycle of odd length (i.e. it is not bipartite) then all nontrivial eigenvalues have absolute values smaller than one which means that the coefficient of $\chi \neq \chi_0$ in p_t tends to zero, and so p_t tends to u.

When the graph is on the other hand bipartite, it may be the case that p_t does not converge. Indeed, if p_0 is the uniform distribution on one of the parts of the graph, then p_1 is uniform on the other part, p_2 on the first part and so forth (this is called a periodic chain).

Mixing times. A basic property of a random walk its *mixing time*. There are several variants of it, and we focus on the L_2 version of it:

$$T(\epsilon; p) = \min\{t \geq 0 : \|p_t - u\|_2^2 \leq \epsilon/|G|\}$$

and

$$T(\epsilon) = \max\{T(\epsilon; p) : p\},$$

where the norm is according to the normalised inner product on G. Sometimes, we may know $\chi(s)$ for $s \in S$. In such cases, we may obtain bounds on the nontrivial eigenvectors and on the mixing time.

The discrete cube. Take for example the discrete cube. That is, the Cayley graph of $G = \mathbb{Z}_2^n$ with $S = \{e_i : e \in [n]\}$ where e_i is the i'th unit vector. We already know the characters of G:

$$\chi_y(g) = (-1)^{\sum_{i \in [n]} y_i g_i}.$$

The eigenvalues of M are therefore

$$\lambda_y = \sum_{i \in [n]} \frac{1}{n}(-1)^{y_i} = |\{i : y_i = 0\}| - |\{i : y_i = 1\}| = n - 2|y|,$$

where $|y|$ is the hamming weight of y (the number of ones in it). The largest nontrivial eigenvalue of M is thus $1 - 2/n$, which implies that for every p,

$$\begin{aligned} \|p_t - u\|_2^2 &= \sum_{\chi \neq \chi_0} |\widehat{p_t}(\chi)|^2 \\ &= \sum_{\chi \neq \chi_0} |\lambda_\chi^t \hat{p}(\chi)|^2 \\ &\leq (1 - 2/n)^t \|p - u\|_2^2 \leq (1 - 2/n)^t. \end{aligned}$$

So, $T(\epsilon)$ is at most order $n(n + \log(1/\epsilon))$ which is poly-logarithmic in the size of the graph.

As we have seen before if $|\lambda_\chi|/|S| < 1$ for all nontrivial χ then the random walk on H converges to a uniform element of G. If $|\lambda_\chi|/|S| < 0.9$ for $\chi \neq \chi_0$ then this convergence is exponentially fast, and such constant-degree regular graphs are called expanders (every set has a large boundary in such regular graphs). An exercise is to show that e.g. $Cay(G, S)$ for $|S| \leq 6$ and $|G|$ large is not an expander (G is abelian).

2.4 Error correcting codes

Error correction codes are very useful objects that allow communication in noisy channels and are also related to many other questions concerning groups, geometry and more. A different point of view of the following discussion is as a spectral approach to geometry (an idea that may deserve its own part in the course).

In this part, we study limitations of such codes, following [McEliece-Rodemich-Rumsey-Welch, Navon-Samorodnitzky]. The approach uses harmonic analysis, and we take a more general points of view (instead of $\mathbb{Z}_2^n$ we look at general groups).

Let G be a finite group, let $S \subset G$ be a generating and symmetric, and let $H = Cay(G, S)$. We are interested in the geometry of H.

Definition 5. *A set $C \subset G$ is a* code *with distance d if for every $c \neq c'$ in C,*

$$dist(c, c') \geq d,$$

where dist is graph distance in H.

The definition of a code is clear: When we are given a point $g \in G$ that we know is a "noisy" version of some $c \in C$, if the distance is large then we can recover the correct c.

The goal is to find codes C that are as large as possible with as large as possible distance. The meaning of large is typically measured by its *asymptotic rate*

$$\rho(C) = \frac{\log_2 |C|}{\log_2 |G|}.$$

Roughly, it measures the amount of information C contains, compared to the length of communication needed. (elements of G can be described using order $\log_2 |G|$ bits). There is obviously a tradeoff here. The following is called the sphere-packing bound: If we denote

$$B = B_{d/2} = \{g \in G : dist(g, 1) < d/2\},$$

the H-ball of radius $(d-1)/2$ around $1 \in G$, then

$$|C||B| \leq |G|,$$

since $cB \cap c'B = \emptyset$ for all $c \neq c'$ in C. The size of $B_{d/2}$ of course increases as d increases, so as d increases, the size of a code must decrease.

If the group G is the discrete cube, then there is a natural inner product defined on G. For general groups G, we use the dual group (for the discrete cube there is a standard identification between G and $\hat{G}$).

Definition 6. *The* dual code[2] *of a code $C \subset G$ is $C^\perp \subset \hat{G}$ defined as the set of characters χ so that $\chi(c) = 1$ for all $c \in C$. The* dual distance *of C with respect to $S^\perp \subset \hat{G}$ is the distance of $C^\perp$ in $Cay(\hat{G}, S^\perp)$.*

Consider $G = \mathbb{Z}_2^n$ for example. It is natural to look for linear (i.e. subgroup) codes C of large distance. The dual code $C^\perp$ is in this case the dual subspace of C as well. A reasonable choice in this case is $S^\perp = S = \{e_i\}$, the standard basis. Then, the dual distance is the distance of $C^\perp$ as a code in the discrete cube itself.

Duality and subgroups. When C is a subgroup (for the discrete cube, a linear code),

$$|\widehat{\mathbf{1}_C}(\chi)|^2 = \frac{1}{|G|^2} \sum_{c,c' \in C} \chi(c)\chi(c'^{-1}) = \frac{|C|}{|G|^2} \sum_{c \in C} \chi(c) = \frac{|C|}{|G|^2} \widehat{\mathbf{1}_C}(\chi),$$

which implies that

$$\widehat{\mathbf{1}_C}(\chi) = \frac{|C|}{|G|^2} \mathbf{1}_{C^\perp}.$$

Spectral aspects of sets. Denote by M the adjacency matrix of the Cayley graph H. We know that $\hat{G}$ are the eigenvectors of M. We in fact have a formula for the eigenvalues:

$$\lambda_\chi = \sum_{s \in S} \chi(s).$$

It will be useful to associate a spectral value to a set $B \subset G$:

$$\lambda_B = \max \left\{ |\langle Mf, f \rangle| \ : \ f : B \to \mathbb{C}, \|f\|_2 = 1 \right\}.$$

A different way to view λ_B is as follows. Considering the subgraph H_B of H induced on the elements of B, and let M_B be its adjacency matrix. Thus, λ_B is the maximal eigenvector of H_B, with corresponding eigenvector f_B. The Perron-Frobenius theorem tells us that $\lambda_B \geq 0$ and that the entries of f_B are nonnegative.

It will also be useful to measure the spectral gap of the dual of a set with respect to the Cayley graph. For $C \subset G$, denote

$$\lambda^\perp(C) = \max\{|\lambda_\chi| : \chi \in C^\perp, \chi \neq \chi_0\}.$$

The main theorem we shall discuss is:

[2] Not to be confused with $\hat{C}$, the Fourier transform of the characteristic function of C.

Theorem 7. *Let $C \subset G$ be a subgroup with dual spectral value $\lambda^{\perp}(C)$. Let $B \subset C$ be a set of spectral value λ_B. Then,*

$$|B \cdot C| \geq |G| \frac{\lambda_B - \lambda^{\perp}(C)}{|S|}.$$

In words, if B has a large spectral value and C has a small dual spectral value then the cosets of C defined by B contain most of G. In some sense, this shows that B and C are highly non-similar. Before proving the theorem, we give an example and preliminary results in harmonic analysis.

Example: the cube. We give an example from $G = \mathbb{Z}_2^n$ over the discrete cube. Let C be a linear code in G of distance d. This means that all points in C are far away from each other. So if B is a hamming ball of small radius then BC will be much larger than C. The theorem tells use for which radius BC is almost all of G.

Lemma 8. *Let $B = B_r$ be a hamming ball of radius r in the discrete cube. Then, $\lambda_B \geq 2\sqrt{r(n-r)} - o(n)$.*

The proof shall be a guided exercise. □

The last ingredient is a bound on the dual spectral value of $C^{\perp}$, when C has distance d. Since the dual of $C^{\perp}$ is C, the Fourier spectrum of $\mathbf{1}_{C^{\perp}}$ is supported on C. The eigenvalue of χ_y is $n - 2|y|$. Thus,

$$\lambda^{\perp}(C^{\perp}) \leq n - 2d.$$

We deduce that for $r = n/2 - \sqrt{d(n-d)} + o(n)$, denoting $d = \delta n$ for fixed small $\delta > 0$,

$$\begin{aligned} n|C^{\perp}B| &\geq 2^n(n\sqrt{(1 - 2\sqrt{\delta(1-\delta)} + o(1))(1 + 2\sqrt{\delta(1-\delta)} - o(1))} - o(n) - n + 2d) \\ &\geq 2^n(n\sqrt{1 - 4\delta(1-\delta) + o(1)} + o(n) - n + 2d) \\ &\geq 2^n(n(1 - 2\delta(1-\delta) + o(1)) - o(n) - n + 2d), \end{aligned}$$

or

$$|C^{\perp}||B| \geq |C^{\perp}B| \geq 2^n(2\delta - o(1)).$$

Since $|C||C^{\perp}| = 2^n$,

$$|C| \leq \frac{|B|}{2\delta - o(1)}.$$

Standard estimates show

$$|B| \approx 2^{n\mathbb{H}(r/n)},$$

where here $\mathbb{H}$ is the entropy function $\mathbb{H}(\xi) = -\xi\log_2(\xi) + (1-\xi)\log_2(1-\xi)$. The asymptotic rate of C with a given relative distance $\delta > 0$ is therefore at most

$$\rho(C) \leq \mathbb{H}(1/2 - \sqrt{\delta(1-\delta)}).$$

This bound is called the first linear programming bound for linear codes. Compare this to the sphere packing bound

$$\rho(C) \leq 1 - \mathbb{H}(\delta/2).$$

Plotting these function we see that for small δ the sphere packing bound is better than the MRRW bound, but for say $\delta > 0.12$ the MRRW bound becomes better.

Convolution. The space $\mathbb{C}[G]$ is in fact a ring. Multiplication in it is called *convolution* and is defined naturally if we formally think of $f \in \mathbb{C}[G]$ as $f = \sum_{g \in G} f(g)g$. That is,

$$(f * h)(g) = \sum_{g' \in G} f(g')h({g'}^{-1}g).$$

Another useful perspective concerns the matrix representation of the group algebra. Every $f \in \mathbb{C}[G]$ defines a matrix

$$M_f = \sum_{g \in G} f(g) M_g$$

and the map $f \mapsto M_f$ is invertible. In addition,

$$M_f M_h = M_h M_f = M_{f*h}$$

and

$$M_f h = f * h,$$

where we think of $f * h$ as a column vector.

A key property of the Fourier transform is that convolution translates to point-wise multiplication:

$$\begin{aligned}\widehat{(f*h)}(\chi) &= \langle f*h, \chi\rangle \\ &= \mathbb{E}_g \sum_{g'\in G} f(g')h(g'^{-1}g)\chi(g^{-1}g')\chi(g'^{-1}) \\ &= \sum_{g'\in G} f(g')\chi(g'^{-1})\mathbb{E}_g h(g'^{-1}g)\chi(g^{-1}g') \\ &= \sum_{g'\in G} f(g')\chi(g'^{-1})\hat{h}(\chi) \\ &= \hat{f}(\chi)\hat{h}(\chi).\end{aligned}$$

Proof of Theorem 7. By Cauchy-Schwartz, if $F : U \to \mathbb{C}$ then

$$|\hat{F}(\chi_0)|^2 = |\mathbb{E}_g F(g)\mathbf{1}_U(g)|^2 \le \|F\|_2^2 \cdot \frac{|U|}{|G|}.$$

Our goal is proving a lower bound on the size of

$$U = B \cdot C = \{bc : b \in B, c \in C\}.$$

Define $F : U \to \mathbb{C}$ as follows. Let $f = f_B$ be the eigenvector of M_B of eigenvalue $\lambda = \lambda_B$. Define

$$F = \sum_{c\in C} M_c f,$$

which can also be written as

$$F = \mathbf{1}_C * f.$$

The support of F is on element of the form cb for $c \in C$ and $b \in B$ and so indeed $F : U \to \mathbb{C}$.

Since G is abelian, and computation is nonnegative,

$$MF = \sum_{c\in C} M_c M f \ge \lambda \sum_{c\in C} M_c f = \lambda F,$$

where here $\ge$ means entry-wise (since M has nonnegative entries). Thus,

$$\lambda\|F\|_2^2 \le \langle MF, F\rangle = \sum_{\chi} \lambda_\chi |\hat{F}(\chi)|^2 = |S||\hat{F}(\chi_0)|^2 + \sum_{\chi\ne\chi_0} \lambda_\chi |\hat{F}(\chi)|^2.$$

Recall

$$\hat{F}(\chi) = \widehat{\mathbf{1}_C}(\chi) \cdot \hat{f}(\chi).$$

Since C is a subgroup, the support of $\widehat{\mathbf{1}_C}$ is contained in $C^\perp$, so

$$\begin{aligned} |S| \frac{|\hat{F}(\chi_0)|^2}{\|F\|_2^2} &\geq \lambda - \frac{\sum_{\chi \neq \chi_0} \lambda_\chi \cdot |\widehat{\mathbf{1}_C}(\chi)|^2 \cdot |\hat{f}(\chi)|^2}{\|F\|_2^2} \\ &\geq \lambda - \frac{\sum_{\chi \in C^\perp : \chi \neq \chi_0} \lambda_\chi \cdot |\hat{F}(\chi)|^2}{\|F\|_2^2} \\ &\geq \lambda - \lambda^\perp(C). \end{aligned}$$

□

To summarise, we have seen applications of harmonic analysis in (finite) geometries and error correcting code. A key idea was assigning spectral information to sets and using analysis instead of combinatorics.

2.5 Codes and k-wise independence

As mentioned, it is useful to have small support distributions that are pseudorandom. An example is k-wise independent distributions, which are useful e.g. in derandomization of algorithms. Here we restrict our attention to $G = \mathbb{Z}_2^n$.

Definition 9. *A k-wise independent distribution p on $\mathbb{Z}_2^n$ is so that for every $S \subset [n]$, the marginal of p on coordinates in S is uniform.*

The uniform distribution on G is n-wise independent. The uniform distribution on

$$\chi^{-1}_{(1,1,1,\ldots,1)}(0)$$

is $(n-1)$-wise independent. There are interesting constructions of such distributions and it is also known that they can not have too small support.

An observation is that this notion can be formalised in analytic terms as well, which gives a correspondence between linear codes and k-wise independent distributions supported on linear spaces.

First, every linear k-wise independent distribution yields a linear code of distance greater than k. If p is supported on a subgroup C of $\mathbb{Z}_2^n$ and is k-wise independent then for every y so that $0 < |y| \leq k$,

$$\hat{p}(\chi_y) = \hat{u}(\chi_y) = 0.$$

That is, the dual $C^\perp$ of C, which is also a subgroup, is actually a code of distance at least k.

Second, if $C^\perp$ is a linear code of distance d, then $C = C^{\perp^\perp}$ is a subgroup so that for all $0 < |y| < d$,

$$\widehat{\mathbf{1}_C}(\chi_y) = 0.$$

Recalling the u is the only distribution that is 0-biased, we conclude that the uniform distribution on C is $(d-1)$-wise independent.

This correspondence can also be proved without using hormonic analysis, but harmonic analysis is useful if we replace sets with functions, and perfect k-wise independence to close to being k-wise independent.

2.6 Additive combinatorics

We now explain the main ideas in the proof of Roth's theorem. Roth's theorem states that a subset A of $[n]$ of constant density contains 3-term arithmetic progression, that is, if $|A| \geq \delta n$ for fixed $\delta > 0$ and n large then there are $a, b, c \in A$ so that $a + c = 2b$. This theorem can be proved using harmonic analysis. Gower's found a proof of the much more general Szemeredi's theorem using harmonic analysis (the case of k-term progressions). Gower's proof is too complicated for us to go over here.

A key idea in proving Roth's theorem is *energy / density increment.* The idea is to use harmonic analysis to distinguish between two cases.

1. The set A is pseudorandom (similarly to the notion of ϵ-biased sets), and then A behaves like a random set and specifically it has 3-term progressions.

2. The set A is not pseudorandom. Then it has a high Fourier coefficient, which means that we can go to a smaller universe (the kernel of the character) and slightly increase the relative size of A.

The point is that step 2 can not happen too many times, which means that step 1 will eventually occur.

Before moving to prove the theorem, recall that in the exercise you have seen that there is a subset of $[n]$ of size $n^{1-o(1)}$ that contains no 3-term progressions. This construction is due to Behrend, and it is related to several other interesting constructions. The first is of a graph with many edges and relatively few triangles. The second is related to the number-on-forehead model in communication complexity and to cylinder intersections (cylints).

How to construct a graph with many edges are few triangles? It is a 3-partite graph with parts V_1, V_2, V_3 each of size $2n$ define by a Behrend subset X of $[n]$ of size $n^{1-o(1)}$. Edges between V_1, V_2 are of form $\{v_1, v_1 + x\}$ for $x \in X$, between V_2, V_3 of the form $\{v_2, v_2 + x\}$ for $x \in X$, and between V_1, V_3 of the form $\{v_1, v_1 + 2x\}$ for $x \in X$. The point is that v_1, v_2, v_3 is a triangle iff $v_2 = v_1 + x$, $v_3 = v_2 + x'$ and $v_3 = v_1 + 2x''$ which means

$$x + x' = 2x''.$$

This implies that $x = x' = x''$. So every edges supports exactly one triangle, but still the number of edges is $|V|^{2-o(1)}$.

The number-on-forehead model in communication complexity models a certain kind of communication. There are three players. Each player i has an input X_i written on her forehead, so she can not see X_i but can see the other two inputs. The parties wish to compute $f(X_1, X_2, X_3)$. To do so they communicate. We are interested in understand the most efficient way to achieve this goal. This is related to many other topics, like circuit and proof complexity. Every protocol corresponds to a disjoint union of cylints, which we soon define.

Let $F : [n]^3 \to \mathbb{R}$ be a 3-dimensional tensor. A well-known notion is the tensor rank of F, which is a generalisation of matrix-rank. A tensor of rank one is of the form $F(x,y,z) = F_x(x)F_y(y)F_z(z)$. The tensor rank of F is the minimal r so that $F = \sum_{i \in [r]} F_i$ with F_i of rank one.

A tensor of rank one corresponds to a combinatorial cube. The structure of cube is quite straightforward (still computing the tensor rank is a difficult task (it is NP-hard [Hastad])). For example, if F is a cube of measure $pn^3 = \sum_{x,y,z} F(x,y,z)$ with $F(x,y,z) \in \{0,1\}$ then there is a choice for a plane, say the $x - y$ plane, so that $\sum_z F(x_0, y_0, z)$ is small, at most $p^{1/3}n$. This is of course tight.

One may wonder if a similar geometric property holds for cylints as well. A *cylint* F is of the form

$$F(x,y) = F_x(y,z)F_y(x,z)F_z(x,y).$$

It is a product/intersection of three cylinders, each over one of the three planes. The geometry of cylints is quite complicated. For example, Behrend's construction shows that there is a cylint of measure $n^{1+o(1)}$ so that for every point in every plane, the line over the point contains at most one point in F. Roughly, it is a one-dimensional shape (measure-wise) with projection that is of density is almost full on each of the three planes.

Back to the proof of Roth's theorem. Let $A \subset [n]$ of size $|A| = \delta n$. Our goal is to show that if A does not contain 3-term progressions (i.e. $a + a' - 2a'' = 0$ for $a, a', a'' \in A$ implies $a = a' = a''$) then there is an interval $[0, n']$ for n' large and $c = c(\delta) > 0$ so that

$$|A \cap [n']|/n' = \delta + c.$$

This can clearly happen at most $1/c$ times. We shall not give the full proof here, but only the first step of it (of the two). (Instead of the interval $[n']$ we shall have an a long arithmetic progression, which is good enough.)

Measure the amount of 3-terms progressions in A:

$$\mu = \mathbb{E}_{g,g' \in G} \mathbf{1}_A(g)\mathbf{1}_A(g+g')\mathbf{1}_A(g+2g').$$

Since A contains no 3-AP,

$$\mu = |A|/|G|^2.$$

It will be useful to generalise this

$$\Lambda(f,h,r) = \mathbb{E}_{g,g'} f(g)h(g+g')r(g+2g').$$

In the Fourier basis:

$$\begin{aligned}\Lambda(f,h,r) &= \sum_{\chi_1,\chi_2,\chi_3} \mathbb{E}_{g,g'\in G}\widehat{f}(\chi_1)\chi_1(g)\widehat{h}(\chi_2)\chi_2(g+g')\widehat{r}(\chi_3)\chi_3(g+2g') \\ &= \sum_{\chi_1,\chi_2,\chi_3} \widehat{f}(\chi_1)\widehat{h}(\chi_2)\widehat{r}(\chi_3)\mathbb{E}_{g,g'\in G}\chi_1(g)\chi_2(g)\chi_2(g')\chi_3(g)\chi_3(2g') \\ &= \sum_{\chi_1,\chi_2,\chi_3} \widehat{f}(\chi_1)\widehat{h}(\chi_2)\widehat{r}(\chi_3)\mathbb{E}_{g}\chi_1(g)\chi_2(g)\chi_3(g)\mathbb{E}_{g'}\chi_2(g')\chi_3(2g').\end{aligned}$$

The only terms that contribute are for $\chi_2(g) = \chi_3(-2g)$ and so $\chi_2(g)\chi_3(g) = \chi_3(-g)$ which means $\chi_1(g) = \chi_3(g)$. Thus, using Parseval,

$$\Lambda(f,h,r) = \sum_{\chi_3} \widehat{f}(\chi_3)\widehat{h}(\chi_3(-2\cdot))\widehat{r}(\chi_3) \leq \|\widehat{r}\|_\infty \|f\|_2 \|h\|_2.$$

This clearly holds for any permutation of f, h, r.

We started with that μ is very small. This is not good for us since we want a lower bound on the Fourier coefficients. Write

$$\mathbf{1}_A = f + \delta\chi_0.$$

Write $\Lambda(\mathbf{1}_A, \mathbf{1}_A, \mathbf{1}_A)$ as a sum of eight terms. One of them is

$$\Lambda(\delta\chi_0, \delta\chi_0, \delta\chi_0) = \delta^3,$$

which means that one of the other terms is at least $\delta^3/8$, since $\Lambda(\mathbf{1}_A, \mathbf{1}_A, \mathbf{1}_A)$ is too small to matter. Assume (e.g.)

$$\Lambda(\delta\chi_0, f, f) \geq \delta^3/8.$$

Notice that

$$\|f\|_2, \|\mathbf{1}_A\|_2 \geq \Omega(\delta^{1/2}).$$

Hence, there is χ so that

$$|\widehat{f}(\chi)|\delta \geq \Omega(\delta^3),$$

so

$$|(\widehat{\mathbf{1}_A - \delta}\chi_0)(\chi)| \geq \Omega(\delta^2).$$

This is almost what we want. The above means that A is not fully balanced on χ. If the level sets of χ were intervals, this would mean that A is not balanced on all intervals. The full proof, which we do not include here, follows by showing that this information about characters provides information about a long interval or a long arithmetic progression. The intuition is that characters are well correlated with intervals/arithmetic progressions.

Chapter 3

Representations

We have seen several applications of harmonic analysis on abelian groups. We now move to discuss a generalisation to general groups. We shall focus on finite groups, but the ideas can be applied in a more general context.

3.1 Basic definitions

Again, we start by studying $\mathbb{C}[G]$ for a general finite group G. We will not be able to diagonalise it, but we shall do the best we can. We know that if G acts on X then G acts on $\mathbb{C}[X]$ by

$$g(f(x)) = f(g^{-1}(x)).$$

There is a thus a group homomorphism from G to the invertible linear transformations on $\mathbb{C}[X]$. For a vector space V, denote by $GL(V)$ the set of invertible linear transformations on V.

Definition 10. *A representation π of G is a group homomorphism π from G to $GL(V)$.*

Here we consider only finite dimensional spaces. We shall focus on V over $\mathbb{C}$, but a similar and more technical discussion can be made over other fields (not all statement hold). It will sometimes be convenient to consider matrix representations, which depend on a choice of a basis for V. Then, π_g is $d \times d$ complex matrix where $d = \dim(V)$.

Examples: trivial representation, regular representation, characters of abelian groups, and sign of permutations.

3.2 Irreducibles and decomposition

The first goal of our discussion will be finding the building blocks of all representations of G. Similarly to the way characters decompose $\mathbb{C}[G]$ in the abelian case. The building blocks will be called *irreducible* representations. Irreducible representation are to representation what prime numbers to integers are. We shall see that any finite group has a finite list of building block (up to equivalence).

An eigenspace of a matrix M is invariant under the action of M.

Definition 11. *A subspace U of V is G-invariant if $\pi_g(U) \subseteq U$ for all $g \in G$.*

Definition 12. *A representation π from G to V is irreducible if the only G-invariant subspaces of V are trivial (i.e. $\{0\}$ and V).*

The space V can be decomposed as a direct sum of minimal G-invariant subspaces (details follow).

Theorem 13 (Maschke's theorem)**.** *Let π be a representation of G on V, and let W be a G-invariant subspace of V. Then, there is a complement W_0 of W that is also stable under G.*

Proof. A key idea is to symmetrize objects. Write $V = W \oplus W'$ for some W'. Denote by p the projection from V to W. Define

$$p_0 = \frac{1}{|G|} \sum_{g \in G} \pi_g p \pi_{g^{-1}}.$$

We have

$$p_0(V) \subset W.$$

For $w \in W$, we have

$$p_0(w) = w.$$

The map p_0 is also a projection from V to W. By symmetry,

$$\pi_g p_0 = p_0 \pi_g,$$

for all $g \in G$. Let W_0 be the kernel of p_0. Thus, for all $w_0 \in W_0$,

$$0 = \pi_g(p_0(w_0)) = p_0(\pi_g(w_0)).$$

So $\pi_g(W_0) \subset W_0$ for all $g \in G$. □

Corollary 14. *Every representation is a direct sum of irreducible representations.*

Proof. Let $\pi : G \to GL(V)$. If V is irreducible, we are done. Otherwise, let W be a nontrivial invariant subspace of V. Let W_0 be its complement by the theorem above. We have

$$V = W \oplus W_0$$

with W, W_0 are representation of G of smaller dimension. Continue by induction. □

Weyl's trick. The case of unitary representations (when the image is contained is the set of unitary maps in $GL(V)$ so inner products are preserved): The symmetrisation above can be done also to unitarize representations. If $\langle v, u\rangle$ is an inner product over V, then

$$\langle v, u\rangle_0 = \frac{1}{|G|}\sum_{g\in G}\langle \pi_g v, \pi_g u\rangle$$

is a new inner product over V for which π is unitary.

We aim at classifying all irreducible matrix representations. There are of course infinitely many since we can choose the basis arbitrarily.

Definition 15. *Two representation π, π' from G to $GL(V)$ are equivalent if there is $h \in GL(V)$ so that $\pi h = h\pi'$.*

Let V be a G-space (that is, there is a representation $\pi : G \to GL(V)$). We now know that we can decompose V as

$$V = \bigoplus_a m_{\pi_a} V_{\pi_a}.$$

The decomposition is to non-equivalent irreducible representations of G, where m_{π_a} is the number of copies of representations that are equivalent to π_a.

3.3 Schur's lemma

In general matrices do not commute. There are however matrices that do commute. Scalar matrices (homotheties) commute with all other matrices. Schur's lemma says that this is the only option for irreducible representations.

Lemma 16. *Let $\pi : G \to GL(V)$ be an irreducible representation. Let $A : V \to V$ be a linear map that commutes with G (such maps are called* intertwiners*). Then, $A = \lambda I$ for some $\lambda \in \mathbb{C}$.*

The lemma is false if π is reducible (e.g. abelian groups). We learn that the only irreducible representations of an abelian group are 1-dimensional: If π is irreducible, then every π_h is an intertwiner, which means that $\pi_h = \lambda_h I$. Since π is irreducible, $\mathrm{span}(v) = V$ for all $v \neq 0$.

We prove the lemma over algebraically closed fields but there are analogs for other fields as well.

Proof. Let λ be an eigenvalue of A, and let V_λ be the corresponding subspace. Thus, for $v \in V_\lambda$,

$$A\pi_g v = \pi_g A v = \lambda \pi_g v.$$

This means that V_λ is invariant so it is either 0 or V. □

We may deduce the following.

Lemma 17. *Let π_1, π_2 be two irreducible representation of G. If π_1, π_2 are not equivalent then the space of intertwiners is $\{0\}$. If $\pi_1 = \pi_2$ then the space of intertwiners is one-dimensional.*

Proof. Let $A \in Hom_G(V_1, V_2)$. Thus, $ker(A)$ is a trivial subspace of V_1 and $im(A)$ is a trivial subspace of V_2. So, either $A = 0$ or A is injective and surjective (invertible).

Therefore, if V_1, V_2 are not isomorphic then $Hom_G(V_1, V_2) = \{0\}$. Otherwise, they are isomorphic. Let A_1, A_2 be invertible linear maps from V_1 to V_2 that commute with G. Let $A = A_1 A_2^{-1}$. Thus, $A \in End_G(V, V)$ which means that A is a homothety, and $A_2 = \lambda A_1$. □

3.4 Orthogonality of matrix entries

Assume V is equipped with an inner product. Given $v, u \in V$, define $\pi_{v,u} : G \to \mathbb{C}$ by

$$\pi_{v,u}(g) = \langle \pi_g(v), u \rangle.$$

This is a generalisation of matrix entries. Recall the inner product on $\mathbb{C}[G]$:

$$\langle f, h \rangle = \frac{1}{G} \sum_{g \in G} f(g)\overline{h(g)}.$$

Theorem 18. *Let π_1, π_2 be two unitary irreducible representations of G. For all relevant $v_1, u_1 \in V_1$ and $v_2, u_2 \in V_2$,*

$$\langle (\pi_1)_{v_1,u_1}, (\pi_2)_{v_2,u_2} \rangle = \begin{cases} 0 & \pi_1 \text{ is not isomorphic to } \pi_2, \\ \frac{1}{dim(V_1)} \langle v_1, v_2 \rangle \overline{\langle u_1, u_2 \rangle} & \pi_1 = \pi_2. \end{cases}$$

This means that the matrix entries of two non-isomorphic representations are orthogonal, and that the matrix entries of a given representation are orthonormal (with a given normalisation).

Proof. For fixed u_1, u_2, the l.h.s. is a bilinear map on (v_1, v_2) which is G-invariant. There is thus an $A \in Hom_G(V_1, V_2)$ so that l.h.s. is equal to $\langle Av_1, v_2 \rangle$. Schur's theorem implies that if V_1, V_2 are not isomorphic then $A = 0$.

Assume $\pi = \pi_1 = \pi_2$. The l.h.s. is equal to $\langle A_{u_1,u_2} v_1, v_2\rangle$ for $A_{u_1,u_2} = \lambda_{u_1,u_2} I$. The map $(u_1, u_2) \mapsto \lambda_{u_1,u_2}$ is also bilinear and G-invariant. So, by Schur's theorem, $\lambda_{u_1,u_2} = \lambda\overline{\langle u_1, u_2\rangle}$. It remains to compute λ. Let $e_1, \ldots, e_d$ be the standard basis. Thus, for all $i \in [d]$ and $v \in V$,

$$\langle \pi_{v,e_i}, \pi_{v,e_i}\rangle = \lambda\langle v, v\rangle.$$

Sum over i, since π is unitary,

$$\langle v, v\rangle = \frac{1}{G}\sum_{g\in G}\sum_{i\in[d]} \left|\langle \pi_g(v), e_i\rangle\right|^2 = \sum_{i\in[d]}\langle \pi_{v,e_i}, \pi_{v,e_i}\rangle = \lambda d\langle v, v\rangle.$$

□

3.5 Characters

Instead of studying $\mathbb{C}[G]$, it will be somewhat easier to study $\mathbb{C}(G)$, the space of class functions. That is, the space of complex functions on G that are constant on conjugacy classes (for all $t, g \in G$, we have $f(tgt^{-1}) = f(g)$). Over fields of characteristic zero this is almost equivalent to $\mathbb{C}[G]$ ([Frobenius] initiator). Over fields of finite characteristic the discussion is more complicated ([Brauer]).

Recall that the conjugacy classes of G are equivalence classes, which means that they partition G. The dimension of $\mathbb{C}(G)$ is therefore the number of conjugacy classes of G.

A basic useful fact is that matrix trace is independent of conjugation. That is, trace does not depend on choice of basis. Let A be a $n \times n$ matrix. Its trace is defined as

$$Tr(A) = \sum_i A_{i,i}.$$

It is a linear map. The following property is very useful:

$$Tr(AB) = Tr(BA) = \sum_{i,j} A_{i,j}B_{j,i}.$$

This means e.g. that it does not depend on choice of basis: If U is invertible then

$$Tr(UAU^{-1}) = Tr(U^{-1}UA) = Tr(A).$$

Definition 19. *Let π be a matrix representation of G. The corresponding character $\chi_\pi : G \to \mathbb{C}$ is defined as*

$$\chi_\pi(g) = Tr(\pi_g).$$

Denote by $\widehat{G}$ the set of irreducible characters of G.

This set will not be a full basis for $\mathbb{C}[G]$ as in the abelian case but an important subset of it. It is a basis for $\mathbb{C}(G)$ (as we shall see).

Proposition 20 (Basic properties)**.** *For all $t, g \in G$:*

- *If χ is a character of π of degree (dimension) d then $\chi(1) = d$.*
- $\overline{\chi(g)} = \chi(g^{-1})$.
- *Class functions: $\chi(tgt^{-1}) = \chi(g)$.*

Proof. First, $\chi(1) = Tr(I) = d$. Second, π_g has finite order which means that it has n eigenvalues of absolute value one. So,

$$\overline{\chi(g)} = \sum_i \overline{\lambda_i} = \sum_i \lambda_i^{-1} = \chi(g^{-1}).$$

Third, by the above property of trace,

$$Tr(\pi_t \pi_g \pi_{t^{-1}}) = Tr(\pi_g).$$

□

Proposition 21 (Characters and representation)**.**

- *If π_1, π_2 are irreducible representations then $\langle \chi_{\pi_1}, \chi_{\pi_2} \rangle \in \{0, 1\}$ and the value one is attained iff π_1, π_2 are equivalent.*
- *Given an irreducible representation π_1, and a representation π_2, denote by $m_{\pi_1}(\pi_2)$ the number of copies of π_1 in π_2. Then $\langle \chi_{\pi_1}, \chi_{\pi_2} \rangle = m_{\pi_1}(\pi_2)$.*
- *If $\pi = \oplus_a m_a \pi_a$ with π_a irreducible then $\|\chi_\pi\|^2 = \sum_a m_a^2$. Specifically, $\|\chi_\pi\| = 1$ iff π is irreducible.*
- *π_1, π_2 are equivalent iff $\chi_{\pi_1} = \chi_{\pi_2}$.*

Proof.

- Assume π_1, π_2 are irreducible representations. Then,

$$\begin{aligned} \langle \chi_{\pi_1}, \chi_{\pi_2} \rangle &= \frac{1}{|G|} \sum_g Tr(\pi_1(g)) \overline{Tr(\pi_2(g))} \\ &= \sum_{i,j} \frac{1}{|G|} \sum_g (\pi_1(g))_{i,i} \overline{(\pi_2(g))_{j,j}}. \end{aligned}$$

If π_1, π_2 are not equivalent then for every i, j this is zero due to entry-wise orthogonality. If π_1, π_2 are equivalent then by the orthogonality lemma

$$= \sum_i \frac{1}{\dim(V_1)} = 1.$$

- If $\pi_2 = \oplus_a m_a \pi_a$ with π_a irreducible then $\chi_{\pi_2}(g) = \sum_a m_a \chi_{\pi_a}$. The claim holds by previous bullet.

- If π_1, π_2 are equivalent then clearly $\chi_{\pi_1} = \chi_{\pi_2}$. In the other direction, if $\chi_{\pi_1} = \chi_{\pi_2}$ then for all irreducible π_a we have $m_a(\pi_1) = m_a(\pi_2)$ which implies the claim.

□

3.6 Structure of regular representation

We conclude that the regular representation contains all irreducible representations. In fact, we also know the multiplicities are the dimensions.

Theorem 22. *Let ρ be the regular representation of G. Let π be an irreducible representation of G. Then $m_\pi(\rho) = dim(\pi)$.*

Proof.

$$m_\pi(\rho) = \langle \chi_\pi, \chi_\rho \rangle = \frac{1}{|G|} \sum_g Tr(\pi_g)\overline{Tr(\rho_g)} = Tr(\pi_1) = \dim(\pi),$$

where here $1 \in G$. □

Corollary 23. *Let $\{\pi_a\}$ be the set of (non-equivalent) irreducible representations of G. Then $|G| = \sum_a dim^2(\pi_a)$.*

Proof. Counting dimensions in two ways. □

Corollary 24. *If π is an irreducible representation of G then $dim(\pi) < |G|^{1/2}$.*

3.6.1 Class functions

The following lemma shows that the characters form a basis to the space of class functions. We know it is an orthonormal set (and hence independent), and we prove it is spanning.

Lemma 25 (Characters are basis to class functions)**.** *If f is a class function so that $\langle f, \chi \rangle = 0$ for all $\chi \in \widehat{G}$ then $f = 0$.*

Proof. Fix an irreducible π. Define a linear map $A = A(f, \pi)$ by

$$A = \frac{1}{|G|} \sum_g f(g) \pi_g.$$

It is an intertwiner: for every $g' \in G$,

$$\pi_{g'} A \pi_{g'^{-1}} = \frac{1}{|G|} \sum_g f(g) \pi_{g' g g'^{-1}} = \frac{1}{|G|} \sum_g f({g'}^{-1} g g') \pi_g = A,$$

since f is a class function. We thus conclude that

$$A = \lambda I.$$

Since f is orthogonal to characters, $\lambda = 0$.

Now, let π be the regular representation of G. We know that $\sum_g f(g) \pi_g = 0$ which implies that $f = 0$. □

Corollary 26. *The number of irreducible characters is equal to the number of conjugacy classes.*

3.7 Applications

3.8 Quasi-random groups

We start with an application to additive combinatorics/group theory due to Gowers. Babai and Sos asked whether for every finite group G, there is a sum-free subset $S \subset G$ of size $|S| > c|G|$, $c > 0$ a constant. That is, for all $x, y, z \in S$ we have $xy \neq z$. Motivation for studying this question comes from a result of Erdos, who proved it for $\mathbb{Z}$ and basically all abelian groups.

Erdos actually proved that for all $X \subset \mathbb{Z}$ of size n, there is a sum-free subset $S \subset X$ of size at least $n/3$. Here is the proof. Let p be the smallest prime so that $X \subset [-p/3, p/3]$. For any $r \in \mathbb{Z}$ so that $r \neq 0 \mod p$, the sets Y and rY are sum-free together modulo p. The set $T = [p/3, 2p/3] \cap \mathbb{Z}_p$ is sum-free modulo p. Its fraction of $\mathbb{Z}_p$ is $1/3$ which means that there is r so that $Y' = rX \cap T$ is large $|Y'| \geq n/3$. Set $Y = r^{-1}Y'$.

Gowers shows that the answer to this question is negative. Since it is true for abelain groups, the examples should be far from abelian. One should look at simple non-abelian group. Representation theory allows to measure the "distance" from abelian group. All irreducible representations of abelian groups has dimension one. One should look at groups for which all nontrivial irreducible representations have large dimension.

Definition 27. *A group G is D-*quasi-random *if every non-trivial irreducible representation of G has dimension at least D.*

A well-known example for such a group is $G = PSL_2(p)$, the group of 2×2 matrices over $\mathbb{F}_p$ of determinant one, modulo the normal subgroup $\{I, -I\}$. It is a simple group. Moreover, Frobenius proved that if π is an irreducible representation of G then the dimension of π is at least $(p-1)/2$. We shall not prove this fact here (a classification of *all* finite simple groups is known). This size of G is roughly p^3, which means that all irreducible representation have dimension at most $p^{3/2}$.

(In the exercise (http://tx.technion.ac.il/∼yehuday/analyticMethods/ex3.pdf) we shall see how to prove a lower bound on the dimension of irreducible representations for simple groups. The beautiful idea [Gowers] is to interpret an irreducible representation as a code, and use volume arguments.)

3.8.1 Mixing property

The proof of the theorem relies on a spectral property of quasi-random groups sometimes called the mixing property. This means establishing a bound on the norm of convolutions. The following theorem proved by Babai, Nikolov and Pyber summarizes the statement (using multiplicity in this way is sometimes called Sarnak-Xue multiplicity argument).

The basic idea is to view convolution as a linear map, and to use representation theory to study its spectrum, using our knowledge about the regular representation.

Theorem 28. *Assume G is a D-quasi-random group that acts transitively on X. Let $\mu : G \to \mathbb{R}$ and $f : X \to \mathbb{R}$ be so that $\sum_{x \in X} f(x) = 0$. Then,*

$$\|\mu * f\|_2^2 \leq \frac{|G|}{D} \|\mu\|_2^2 \|f\|_2^2$$

where

$$(\mu * f)(x) = \sum_{g \in G} \mu(g) f(g^{-1}(x)).$$

Proof. Let $\rho_1, \ldots, \rho_t$ be the list of all (non-isomorphic) irreducible representations of G, where ρ_1 is the trivial representation. Let ρ be the regular representation of G. We know that $\rho = \oplus_i c_i \rho_i$. Denote by $d_i = c_i$ the dimension of ρ_i.

We use these properties of the regular representation to bound the relevant singular values. First, some definitions. For a $k \times k$ complex matrix M, denote by M^* the conjugate transpose of M. The matrix MM^* is positive semi-definite, and has k non-negative eigenvalues $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k \geq 0$. Denote

$$\sigma(M) = \sigma_1.$$

Denote

$$M_i = \sum_{g \in G} \mu(g) \rho_i(g)$$

for $1 \leq i \leq t$. After a basis change, the matrix

$$R = \sum_{g \in G} \mu(g) \rho(g)$$

is in block-diagonal form with blocks from the list $M_1, \ldots, M_t$, each M_i has multiplicity c_i. Denote by Q the $|G| - 1 \times |G| - 1$ matrix with blocks of the form $M_2, \ldots, M_t$, after deleting from R a row and a column that correspond to the trivial representation ($c_1 = d_1 = 1$). On one hand, quasirandomness implies (multiplicity method)

$$\text{trace}(QQ^*) = \sum_{2 \leq i \leq t} c_i \text{trace}(M_i M_i^*) \geq D \cdot \sigma(Q).$$

On the other hand,

$$\operatorname{trace}(QQ^*) \leq \operatorname{trace}(RR^*) = \sum_{g_1 \in G} \sum_{g_2 \in G} \mu(g_2 g_1^{-1})^2 = |G| \, \|\mu\|_2^2 \, .$$

We conclude that

$$\max\{\sigma(M_i) : 2 \leq i \leq t\} = \sigma(Q) \leq \|\mu\|_2^2 \, |G|/D.$$

Finally, we use the bound on singular values to prove the theorem. Consider the representation π of G induced by the action of G on X. After a basis change, the matrix $R' = \sum_{g \in G} \mu(g)\pi(g)$ can be written in block-diagonal form with blocks that are copies of $M_1, \ldots, M_t$, each of the blocks appears in R' with multiplicity c_i'.

We claim that $c_1' = 1$, since G acts (one-) transitively on X:

$$c_1' = \langle \chi_\pi, \mathbf{1} \rangle = \frac{1}{|G|} \sum_{g \in G} \sum_{x \in X} \pi(g)_{x,x} = \frac{1}{|G|} \sum_{x \in X} \big|\{g \in G : g(x) = x\}\big| = 1.$$

The last equality holds due to transitivity, which implies that $\big|\{g \in G : g(x) = x\}\big| = \big|\{g \in G : g(x) = x'\}\big|$ for every x, x' in X.

Since $\langle f, \mathbf{1} \rangle = \sum_x f(x) = 0$ and since the all-ones vector spans the subspace corresponding to the trivial representation,

$$\|\mu * f\|_2^2 = \langle R'f, R'f \rangle \leq \sigma(Q) \, \|f\|_2^2 \, .$$

□

3.8.2 All large sets are not sum free in quasi-random groups

Gowers used that quasi-randomness implies mixing to prove the following.

Theorem 29. *Let G be D-quasi-random of size n. Let $A, B \subset G$ be of size $|A| = an$ and $|B| = bn$. Then*

$$|AB| \geq n \left(1 - \frac{1}{Dab}\right).$$

It follows that if A, B, C are of size at least cn, and $D > c^3$ then

$$|(AB) \cap C| \geq |AB| + |C| - n \geq n \left(c - \frac{1}{Dc^2}\right) > 0.$$

This means that in $G = PSL_2(p)$ for example every set of size at least order $p^{3-1/3}$ ($\ll |G|$) is not sum-free.

Proof. Denote

$$f = \mathbf{1}_B - b$$

so $\sum_g f(g) = 0$ and $\|f\|_2^2 = b(1-b)n \leq bn$. By the mixing lemma,

$$\|\mathbf{1}_A * f\|_2 \leq \frac{n}{D} anbn.$$

But

$$\mathbf{1}_A * f = \mathbf{1}_A * \mathbf{1}_B - abn.$$

Denote by m the size of the complement of AB. By the above, if $(\mathbf{1}_A * \mathbf{1}_B)(g) = 0$ then $g \notin AB$ and $(\mathbf{1}_A * f)(g) = -abn$. Thus,

$$m(abn)^2 \leq \frac{n}{D} anbn,$$

or

$$m \leq \frac{n}{Dab}.$$

□

3.9 Card shuffling

We now discuss mixing a deck of cards (following Bayer-Diaconis). Consider a deck of $n = 52$ cards. It is interesting to understand how fast will a specific shuffle of a deck of cards will make it close to random. It will provide some motivation to study the representation theory of the symmetric group.

The Gilbert, Shannon, and Reeds (GSR) model for card shuffling has several equivalent definitions:

1. (Sequential) Cut the deck to two piles, where the size of one of the piles is $Bin(n, 1/2)$. Mix the two piles together so that if at some point in time one of the (smaller) piles have A cards and the other B, then the chance of the next card being from A is $A/(A+B)$.

2. (Geometric) Drop n points $X_1, \ldots, X_n$ independently and uniformly into $I = [0, 1]$. Use the baker's map $X_1 \mapsto 2X_1 \mod I$ to get n new points $Y_1, \ldots, Y_n$. Assume $X_1 < \ldots < X_n$ and $Y_1 < \ldots < Y_n$. We get a permutation on $[n]$ by $X_i \mapsto Y_{\sigma(i)}$.

3. (Maximum entropy) Every partition of the deck to two parts and then interleaving the two parts is equally likely (empty parts are possible).

4. (Inverse) The following random way to generate the inverse. Mark every card $\{0, 1\}$ independently and uniformly at random. Move all cards that are labelled 1 to the top of the deck.

The identity permutation has probability $(n+1)/2^n$ and all other possible permutations have probability 2^{-n}.

It will be useful to generalise this process. Instead of cutting to 2 parts, we shall cut is to $a \geq 2$ parts:

1. (Sequential) Cut the deck to a piles of sizes $j_1, \ldots, j_a$ that are distributed multinomially with parameter 1/2. Mix the a piles together so that if at some point in time the sizes are $a_1, \ldots, a_j$ then then the chance of the next card being from pile i is $a_i/(a_1 + \ldots + a_j)$.

2. (Geometric) Drop n points $X_1, \ldots, X_n$ independently and uniformly into $I = [0, 1]$. Use the baker's map $X_1 \mapsto aX_1 \mod I$ to get n new points $Y_1, \ldots, Y_n$. Assume $X_1 < \ldots < X_n$ and $Y_1 < \ldots < Y_n$. We get a permutation on $[n]$ by $X_i \mapsto Y_{\sigma(i)}$.

3. (Maximum entropy) Every partition of the deck to a parts and then interleaving the two parts is equally likely (empty parts are possible).

4. (Inverse) Mark every card $\{1, 2, \ldots, a\}$ independently and uniformly at random. Reorder the deck according to marks.

Claim 30. *The 4 ways above to define Q are equivalent.*

Proof. We shall not prove it here. □

Denote by $Q = Q_a \in \mathbb{C}[G]$ the probability distribution that corresponds to this shuffling, for the permutation group $G = S_n$. Repeating this shuffling process corresponds to convolution. After k times, the distribution is Q^{*k}. It can be shown that the relevant graph is connected and aperiodic and so $Q^{*k} \to u$ when $k \to \infty$, where u is the uniform distribution. It is interesting to understand get a quantitative bound. We will look for the L_1 mixing time.

Theorem 31 (Bayer-Diaconis)**.** *For $a = 2$,*

k	4	5	6	7	8	9	10
$\lvert Q^{*k}(p) - u\rvert_1$	1.00	0.92	0.61	0.33	0.16	0.08	0.04

For general n, and $k = (3/2)\log_2(n) + c$, they show that distance is roughly one minus the probability that a normal gaussian gets a value larger than 2^{-c}.

The following property is crucial. It says that the matrices Q_a, $a \geq 2$, commute. Moreover, $Q_2^{*k} = Q_{2^k}$.

Claim 32. *An a-shuffle followed by a b-shuffle is equivalent to an ab-shuffle. That is, $Q_a * Q_b = Q_{ab}$.*

Proof. Follows from the geometric description. □

It thus suffices to understand Q_a for general a. The following definition is useful:

Definition 33. *A* rising sequence *in σ is a maximum interval $i, i+1, \ldots, i+k-1$ of $[n]$ so that $\pi(j) < \pi(j+1)$ for all j in the sequence.*

Example: 716248359 has four rising sequences: 123, 45, 6, and 789. They are basic concepts in some magic tricks (see page 297 in [BD]). They form a partition of the universe.

Claim 34. *If σ has r rising sequences then $Q_a(\sigma) = \binom{a+n-r}{n}/a^n$.*

Proof. The number of rising sequences r is at most a. The r rising sequences in σ surely come from cuts. The rest $a - r$ cuts must be chosen. We think of them as "markers" between $n + a - r$ objects. The total number of choices is thus $\binom{n+a-r}{n}$ and the number of a-shuffles is a^n. □

Corollary 35. *For every permutation σ, we have $Q_2^{*k}(\sigma) = \binom{2^k+n-r}{n}/2^{kn}$ where r is the number of rising sequences in σ. Specifically,*

$$\left|Q_2^{*k} - u\right|_1 = (1/2)\sum_r k(r)\left|\binom{2^k+n-r}{n}/2^{kn} - 1/n!\right|,$$

where $k(r)$ is the number of permutation with r rising sequences.

The numbers $k(r)$ are the Eulerian numbers [Stanley]. Finding the asymptotic behaviour of the mixing time requires elaborate calculations.

3.9.1 Descends

We briefly discuss a connection to descend theory, studied by Stanley and others. It has connections to shuffling, juggling and more.

A permutation σ has a *descent* at i if $\sigma(i) > \sigma(i+1)$. By $D(\sigma) \subseteq [n-1]$ we denote the set of descends of σ. Solomon observed that descends yield a sub algebra of $\mathbb{C}[S_n]$. For $S \subseteq [n-1]$, denote

$$a_S = \sum_{\sigma : D(\sigma)=S} \sigma \in \mathbb{C}[S_n].$$

Solomon showed that

$$a_S a_T = \sum_R c_{S,T}^R a_R$$

for $c_{S,T}^R \in \mathbb{Z}$. This is a topic by itself that we shall not address here.

We may consider a more symmetric algebra generated by $A_0, \ldots, A_{n-1}$ defined by $A_0 = id$ and

$$A_i = \sum_{\sigma : |D(\sigma)|=i} \sigma.$$

This algebra is closely related to GSR shuffling. The first step is to note that the set of permutations σ with $D(\sigma) = \{i\}$ is exactly the set of permutations σ^{-1} obtained by choosing a subset of size i from $[n]$ and moving it to the top of the deck (except for id): Denote by τ a permutation defined by choosing $J = \{j_1 < \ldots < j_i\} \subset [n]$ and moving it to top. Thus, τ^{-1} preserves the order on $[i]$ and on $\{i+1, \ldots, n\}$, and switches the order between $i, i+1$, except when $J = [i]$. This is exactly the process done in the inverse way to sample a GSR shuffling. Thus,

$$\sum_\sigma Q(\sigma^{-1})\sigma = \frac{n+1}{2^n} + A_1.$$

This means that repeated GSR shuffling corresponds to multiplication inside this algebra.

It enjoys some interesting algebraic topology properties.

3.10 Logrank and the symmetric group

We describe an application of representation theory to communication complexity due to Raz and Spieker. We will start exploring the representation theory of S_n on the way.

The setup is two party deterministic communication complexity. Let $f : [n]^2 \to \{0,1\}$. There are two parties, Alice and Bob. Alice gets $a \in [n]$ and bob gets $b \in [n]$, and they wish to compute $f(a,b)$. A protocol P is an exchange of bits (messages) between Alice and Bob. The message Alice sends e.g. depend on a and on the messages previously sent. The last bit sent in P is the function it computes. Denote by $|P|(a,b)$ the number of bits exchanged with inputs a, b. Denote $|P| = \max_{a,b} |P(a,b)|$. Denote

$$D(f) = \min\{|P| : P \equiv f\}.$$

Understanding $D(f)$ is theoretically interesting and related to other theoretical and practical problems. A standard way to lower bound $D(f)$ is using matrix rank. Think of f as a boolean $n \times n$ matrix M_f. A protocol for f with complexity c yield a partitions of the ones of the matrix to at most 2^c disjoint rectangles. Thus,

$$D(f) \geq \log rank(M_f),$$

where $rank$ is over say the reals. The log rank conjecture [Lovasz-Saks] states that this is basically tight: there is $c > 0$ so that for all boolean f,

$$D(f) \leq c \log^c rank(M_f).$$

This has been a long standing open problem.

The first question is whether c can be one (at least in the exponent). The first result refuting this option is the work of [RS] that uses representation theory of S_n we discuss. Later, Kushilevitz showed that c must be at least $\log_3 6 \approx 1.6$ using Fourier analysis.

On a high level, the conjecture is about establishing a structure for low rank boolean matrices. Refuting the conjecture is about finding “non trivial” low rank boolean matrices. We shall see how to use representation theory to achieve this.

3.10.1 The example

In the example both Alice and Bob get permutations. So $a, b \in S_n$. The underlying matrix is thus $n! \times n!$.

Denote by $K \subset S_n$ the set of n-cycles. It is a conjugacy class, so $\mathbf{1}_K$ is spanned by the characters, which will be helpful. From it generate a two-input function by:

$$f(a,b) = \mathbf{1}_K(a^{-1}b).$$

Theorem 36. *$D(f) \geq \Omega(n \log\log(n))$ and $rank(M_f) = \binom{2n-2}{n-1} \leq 2^{2n}$.*

We shall not prove the lower bound on $D(f)$ here.

3.10.2 Computing the rank

Representation theory allows to compute the rank. Here are two useful properties. Let ρ be the regular matrix representation of S_n. The matrix $M = M_f$ is

$$M = \sum_{\sigma \in K} \rho(\sigma).$$

Since K is a conjugacy class, M is an interwiner: for all σ,

$$\rho(\sigma) M \rho(\sigma^{-1}) = M_f.$$

(A version of this holds for all class functions.) We are able to use Shcur's lemma as follows.

Let $\rho_0, \ldots, \rho_k$ be the list of irreducible representations of S_n. Denote by $1 = d_0, \ldots, d_k$ their dimensions. We know

$$n! = d_0^2 + \ldots + d_k^2.$$

We also know the multiplicity of ρ_i in ρ is d_i as well.

We know that there is a unitary matrix U so that UMU^{-1} is block diagonal, where every matrix on the diagonal is a copy of

$$M_i = \sum_{\sigma} \mathbf{1}_K(\sigma) \rho_i(\sigma).$$

Since M is an interwiner, M_i is also an interwiner. By Schur's lemma, we learn that M_i is a scalar matrix for all i. Denote

$$M_i = c_i I$$

for all i.

The rank of M is now well understood. Denote by $0, 1, \ldots, \ell$ the set of i so that $c_i \neq 0$. Then,

$$rank(M) = \sum_{i=0}^{\ell} d_i^2.$$

We just need to understand for which i the matrix M_i is nonzero. This what we shall do next.

3.11 Representations of the symmetric group

We have seen that there is a one-to-one correspondence between irreducible representations and characters, and between characters and conjugacy classes.

Conjugacy classes of S_n are determined by the cycle structure of S_n. The number of cycle structures is exactly the number of partitions of n. That is, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ so that $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_\ell \geq 1$ so that $|\lambda| := \sum_i \lambda_i = n$. This is denoted by $\lambda \vdash n$.

We shall describe a concrete correspondence between partitions and irreducible representation [Sagan, James].

We start by a concrete description of the vector spaces on which G shall act. We thus describe representation of G. We shall start with "big" representations and shall then reduce the dimension until we get irreducible ones. The dimension reduction will be done by dividing by certain subgroups.

Definition 37. *A* Ferrer *diagram of shape $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ is a collection of n cells with ℓ rows, where in row i from top there are exactly λ_i cells (aligned to left).*

A (Young) tableau *is a (one-to-one) filling of a diagram with the numbers* 1 *to* n.

A standard *tableau is such that the number if every row increase (to right) and in every column increase (to bottom).*

Examples for $\lambda = (3, 2, 2, 1)$.

A permutation σ acts on a tableau t: The entry $t(i, j)$ is replaced by $\sigma(t(i, j))$ for suitable (i, j). We may thus consider the action of S_n on the space of complex functions on λ-tableaux. This is a representation of S_n, but not an irreducible one.

Definition 38. *Define an (row) equivalence relation on the space of λ-tableaux: $t_1 \equiv t_2$ if there is a permutation σ that fixes the rows of t_1 so that $\sigma t_1 = t_2$. An equivalence class $\{t\}$ is called a* tabloid.

The number of tableaux in a λ-tabloid $\lambda! := \lambda_1!\lambda_2! \ldots \lambda_\ell!$. The number of λ-tabloids is $n!/\lambda!$.

Permutations act on tabloids as well

$$\sigma\{t\} = \{\sigma t\}$$

(needs to be verified). This gives a representation of S_n on

$$M^\lambda := \mathbb{C}(\{t_1\}, \ldots, \{t_k\}),$$

where $\{t_1\}, \ldots$ is the list of λ-tabloids.

Examples: $M^{(n)} \cong \mathbb{C}$ with trivial representation, $M^{(1^n)} \cong \mathbb{C}[S_n]$ with regular representation, and $M^{(n-1,1)} \cong \mathbb{C}[\{1, \ldots, n\}]$ with the *defining* representation.

These are still not necessarily irreducible. We shall now define the irreducible representations of S_n.

3.11.1 Specht modules

The Specht modules which we now define are the irreducible representations.

Definition 39. *Let t be a tableau with rows $R_1, \ldots, R_t \subset [n]$ and columns $C_1, \ldots, C_k \subset [n]$. Define the* row stabiliser *as*

$$R_t = S_{R_1} \times \cdots S_{R_t}$$

Define the columns stabiliser *as*

$$C_t = S_{C_1} \times \cdots S_{C_k}.$$

Example: for $\lambda = (3, 2)$, we have $|R_t| = 3! \cdot 2!$ and $|C_t| = 2! \cdot 2! \cdot 1!$.

Definition 40. *Given t, define*

$$\kappa_t = \sum_{\sigma \in C_t} sign(\sigma)\sigma \in \mathbb{C}[S_n]$$

and a polytabloid *as*

$$e_t = \kappa_t\{t\} \in M^\lambda.$$

Example for λ as above, where κ_t, e_t are sums of four tabloids.

Definition 41. *The Specht modules are of the form*

$$\mathcal{S}^\lambda = span_{\mathbb{C}}\{e_t : t \text{ a } \lambda\text{-tableau}\}.$$

Begin by showing that they are representations of S_n.

Lemma 42.

1. *$R_{\sigma t} = \sigma R_t \sigma^{-1}$, $C_{\sigma t} = \sigma C_t \sigma^{-1}$. and $\kappa_{\sigma t} = \sigma \kappa_t \sigma^{-1}$*

2. *$\sigma e_t = e_{\sigma t}$.*

Proof.

1. Consider R for example: $\tau \in R_{\sigma t}$ iff $\tau\{\sigma t\} = \{\sigma t\}$ iff $\sigma^{-1}\tau\sigma\{t\} = \{t\}$ iff $\tau \in \sigma R_t \sigma$.

2.

$$e_{\sigma t} = \kappa_{\sigma t}\{\sigma t\} = \sigma \kappa_t \sigma^{-1}\{\sigma t\} = \sigma \kappa_t \{t\} = \sigma e_t.$$

□

The lemma implies that S_n acts on $\mathcal{S}^\lambda$ by permuting the e_t's. This implies that M^λ is indeed a representation.

Theorem 43. *The subspace $\mathcal{S}^\lambda \subset M^\lambda$ is irreducible.*

We shall not prove this right now (see below).

The set $\{e_t\}$ spans $\mathcal{S}^\lambda$ but it is not a basis. The following theorem describes a basis.

Theorem 44. *The set $\{e_t :\ t \text{ is a standard } \lambda\text{-tableau}\}$ is a basis for $\mathcal{S}^\lambda$.*

We shall not prove the theorem right now either. The proof is very informative. It shows e.g. that the action of G on $\mathcal{S}^\lambda$ can be represented by upper triangular matrices with ±1 on the diagonal.

Examples: $(1^n), (n), (n, 1)$.

3.11.2 Back to logrank

Recall that we started this discussion by trying to compute the rank of a given matrix (corresponding to the class of n-cycles K). We need to examine which characters have nonzero inner product with $\mathbf{1}_K$. This is known (perhaps we shall prove it later on):

$$\langle \chi_{\mathcal{S}^\lambda}, K\rangle \neq 0 \ \Leftrightarrow\ \lambda \text{ has one turn (i.e. } \lambda = (n-j, 1^j) \text{ for some } j \in \{0, \ldots, n-1\}).$$

The last step towards understanding the rank is figuring out the dimension d_λ of $\mathcal{S}^\lambda$ for $\lambda = (n-j, 1^j)$. It is exactly the number of standard tableau of this shape. That is,

$$d_\lambda = \binom{n-1}{j}$$

(which j elements appear in the single column). Overall,

$$rank(M) = \sum_{j=0}^{n-1} \binom{n-1}{j}^2 = \sum_{j=0}^{n-1} \binom{n-1}{j}\binom{n-1}{n-1-j} = \binom{2n-2}{n-1},$$

as claimed.

3.11.3 Orders on partitions

There are two orders on partitions that are useful in understanding the Specht modules.

Definition 45 (Dominance order)**.** *Let λ, λ' be partitions of n. We write $\lambda \gg \lambda'$ if for all i,*

$$\lambda_1 + \ldots + \lambda_i \geq \lambda'_1 + \ldots + \lambda'_i.$$

That is, if "λ is wider than λ' on every height." Example $(4,3)$ is larger than $(3,2,1^2)$. Sometimes a different notation is used.

Lemma 46 (Dominance)**.** *Let t, t' be tableaux of shape λ, λ'. If for every i the element in row i of t' appear in different columns in t then $\lambda \gg \lambda'$.*

Proof. By reordering the columns of t, going over λ' row-by-row, move the elements that appear in first i rows in λ' so that they appear in first i rows of λ. □

The dominance order is partial. A linear order that extends it is (exercise)

Definition 47 (Lexicographic order)**.** *We write $\lambda \geq \lambda'$ if $\lambda_i \geq \lambda'_i$ where i is the smallest integer in which $\lambda_i \neq \lambda'_i$.*

3.11.4 Irreducability

Lemma 48. *If $u \in M^\lambda$ and t is of shape λ then $\kappa_t u = c_u e_t$ for some $c_u \in \mathbb{C}$.*

Proof. [1] Let t' be of shape λ as well. If there are two elements a, b in the same row of t' and in the same column of t, then

$$\kappa_t\{t'\} = sign((a,b))\kappa_t(a,b)\{t'\} = -\kappa_t\{t'\},$$

which implies

$$\kappa_t\{t'\} = 0 = 0e_t.$$

Otherwise, by the dominance lemma there is $\pi \in C_T$ so that $\{t'\} = \pi\{t\}$, which implies

$$\kappa_t\{t'\} = sign(\pi)e_t.$$

Hence,

$$\kappa_t u = \sum_{\{t'\}} \alpha_{\{t'\}}\kappa_t\{t'\} = e_t \sum_{t'} \alpha_{\{t'\}} c_{\{t'\}}.$$

□

[1] There seems to be a problem with [Sagan]: "The reasoning of the dominance lemma shows that $\{t'\} = \pi\{t\}$ for some $\pi \in C_t$."

The following theorem of [James] shows that $\mathcal{S}^\lambda$ is irreducible[2] over $\mathbb{C}$.

Theorem 49. *Let U be a submodule of M^λ. Then either $U \subset \mathcal{S}^\lambda$ or ${\mathcal{S}^\lambda}^\perp \subset U$.*

Here the inner product is the unique one so that

$$\langle \{t\}, \{t'\} \rangle = \mathbf{1}_{\{t\}=\{t'\}}.$$

Proof. If there is $u \in U$ so that $c_u \neq 0$ then $e_t \in U$. But $\mathcal{S}^\lambda$ is *cyclic*, that is,

$$\mathcal{S}^\lambda = \mathbb{C}[S_n] \cdot e_t$$

(since every tableaux can obtained from t by reordering).

Otherwise, $\kappa_t u = 0$ for all $u \in U$. The claim is that every $u \in U$ is in the dual of $\mathcal{S}^\lambda$. Indeed, since C_t is a subgroup, the linear map $x \mapsto \kappa_t x$ is unitary, which means that for every e_t,

$$\langle u, e_t \rangle = \langle \kappa_t u, \{t\} \rangle = 0.$$

The set $\{e_t\}$ spans $\mathcal{S}^\lambda$. □

Corollary 50. *The set $\{\mathcal{S}^\lambda\}$ is a complete list of the irreducible representation of S_n.*

Proof. We just need to show that they are pairwise inequivalent (since the dimensions sum up correctly). For this, we use the order on partitions. Assume $\mathcal{S}^\lambda$ is equivalent to $\mathcal{S}^{\lambda'}$. This implies that there is a nonzero $\theta \in Hom(\mathcal{S}^\lambda, M^{\lambda'})$ (a module homomorphism). We claim that then $\lambda \gg \lambda'$. By symmetry, this completes the proof.

There is some t of shape λ so that $\theta(e_t) \neq 0$. Over $\mathbb{C}$, we know that θ can be extended to $Hom(M^\lambda, M^{\lambda'})$ by setting it to be zero on ${\mathcal{S}^\lambda}^\perp$. So,

$$0 \neq \theta(e_t) = \kappa_t \theta(\{t\}) = \sum_i c_i \kappa_t \{t'_i\}.$$

Specifically, there is $\{t'\}$ so that $\kappa_t\{t'\} \neq 0$. By the argument of Lemma 48, there are no a, b are in the same row of t' that are in the same column of t, and the dominance lemma applies. □

Corollary 51.

$$M^{\lambda'} = \bigoplus_{\lambda \gg \lambda'} m_{\lambda',\lambda} \mathcal{S}^\lambda.$$

The numbers $m_{\lambda',\lambda}$ have combinatorial interpretation.

[2] Can actually be proved for any field.

3.12 All groups yield expanders with log many generators

In this part we prove the Alon-Roichman theorem that states that order $\log |G|$ elements of a group yield an expander graph (we have seen this in the exercise for abelian groups). We follow the proof of [Landau and Russel]. The proof is similar to the simple proof for abelian group, except that is uses a matrix version of the Chernoff bound proved by [Ahlswede and Winter].

Let G be a finite group. Recall the definition of $Cay(G,S)$ for a symmetric subset of G. Denote by M the normalised adjacency matrix of S. In the following let $S = \{s_1, s_1^{-1}, \ldots, s_k, s_k^{-1}\}$ be random so that each s_i is uniform and independent. The matrix M is symmetric and so has $n = |G|$ real eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

Let ρ be the regular (matrix) representation of G. So,

$$M = \frac{1}{2k} \sum_{i \in [k]} \rho(s_i) + \rho(s_i^{-1}).$$

We know that $\rho = \oplus_a d_a \pi_a$ where the multiplicity d_a of the irreducible π_a is its dimension.

Theorem 52. *If $k \geq c(b + \log D)$ with $D = \sum_a d_a$ for some constant $c = c(\epsilon) > 0$ then*

$$\Pr[\lambda_2 > \epsilon] \leq 2^{-b}.$$

Observe $D^2 \leq |G| \sum_a d_a^2 \leq |G|^2$.

Proof. Denote

$$M_a = \frac{1}{k} \sum_{i \in [k]} \frac{\pi_a(s_i) + \pi_a(s_i^{-1})}{2}.$$

We know that (up to a basis change) M has the matrices M_a on its "diagonal".

Fix a so that π_a is not trivial. The orthogonality lemma implies

$$\mathbb{E} M_a = 0.$$

We would like to say that M_a is close to the zero matrix w.h.p. so that all of its eigenvalues are small. This is what Ahlswede and Winter proved (which is a matrix analog of Chernoff equality we used in abelian case).

Proposition 53. *Let $A_1, \ldots, A_k$ be i.i.d. random $d \times d$ matrices so that*[3] $\mathbb{E} A_1 = A \geq \mu I$

[3] Here $A \leq B$ means that $B - A$ is positive semidefinite (all eigenvalues are nonnegative).

and a.s. $A_1 \leq I$. Then, for every $\epsilon \in (0, 1/2]$,

$$\Pr\left[\frac{1}{k}\sum_{i\in[k]} A_i \notin [(1-\epsilon)A, (1+\epsilon)A]\right] \leq 2de^{-\frac{\epsilon^2 \mu k}{2}}.$$

Remark 54. *The proof of the proposition uses exponential moments as does the proof of Chernoff. A key property that is used is the Golden-Thompson inequality $Tr(e^{A+B}) \leq Tr(e^A e^B)$ that holds for all Hermitian ($A = A^*$) matrices A, B.*

Use the proposition with

$$A_i = \frac{1}{2}I + \frac{\pi_a(s_i) + \pi_a(s_i^{-1})}{2}.$$

We may use the proposition with $\mu = 1/2$ since π_a is unitary:

$$\Pr\left[\frac{1}{k}\sum_{i\in[k]} A_i \notin [(1-\epsilon)I/2, (1+\epsilon)I/2]\right] \leq 2d_a e^{-\frac{\epsilon^2 k}{2}} \leq \frac{d_a}{D}2^{-b}.$$

When $\frac{1}{k}\sum_{i\in[k]} A_i \in [(1-\epsilon)I/2, (1+\epsilon)I/2]$, we know that M_a is in $[-\epsilon I, \epsilon I]$ and so its eigenvalues are at most ϵ in absolute value.

The union bound over a completes the proof. □

3.13 Sums of squares

We now discuss and prove Hurwitz's theorem. We consider the field $\mathbb{C}$ but the discussion can be carried over an field with $char \neq 2$. We follow notes of [Conrad].

Definition 55. *Define $\sigma(n)$ as the smallest integer k so that*

$$\sum_{i \in [n]} x_i^2 \cdot \sum_{j \in [n]} y_j^2 = \sum_{\ell \in [k]} z_\ell^2,$$

where z_ℓ is bilinear in x, y.

Examples: $1, 2, 4, 8$. Interesting history. Loss of commutativity, associativity.

Upper bounds: obvious $\sigma(n) \leq n^2$, less obvious $\sigma(n) \lesssim n^2 / \log_2(n)$.

Lower bounds: $\sigma(n) \geq n$.

Connection to computational complexity: if $\sigma(n) \geq n^{1+\epsilon}$ then the permanent is hard for non commutative circuits and $ncVNP \neq ncVP$ [Hrubes,Wigderson,Y].

Connections to division algebras.

Theorem 56 (Hurwitz)**.** *$\sigma(n) = n$ iff $n \in \{1, 2, 4, 8\}$.*

Assume $\sigma(n) = n$. The first part of the proof is representing the problem by matrices. Each z_ℓ is bilinear, that is,

$$z_\ell = x A_\ell z$$

for some matrix A_ℓ. The coefficient of $x_i x_{i'} y_j y_{j'}$ is of form (when $i \neq i'$)

$$\mathbf{1}_{i=i', j=j'} = \sum_{\ell \in [n]} A_\ell(i, j) A_\ell(i', j') + A_\ell(i, j') A_\ell(i', j).$$

We have an $n \times n \times n$ tensor $(A_\ell(i, j))$. For fixed i, define the matrix

$$B_i(j, \ell) = A_\ell(i, j).$$

Thus,

$$B_i B_i^T = I$$

and for $i \neq i'$,

$$B_i B_{i'}^T + B_{i'} B_i = 0.$$

We can already prove that

Claim 57. *n is even.*

Proof. $det(B_1 B_2) = (-1)^n det(B_2 B_1)$. □

Consider the group of matrices generated by $B_1, \ldots, B_n$. We may assume that $B_n = I$ by considering if needed $\{B_i B_n^T\}$ instead. This means that

$$B_i^T = -B_i$$

so

$$B_i^2 = -I.$$

It consists of matrices of the form

$$\pm B_1^{a_1} B_2^{a_2} \cdots B_{n-1}^{a_{n-1}}$$

for $a \in \{0,1\}^{n-1}$. Abstractly expressed this group (if it exists) as G defined by $\epsilon, g_1, \ldots, g_{n-1}$ satisfying $g_i g_j = \epsilon g_j g_i$ if $i \neq j$ and $g_i^2 = \epsilon \neq 1$.

Claim 58.

1. *$|G| = 2^n$.*

2. *$[G, G] = \{1, \epsilon\}$.*

3. *If $g \notin Z(G)$ then the conjugacy class of g is $\{g, \epsilon g\}$.*

4. *$Z(G) = \{1, \epsilon, g_1 g_2 \cdots g_n, \epsilon g_1 g_2 \cdots g_n\}$.*

Proof.

1. The claim is that $g_1^{a_1} \ldots g_{n-1}^{a_{n-1}} = 1$ implies $a = 0$. By induction on n. (If $g_1 g_2 \ldots g_{n-1} = 1$ then we get g_{n-1} is a word in the other elements, and substituting it provides a contradiction.)

2. By definition.

3. By definition.

4. If $g = \epsilon^{a_0} g_1^{a_1} \cdots g_{n-1}^{a_{n-1}}$ is in center then for all i,

$$\epsilon^{a_0} \epsilon^{\sum_{j<i} a_j} g_1^{a_1} \cdots g_i^{a_i+1} \cdots g_{n-1}^{a_{n-1}} \epsilon^{n-i+1} \epsilon^{\sum_{j>i} a_j} g_1^{a_1} \cdots g_i^{a_i+1} \cdots g_{n-1}^{a_{n-1}}.$$

Thus, $(J - I)a' = 0 \mod 2$ where $a' \in \{0,1\}^{[n-1]}$ and J is the all-ones matrix. This means that $a \in \{0, \vec{1}\}$.

□

We are ready to use representation theory. We have a n-dimensional representation of G that we investigate.

Start by analysing G. The group $G/[G,G]$ is abelian of size 2^{n-1}, which means that it has 2^{n-1} one-dimensional representations. These can be extended to 2^{n-1} one-dimensional representation of G. The number of irreducible representation is equal to number of conjugacy classes which is $4+(2^n-4)/2 = 2^{n-1}+2$. So, there are exactly 2 irreducible representation of dimension greater than one. The dimension of these must divide the size of G (exercise), and hence are powers of two: $2^{c_1}, 2^{c_2}$. In addition, the sum of squares of dimensions is equal to $|G|$:

$$2^{2c_1} + 2^{2c_2} = 2^n - 2^{n-1} = 2^{n-1}$$

which implies

$$2^{c_1} = 2^{c_2} = 2^{n/2-1}.$$

We started with $B_1, \ldots, B_n$, which is an n-dimensional representation of G. It remains to observe that there are no copies of a one-dimensional representation in G, since commutativity holds on this part. So,

$$2^{n/2-1} \Big| n.$$

If $n = 2^r s$ with s odd then $2^r \le n \le 2r+2$ which implies $r \in \{2,3\}$.