LECTURE NOTES FOR THE 22ND MCGILL INVITATIONAL WORKSHOP ON COMPUTATIONAL COMPLEXITY

Bellairs Institute
Holetown, Barbados

Lecturers:
Ben Green
Avi Wigderson

Contents

Foreword

These notes reflect a series of lectures given by Ben Green and Avi Wigderson at the 22nd McGill Invitational Workshop on Computational Complexity. The workshop was held at the Bellairs Research Institute in Holetown, Barbados in February, 2010.

The two lecturers alternated presentations, coving related, but disjoint material. Odd numbered lectures, given by Ben Green, focused on topics in additive combinatorics; even numbered lectures, given by Avi Wigderson, focused on applications of the theory of representations of groups to theoretical computer science.

Chapter 1

Additive Combinatorics

A series of 5 lectures by Ben Green.

LECTURE 1

Lecturer: Ben Green *Scribe: Alexander Russell*

1.1 Introduction: Approximate structure

We will study approximate algebraic structure: approximate groups, homomorphisms, polynomials, etc. We can roughly divide our study into three tasks: finding good definitions for such objects, determining what can be said about them, and exploring their applications. The theory can be divided into two regimes: the 99% regime (studying structures that "very close" to their genuine counterparts) and the 1% regime (studying structures that may only weakly resemble their genuine counterparts).

Notation Let A and B be two subsets of a universal group G. We write

$$A \cdot B = \{ab \ : \ a \in A, b \in B\},$$
$$A^{-1} = \{a^{-1} \ : \ a \in A\}.$$

(In abelian groups, we express such sets additively: $A + B$ and $-A$.)

Theorem 1.1.1 (Freiman)**.** *Suppose $A \subset \mathbb{F}_2^n$, the vector space of dimension n over $\mathbb{F}_2$. If $|A + A| \leq \frac{3}{2}|A|$ then A is a subspace.*

Remark 1.1.2. The condition that $|A + A| < \frac{3}{2}|A|$ is a notion of approximate subgroup.

Proof. Suppose that $x, y \in A$. The cosets $x + A$ and $y + A$ are both subsets of $A + A$. As $|A + A| < \frac{3}{2}|A|$ is follows that $|x + A \cap y + A| > |A|/2$. Hence there are more than $|A|/2$ pairs (a_1, a_2) for which $x + a_1 = y + a_2$; in this case $x + y = a_1 + a_2$. Now, if x', y' are two further elements we likewise have more than $|A|/2$ pairs $a_1', a_2' \in A$ for which $x' + y' = a_1' + a_2'$. It follows that there is a pair (a_1, a_2), for which $x + y = a_1 + a_2$, and a pair (a_1', a_2'), for which $x' + y' = a_1' + a_2'$, so that $a_2 = a_1'$. We conclude that

$$(x + y) + (x' + y') = (a_1 + a_2) + (a_1' + a_2') = a_1 + a_2' \in A + A,$$

as desired. □

Remark 1.1.3.

1. This same argument implies that if $|A + A| \leq (1 + \epsilon)|A|$ with $\epsilon < 1/2$ then there is a subgroup H (in fact equal to $A + A$) and an element $x \in G$ so that
$$\left|A \ \triangle \ (H + x)\right| = O(\epsilon) \cdot \min(|A|, |H|).$$
2. The theorem is true for nonabelian groups as well.
3. A more complicated argument can extend the values of ϵ for which the implication is true to $\phi - 1 \approx .618$ (where ϕ is the golden ratio).

It is unknown if the statement is true for all $\epsilon < 2$. In particular, the following problem is open.

Question 1.1.4. Is there a function $f : [0,1] \to \mathbb{R}_+$ so that for any set A of a group G for which $|A+A| \leq (2-\delta)|A|$ there is a subgroup H and an element $x \in G$ so that

$$|A \cap Hx| > f(\delta)\max(|A|,|H|)\,?$$

Remark 1.1.5. The range for which this applies is no larger than $[0,2)$. If A is an arithmetic progression in $\mathbb{Z}$ then $|A+A| < 2|A|$. However, subgroups in $\mathbb{Z}$ are either infinite or have cardinality 1.
The question is known for abelian groups; this is Kneser's theorem.

1.2 The Blum-Luby-Rubinfeld test; approximate homomorphisms

Theorem 1.2.1. *Let $\phi : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function. Suppose that*

$$\mathbb{P}_{x,y}[\phi(x+y) = \phi(x)+\phi(y)] \geq 1-\epsilon\,.$$

Then there is a homomorphism $\phi' : \mathbb{F}_2^n \to \mathbb{F}_2$ so that

$$\mathbb{P}_x[\phi(x) = \phi'(x)] \geq 1 - O(\epsilon^{1/2})\,.$$

With a more careful analysis it is possible to achieve the bound $1-O(\epsilon)$ above.

Proof sketch. Define

$$\tilde{\phi}(h) = \mathrm{maj}\{\phi(x+h)-\phi(x) \;:\; x \in \mathbb{F}_2^n\}\,,$$

where maj S denotes the element x appearing with highest multiplicity in the multiset S.

Step 1. There is always an element z that forms the overwhelming majority of $\{\phi(x+h)-\phi(x) \;:\; x \in \mathbb{F}_2^n\}$.

Proof. Observe that

$$\mathbb{P}_{x,y}[\phi(x+h)+\phi(y) = \phi(x+y+h)] = \mathbb{P}_{x,y}[\phi(y+h)+\phi(x) = \phi(x+h+y)] \geq 1-\epsilon$$

and hence both events occur with probability at least $1-2\epsilon$. In this case, subtracting the two equations yields

$$\phi(x+h)-\phi(x) = \phi(y+h)-\phi(y)\,,$$

as desired. Note, however, that if the most likely element of $\{\phi(x+h)-\phi(x) \;:\; x \in \mathbb{F}_2^n\}$ appears with probability $1-\delta$ we must have

$$\mathbb{P}_{x,y}[\phi(x+h)-\phi(x) = \phi(y+h)-\phi(y)] \leq (1-\delta)^2+\delta^2$$

and thus $1-2\delta+2\delta^2 \geq 1-2\epsilon$. Hence the most likely element appears with probability $1-2\epsilon$. (As $\epsilon \to 0$, this yields $\delta = (1+o(1))\epsilon$.)

□

Step 2. $\tilde{\phi}$ is linear: $\tilde{\phi}(h+k) = \tilde{\phi}(h)+\tilde{\phi}(k)$.

Proof. For most triples x, y, z

$$\tilde{\phi}(h+k) = \phi(x+h+k) - \phi(x)\,;$$
$$\tilde{\phi}(h) = \phi(y+h) - \phi(y)\,;$$
$$\tilde{\phi}(k) = \phi(z+k) - \phi(z)\,.$$

In particular, applying the result of step 1, this occurs with probability at least $1 - 6\epsilon$. In fact, the same can be said if we merely select x at random, and assign $z = x$ and $y = x + k$. In this case, we conclude

$$\tilde{\phi}(h+k) = \tilde{\phi}(h) + \tilde{\phi}(k)\,.$$

□

N.b. If ϕ is an approximate homomorphism then $(\phi(x), x)$ is an approximate group.

Step 3. For most h, $\phi(h) = \tilde{\phi}(h)$.

Proof. For most x, h

$$\tilde{\phi}(h) = \phi(x+h) - \phi(x) \quad \text{and} \quad \phi(h) = \phi(x+h) - \phi(x).$$

(Note that the first event occurs with probability at least $1 - 2\epsilon$ even conditioned on a particular choice of x; the second event occurs with probability $1 - \epsilon$. Thus both occur with probability $1 - 3\epsilon$.) Now apply Markov's inequality. □

□

Remark 1.2.2. Let $A = \{(x, \phi(x)) \ : \ x \in \mathbb{F}_2^n\}$ be the graph of ϕ. Then the BLR condition is equivalent to

$$\mathbb{P}[a_1 + a_2 \in A \ : \ a_1, a_2 \in A] \geq 1 - \epsilon\,,$$

an alternate notion of approximate subgroup.

1.3 Multiplicative energy

Definition 1.3.1. Let A be a subset of a group G. We define the *multiplicative energy* of A to be the quantity

$$E(A) = |\{(a_1, a_2, b_1, b_2) \ : \ a_1 a_2^{-1} = b_1 b_2^{-1}, a_i \in A, b_i \in A\}|\,.$$

Remark 1.3.2. Observe that $E(A) \leq |A|^3$ and that if $U : A \to \mathbb{R}$ is the uniform distribution on A, $E(A)$ is proportional to the collision probability of $U * U$.

The condition that $E(A) < |A|^3$ is another notion of approximate subgroup.

Theorem 1.3.3 (Fournier [10])**.** *If $E(A) \geq (1-\epsilon)|A|^3$ then there is a subgroup H and an element $x \in G$ such that*

$$|A \cap Hx| \geq (1 - O(\epsilon^{1/4})) \max(|A|, |H|)\,.$$

Remark 1.3.4. By bootstrapping, you can achieve $O(\epsilon)$ error (rather than $O(\epsilon^{1/4})$). This implies the BLR result, even for general groups.

Fournier was interesting in constructing tight cases for Young's inequality:

Theorem 1.3.5. *If* $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ *then* $\|f * g\|_r \leq \|f\|_p \|g\|_q$.

A particularly insteresting case in additive combinatorics occurs when $r = 1/2$ and $p = q = 4/3$. As a consequence of the theorem above, Fournier established:

Theorem 1.3.6 (Fournier)**.** *Let G be a group with no compact open subgroups. Then there exists a constant $c < 1$ so that*

$$\|f * g\|_2 \leq c\|f\|_{\frac{4}{3}} \|g\|_{\frac{4}{3}} .$$

Remark 1.3.7. This constant c over $\mathbb{R}$ is $c = \sqrt[4]{16/27}$ and is attained for Gaussian distributions $f = g = e^{-x^2/2}$.

Definition 1.3.8. Let A be a finite set and $\delta > 0$ a parameter. Then

$$\mathrm{Sym}_{1-\delta} = \{x \ : \ |A \cap Ax| \geq (1-\delta)|A|\} .$$

When $x \in \mathrm{Sym}_{1-\delta}(A)$, the element x has $(1-\delta)|A|$ representations of the form $a_1^{-1}a_2$.

Remark 1.3.9.

1. $\mathrm{Sym}_{1-\delta}(A)$ is symmetric: $x \in \mathrm{Sym}_{1-\delta}(A) \Leftrightarrow x^{-1} \in \mathrm{Sym}_{1-\delta}(A)$.
2. Sym possesses "weak additive closure": $\mathrm{Sym}_{1-\delta_1}(A) \cdot \mathrm{Sym}_{1-\delta_2}(A) \subset \mathrm{Sym}_{1-\delta_1-\delta_2}(A)$.

   Proof. Let $x \in \mathrm{Sym}_{1-\delta_1}(A)$ and $y \in \mathrm{Sym}_{1-\delta_2}(A)$. Then x has $(1-\delta_1)|A|$ representations of the form $a_1 a_2^{-1}$; likewise, y has $(1-\delta_2)|A|$ representations of the form $b_1 b_2^{-1}$. In at least $(1-\delta_1-\delta_2)$ of these pairs of representations, we have $a_2 = b_1$ and thus $xy = a_1 a_2^{-1} b_1 b_2^{-1} = a_1 b_2^{-1}$. □

Claim 1.3.10. *Suppose that $E(A) \geq (1-\epsilon)|A|^3$. Then $|\mathrm{Sym}_{1-\delta}(A)| \geq (1-\epsilon/\delta)|A|$.*

N.b. $|\mathrm{Sym}_{1-\delta}(A)|$ could exceed $|A|$, but is always $(1 + O(\delta))|A|$.

Proof. Write $r(x) = |\{(a_1, a_2) \in A \times A \ : \ a_1 a_2^{-1} = x\}|$. Then $\mathrm{Sym}_{1-\delta}(A) = \{x \ : \ r(x) \geq (1-\delta)|A|\}$ and $E(A) = \sum_x r(x)^2$. Define $S = \mathrm{Sym}_{1-\delta}(A) = \{x \ : \ r(x) \geq (1-\delta)|A|\}$ and σ so that $\sum_{x \in S} r(x) = \sigma|A|^2$. Then

$$\begin{aligned} E(A) &\leq |A| \cdot \sum_{x \in S} r(x) + (1-\delta)|A| \sum_{x \notin S} r(x) \\ &= \sigma|A|^3 + (1-\delta)(1-\sigma)|A|^3 , \end{aligned}$$

since $\sum_x r(x) = |A|^2$.

As $E(A) \geq (1-\epsilon)|A|^3$ by hypothesis, we conclude that

$$(1-\epsilon) \leq \sigma + (1-\delta)(1-\sigma) \quad \Rightarrow \quad -\epsilon \leq -\delta + \delta\sigma \quad \Rightarrow \quad \sigma \geq 1 - \epsilon/\delta .$$

Thus $|S| \geq \sum_{x \in S} r(x)/|A| = \sigma|A| \geq (1-\epsilon/\delta)|A|$. □

Proof of Theorem 1.3.3. Suppose ϵ is small; in this case, $E(A) \approx |A|^3$ and, by the preceding argument,

$$|\operatorname{Sym}_{1-\eta}(A)| \approx |\operatorname{Sym}_{1-2\eta}(A)| \approx |\operatorname{Sym}_{1-4\eta}(A)| \approx |\operatorname{Sym}_{1-5\eta}(A)| \approx |A|\,.$$

for $\eta = 10^{-4}\epsilon^{1/2}$. With such η, the symbol $\approx$ above denotes equality up to a multiplicative factor of $1 - O(\epsilon^{1/2})$.

We shall establish that $\operatorname{Sym}_{1-2\delta}(A) = \operatorname{Sym}_{1-4\delta}(A)$ so that, by "weak additive closure,"

$$\operatorname{Sym}_{1-2\delta}(A) \cdot \operatorname{Sym}_{1-2\delta}(A) \subset \operatorname{Sym}_{1-4\delta}(A) = \operatorname{Sym}_{1-2\delta}(A)$$

and $\operatorname{Sym}_{1-2\delta}(A)$ must be a subgroup. It follows that A has large intersection with a left coset of $\operatorname{Sym}_{1-2\delta}(A)$

To establish that $\operatorname{Sym}_{1-2\delta}(A) = \operatorname{Sym}_{1-4\delta}(A)$, consider an element $x \in \operatorname{Sym}_{1-4\eta}(A)$. The sets $\operatorname{Sym}_{1-\eta}(A)$ and $\operatorname{Sym}_{1-4\eta}(A)$ both lie in $\operatorname{Sym}_{1-5\eta}(A)$ but

$$|\operatorname{Sym}_{1-\eta}(A)| \approx |x \operatorname{Sym}_{1-\eta}(A)| \approx |\operatorname{Sym}_{1-5\eta}(A)|$$

and hence $x \operatorname{Sym}_{1-\eta}(A)$ and $\operatorname{Sym}_{1-\eta}(A)$ have a nonempty intersection. It follows that $x \in \operatorname{Sym}_{1-\eta}(A) \cdot \operatorname{Sym}_{1-\eta}(A) \subset \operatorname{Sym}_{1-2\eta}(A)$. □

1.3.1 Cooperman's algorithm

Cooperman's algorithm is a procedure for generating a nearly random element of a black box group (given generators for the group).

Example 1.3.11. Let S_1 and S_2 be two elements of $\mathsf{GL}(\mathbb{F}_q)$. Note multiplication in $\mathsf{GL}(\mathbb{F}_q)$ can be carried out efficiently. Cooperman's algorithm efficiently generates a nearly uniform sample from $\langle S_1, S_2 \rangle$, the subgroup generated by S_1 and S_2.

Algorithm 1.3.12. Let $s_1, \ldots, s_k$ be a sequence of generators for the group G. Define the sequence $g_1, g_2, \ldots$ so that, for $i = 1, \ldots, k$, $g_i = s_i$ and, for $i > k$, g_i is a random element of the "cube"

$$\Sigma(g_1, \ldots, g_{n-1}) = \{ g_1^{\epsilon_1} \cdots g_{n-1}^{\epsilon_{n-1}} \ : \ \epsilon_i \in \{0, 1\} \}\,,$$

determined by selecting each ϵ_i independently and uniformly at random from $\{0, 1\}$.

Theorem 1.3.13 (Cooperman)**.** *For $s = 2k + C_1 \log |G|$, the distribution of g_s is within* 0.01 *of uniform.*

LECTURE 3

Lecturer: Ben Green *Scribe: Ricard Gavaldà*

1.4 Correctness of Cooperman's algorithm

In this lecture we will show that Cooperman's algorithm performs as claimed, that is, given a set of generators of a black-box group G, it generates a random element of $|G|$ in $O(\log |G|)$ iterations.

Recall that Cooperman's algorithm is given as input a set of generators $s_1, \ldots, s_k$ for an otherwise unknown black-box group G. It then generates a sequence of $g_1, g_2, \ldots$ of elements of G, as follows:

- $g_1, \ldots g_{2k}$ are simply $s_1, \ldots, s_k$ and their inverses, taken in some order
- for $i \geq 2k$, g_{i+1} is a random element taken from the Boolean cube of $g_1, \ldots g_i$, that is,

$$\sum(g_1, \ldots, g_i) = \{\ g_1^{\epsilon_1} \cdot \cdots \cdot g_i^{\epsilon_i}\ :\ \epsilon_i \in \{0,1\}\ \}.$$

We devote most of this lecture to proving:

Theorem 1.4.1 ([7])**.** *For some constant c, if $t \geq 2k + c \log |G|$ then the distribution of g_t is 0.01 away from the uniform in ℓ_1.*

We will start by discussing probabilities, and probability measures over groups.

Definition 1.4.2. A function $\nu : G \to \mathbb{R}_{\geq 0}$ is called a *probability measure* if $\mathbb{E}_x\, \nu(x) = 1$. Here $\mathbb{E}_x$ denotes the expected value, i.e.,

$$\mathbb{E}_x\, f(x) = \mathbb{E}_{x \in G}\, f(x) = \frac{1}{|G|} \sum_{x \in G} f(x)\,.$$

See that, by this definition, we are taking the uniform distribution over G from now on for expected values. Write $\mathcal{M}(G)$ for the space of probability measures over G.

A particular probability measure that we will use often is that which is uniform on a support set $A \subseteq G$:

$$M_A(x) = \begin{cases} |G|/|A| & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Particular cases are $\delta_g(x) = M_{\{g\}}$ for any $g \in G$, and M_G, the uniform distribution on G.

We will often want to convolute probability measures. Let $\nu_1, \nu_2 \in \mathcal{M}(G)$. The convolution of ν_1 and ν_2, written $\nu_1 \star \nu_2$, is defined by

$$(\nu_1 \star \nu_2)(x) = \mathbb{E}_y\ \nu_1(y)\nu_2(y^{-1}x)\,.$$

It is easy to check that $\nu_1 \star \nu_2 \in \mathcal{M}(G)$. Note that if $\nu_1 = M_A$ and $\nu_2 = M_B$, $\nu_1 \star \nu_2$ is supported (not necessarily uniformly) on $AB = \{ab\ :\ a \in A, b \in B\}$.

A random walk is really the composition of a probability measure with itself. Take an initial set $S = \{s_1, \ldots, s_k\}$, and then

$$\nu(x) = \frac{1}{2k}\left(\delta_{s_1} + \delta_{s_1^{-1}} + \cdots + \delta_{s_k} + \delta_{s_k^{-1}}\right).$$

Then

$$\underbrace{\nu \star \nu \star \cdots \star \nu}_{t} \star \delta_{\mathrm{id}}$$

has the distribution of the random walk starting at the group identity and, for t times, choosing a random $s_i^{\pm 1}$ and multiplying. Or, in other words, it is a length-t random walk on the Cayley graph $\mathrm{Cay}(G; S)$.

Two more definitions:

- We write $\tilde{\nu}(x) = \nu(x^{-1})$.
- For $p \geq 1$, we write $\|\nu\|_p = (\mathbb{E}_x\ \nu(x)^p)^{1/p}$ for the ℓ_p-norm of ν.

Note that $\|\nu\|_1 = 1$ for every probability measure ν, and that we often write $\|\nu\|$ for $\|\nu\|_2$, since we will use it most often. Observe also that $\|\delta_g\| = \sqrt{|G|}$ and $\|M_G\| = 1$. In general, the more concentrated ν is, the larger $\|\nu\|_2$ becomes, and conversely, when $\|\nu\|$ is small, ν is close to the uniform. Precisely, if $\|\nu\| < 1 + \epsilon$ then

$$\|\nu - M_G\|^2 = \langle \nu - M_G, \nu - M_G \rangle = \|\nu\|^2 - 1 \leq 3\epsilon$$

hence $\|\nu - M_G\| = O(\epsilon^{1/2})$.

The following lemma says that, at any given step of the random walk in Cooperman's algorithm, either we are already close to the uniform, or we get significantly closer to uniform at this step:

Lemma 1.4.3 (ℓ_2-flattening Lemma)**.** *Let $g_1, \ldots, g_i \in G$ generate G. Let g_{i+1} be sampled at random from the cube* [1]

$$M_{\sum(g_1,\ldots,g_i)} (= M_{\{1,g_1\}} \star \cdots \star M_{\{1,g_i\}}).$$

Then

$$\|M_{\sum(g_1,\ldots,g_{i+1})}\| \leq \|M_{\sum(g_1,\ldots,g_i)}\|$$

and either

(i) $\|M_{\sum(g_1,\ldots,g_i)}\| \leq 1.001$ *or*

(ii) with probability $\geq c_1$, $\|M_{\sum(g_1,\ldots,g_{i+1})}\| \leq (1 - c_2) \cdot \|M_{\sum(g_1,\ldots,g_i)}\|$,

for constants c_1 and c_2.

Before we prove the lemma, let us deduce Cooperman's theorem from it. The following is clear:

Lemma 1.4.4. *Let c_1 be as in the flattening lemma, part (ii), c_3 some constant, and c_4 sufficiently larger than c_3. Run Cooperman's algorithm for $t = 2k + c_4 \log |G|$ steps. The probability that (i) in the flattening lemma is never reached and that (ii) happens less than $c_3 \log |G|$ times is less than*

$$\sum_{j \leq c_3 \log |G|} \binom{c_4 \log |G|}{j} (c_1)^j (1 - c_1)^{c_4 \log |G| - j}.$$

[1] Observe that the subscript $\sum(g_1, \ldots, g_i)$ of M should not be taken as a set, but a multiset. That is, each g_i is given mass according to its multiplicity.

By standard tail estimates for the binomial distribution (i.e., Chernoff) the latter quantity is less than $|G|^{-10}$ if c_4 is large enough with respect to c_1 and c_3. Therefore, even assuming that we started from a most concentrated distribution (of norm $\sqrt{|G|}$), with probability greater than $1 - |G|^{-10}$ we have

$$\|M_{\sum(g_1,\dots,g_t)}\| \leq (1-c_2)^{c_3 \log |G|}\sqrt{|G|} \leq 1.001$$

if c_3 is chosen somewhat large w.r.t. c_2. With this probability, Cooperman's algorithm outputs some $g \in G$ from a measure with ℓ_2-norm less than 1.001. The rare events (occurring with probability less than $|G|^{-10}$) in which this does not occur contribute in the worst case distributions whose ℓ_2-norm can be at most $\sqrt{|G|}$, so their effect on the expected value is negligible. Overall, then Cooperman's algorithm produces a probability distribution ν_t with $\|\nu_t\|_2 \leq 1.002$.

We will deduce the ℓ_2-flattening lemma from another lemma, which says that we get some ℓ_2-flattening at one step from ν if ν is not totally concentrated on any one coset.

Lemma 1.4.5. *Let $\nu \in \mathcal{M}(G)$ be a probability measure with the property that $\nu(Hx) < 0.99$ for every coset Hx of a proper subgroup $H \leq G$. Let g be sampled at random from ν. Then either (i) $\|\nu\| < 1.001$ or (ii) with probability at least c, $\|\nu \star M_{\{\mathrm{id},g\}}\| < (1-c)\|\nu\|$.*

Again, before proving this lemma, let us first deduce the ℓ_2-flattening lemma from it. Note that $M_{\sum(g_1,\dots,g_{i+1})} = M_{\sum(g_1,\dots,g_i)} \star M_{\mathrm{id},g_{i+1}}$. So we only need to show the nonconcentration property, i.e. that $M_{\sum(g_1,\dots,g_{i+1})}(Hx) < 0.99$; we will in fact show that it is at most $1/2$. The argument is due to Babai and Erdős [2].

Since $g_1, \dots, g_i$ generate G and H is a proper subgroup, there is a minimal j such that $g_j \notin H$. Consider an element $w = g_1^{\epsilon_1} \dots g_i^{\epsilon^i}$ of the cube $\sum(g_1 \dots g_i)$, and split it as

$$w = w_1\, g_j^{\epsilon_j}\, w_2, \quad \text{with } w_1 = g_1^{\epsilon_1} \dots g_{j-1}^{\epsilon^{j-1}}, w_2 = g_{j+1}^{\epsilon_{j+1}} \dots g_i^{\epsilon_i}.$$

Because $w_1 \in H$, we have that $w \in Hx$ if and only if $g_j^{\epsilon_j} w_2 \in Hx$. If $w_2 \notin Hx$, then $g_j^{\epsilon_j} w_2 \notin Hx$ for $\epsilon_j = 0$. If $w_2 \in Hx$, then $g_j w_2 \notin Hx$, otherwise we have $g_j \in H$ which is not true, and therefore $g_j^{\epsilon_j} w_2 \notin Hx$ for $\epsilon_j = 1$. Since ϵ_j is a random bit, we have $M_{\sum(g_1,\dots,g_{i+1})}(Hx) \leq 1/2$ as claimed.

Let us prove Lemma 1.4.4. Recall that $E(A)$ denotes the multiplicative energy of A,

$$E(A) = \#\{(a_1,a_2,a_3,a_4) \ : \ a_1 a_2^{-1} = a_3 a_4^{-1}\}$$

and that Fournier's theorem asserts:

Theorem 1.4.6 ([10])**.** *Let $A \in G$ be a set with $E(A) \geq (1-\epsilon)|A|^3$. Then there is a subgroup H and an element x with $|A \bigtriangleup Hx| \leq O(\epsilon^{1/2})\ \min(|A|,|H|)$.*

(We could in fact replace $\min(|A|,|H|)$ with, say, $|A|$ because we are precisely saying that $|A|$ and $|H|$ are very close.)

Since we are dealing with probability measures, let us state a measure version of this theorem:

Theorem 1.4.7. *Let $\nu \in \mathcal{M}(G)$ be such that $\|\nu \star \tilde{\nu}\| \geq (1-\epsilon)\|\nu\|$. Then there is a subgroup H and an element x with $\|\nu - M_{Hx}\| \leq O(\epsilon^{1/2})\,|A|$.*

Observe for clarity that if $\nu = M_A$ then $E(A)/|A|^3 = \|\nu \star \tilde{\nu}\|^2/\|\nu\|^2$; the proof is left as an exercise. So claiming that the right-hand-side of this equality is large is the same as claiming that A has large multiplicative energy.

We will omit the proof of this theorem, but mention two possible approaches to the proof. Dixon's approach was essentially to redo Fournier's proof with measures. Fournier's way was to reduce the measure version to the set version. The idea is that as there are few repeated products, there are very few multiplicities, and so ν is close to M_A for some set A.

Let us now finally prove the lemma. We will show that if neither (i) nor (ii) hold then

$$\|\nu \star \tilde{\nu}\| \geq (1-\epsilon)\|\nu\|$$

which implies (by the measure version of Fournier's theorem) that

$$\|\nu - M_{Hx}\| < 0.001\|\nu\|$$

from which we can deduce (exercise) that $\nu(Hx) \geq 0.99$. The following holds in general:

$$\|\nu - \nu \star \tilde{\nu}\| = \|\mathbb{E}[\tilde{\nu}(g)(\nu - \nu \star \delta_g)\| \leq \mathbb{E}_g\|\nu - \nu \star \delta_g\|. \tag{1}$$

It is easy to check that $\nu \star (\delta_{\mathrm{id}} + \delta_g)(= 2\nu * M_{\mathrm{id},g})$ and $\nu \star (\delta_{\mathrm{id}} - \delta_g)(= \nu - \nu * \delta_g)$ are orthogonal (have zero inner product) and their sum is 2ν. Then by Pythagoras

$$\|\nu - \nu \star \delta_g\|^2 = 4\|\nu\|^2 - 4\|\nu \star M_{\mathrm{id},g}\|^2.$$

Plugging this into (1) and assuming (ii) and (i) do not hold we obtain

$$\|\nu - \nu \star \tilde{\nu}\| \leq 2\sqrt{1-(1-c)^2}\,\|\nu\| < 0.001\|\nu\|$$

if c is small enough. This concludes all pending proofs, and Cooperman's algorithm is correct.

We will move now to the world of "1% additive combinatorics." From now on, $K \geq 2$ will be some fixed parameter, with $1/K$ measuring the degree to which approximate objects resemble exact objects. We will then be concerned with questions such as:

- If A is a finite set, what can we say if $|A \cdot A| \leq K|A|$? This resembles Freiman's theorem in the 99% world, but is much harder and unsolved in general. We will ask it on $\mathbb{F}_2^n$, $\mathbb{Z}$, and $\mathsf{GL}_n(\mathbb{C})$.
- Suppose that $\varphi : G \to H$ satisfies

  $$\mathbb{P}[\varphi(xy) = \varphi(x)\varphi(y)] \geq 1/2.$$

  What can we say about φ? Is it close to a homomorphism?

The plan for the rest of the lectures is as follows: In Lecture 5, we will deal with the basic theory of these objects. We will cover subset estimates and the Balog-Szemerédi-Gowers theorem, all over $\mathbb{F}_2^n$. In Lecture 7 we will move to other groups, and cover analogs of Freiman's theorem. In Lecture 9 we will cover applications, and in particular, an application to the construction of expanders.

Here Avi observes that what Ben calls "the 99% world" and "the 1% world" are typically called "unique decoding" and "list decoding" in computer science terms. This is because when an approximate object is 99%-close to some exact object, it in fact is close to a *unique* one, while a 1%-approximate object may be correlated to several exact ones.

LECTURE 5

Lecturer: Ben Green *Scribe: Yara Elias*

1.5 Notation

Definition 1.5.1. Let A and B be two sets of an abelian group G. Define

$$\begin{aligned} A+B &= \{a+b \ : \ a\in A, b\in B\}, \\ A-B &= \{a-b \ : \ a\in A, b\in B\}, \\ A\cdot B &= \{a.b \ : \ a\in A, b\in B\}, \\ A\times B &= \{(a,b) \ : \ a\in A, b\in B\}, \\ kA &= \{a_1+a_2+\cdots+a_k \ : \ a_1,a_2,\dots,a_k\in A\}. \end{aligned}$$

Take A $\subseteq F_2^n$. The aim is to see when $A+A$ has structure ($|A+A|$ is small) and what it implies. We will see first that bounds on $|2A|$ imply bounds on $|kA|$, and that structure in $A+A$ implies (is equivalent) that A lies in some "small" subgroup using a result of Ruzsa. Then, we will look at a result of Balog-Szemerédi-Gowers stating that "large" additive energy in A forces the existence of a "large" subset A' in A such that $2A'$ has structure.

1.6 Ruzsa's theorem

Lemma 1.6.1. *Ruzsa's triangle inequality. Suppose $U,V,W\subseteq F_2^n$. Then*

$$|U|\cdot|V-W|\leq|U-V|\cdot|U-W|.$$

Proof. Define an injection $\psi \ : \ Ux(V-W)\to(U-V)x(U-W)$. That obviously suffices. To do this, fix for each d in $V-W$ a choice of $v(d)$ in V and $w(d)$ in W with $d=v(d)-w(d)$. Define $\psi(u,d)=(u-v(d),u-w(d))$. Observe that if the right hand side is known, we can recover u and d: Suppose $\psi(u',d')=\psi(u,d)$. By substraction, $d'=(u'-w(d'))-(u'-v(d'))=(u-w(d))-(u-v(d))=d$. $u=(u-v(d))+v(d)=(u-v(d'))+v(d')=u'$. □

Remark 1.6.2. This is called the triangle inequality since if we define

$$d_{\text{Ruzsa}}(A,B)=\log\frac{|A-B|}{|A|^{\frac{1}{2}}|B|^{\frac{1}{2}}},$$

then the inequality is equivalent to

$$d_{\text{Ruzsa}}(A,C)\leq d_{\text{Ruzsa}}(A,B)+d_{\text{Ruzsa}}(B,C).$$

Note however that d_{Ruzsa} is not a distance since $d_{\text{Ruzsa}}(A,A)$ is not always 0 and $d_{\text{Ruzsa}}(A,B)$ may be 0 even when $A\neq B$.

Corollary 1.6.3. *Let $k\geq 3$. Suppose $A\subseteq F_2^n$ and $|3A|\leq K|A|$. Then, for any $k\geq 3$, $|kA|\leq K^{k-2}|A|$.*

Proof. Induction using Ruzsa's inequality. For $k = 3$, the inequality is the same as the hypothesis. In characteristic 2, Ruzsa's inequality can be written as: $|U| \cdot |V + W| \leq |U + V| \cdot |U + W|$. Take $W = 2A$; $V = (k-1)A$ and $U = A$. Then, $|A||(k+1)A| \leq |kA||3A| \leq (K^{k-2}|A|)(K|A|)$. Thus $|(k+1)A| \leq K^{k+1-2}|A|$. □

Proposition 1.6.4. *Suppose* $|2A| = |A + A| \leq K|A|$. *Then* $|4A| \leq CK^c|A|$.

Proof. Find a large set S such that $2A + S$ is small then, applying Ruzsa with $U = S$ and $V = W = 2A$, we get $|4A| \geq |2A + S|^2/|S|$. Define

$$r(x) = |\{(a_1, a_2) \in AxA : a_1 + a_2 = x\}|.$$

Take $S = \{x : r(x) \geq |A|/2k\}$. Claim: $|S| \geq |A|/(2k)$. Indeed,

$$\sum_{x \notin S} r(x) \leq \frac{|A|}{2k} k|A| = \frac{|A|^2}{2}.$$

Hence, $\sum_{x \in S} r(x) \geq |A|^2/2$ since $\sum_x r(x) = |A|^2$. So $|S| \geq |A|/2$ since $r(x) \leq |A|$. Then every element $a_1 + a_2 + s$, a_1, a_2 in A, s in S can be written in at least $|A|/2k$ ways as $a_1 + a_2 + a_1' + a_2' = (a_1 + a_1') + (a_2 + a_2')$ that is to say in $\geq |A|/2k$ ways as sum of 2 elements of $A + A$. Note that the $a_1 + a_1'$ are distinct since we have at least as many distinct a_1'; this is also true for the $a_2 + a_2'$. Hence,

$$\frac{|2A + S||A|}{2k} \leq |A + A|^2, \qquad |2A + S| \leq 2k^3|A|$$

(using the hypothesis), and $|4A| \leq 8k^6|A|$ (using Ruzsa as first indicated). □

Remark 1.6.5. This is not true in non abelian case. Let $A = H \cup \{x\}$ (non abelian). Then $A \cdot A = H \cup xH \cup Hx \cup \{x^2\}$. $|A \cdot A| \leq 3|A| - 2$. Thus $A \cdot A \cdot A$ contains $H \cdot H$ and there is no reason why this should be small.

Fact 1.6.6. If A $\subseteq$ abelian group such that $|A + A| \leq k|A|$, then $|mA - \ell A| \leq k^{m+\ell}|A|$.

Theorem 1.6.7 (Ruzsa)**.** *Suppose* $A \subseteq F_2^n$ *is a finite set with* $|A + A| \leq k|A|$. *Then there is a subspace* $H \leq F_2^\infty$ *containing* A *with* $|H| \leq F(k)|A|$. *We'll get* $F(k) = \exp(k^c)$.

Remark 1.6.8. If, by contrast, A is a subset of some subgroup H wih $|A| \geq \delta|H|$ then $|A+A| \leq |H| \leq |A|/\delta$. Thus, in some sense, Ruzsa's theorem gives a complete classification of sets with small doubling in $\leq F_2^\infty$.

Proof. Let X be a subset of $3A$ for which the translates $A + x$, $x \in X$ are all disjoint and which is maximal with respect to this property. Observe that the disjoint union $\bigcup_{x \in X}(A + x) \subseteq 4A$. Hence, $|X||A| \leq 8k^6|A|$ implies $|X| \leq 8k^6$. Now suppose $y \in 3A$. By maximality, $(A + y) \cap (A + x) \neq \emptyset$ for some $x \in X$. Hence $y \in 2A + X$ (characteristic 2). That is $3A \subseteq 2A + X$. So, $4A \subseteq_{\text{adding } A} 3A + X \subseteq 2A + 2X$ and $5A \subseteq 2A + 3X$. We conclude that $\langle A \rangle$, the subgroup of F_2^∞ generated by A, is contained in $2A + \langle X \rangle$. This implies that $|\langle A \rangle| \leq |2A| \cdot |\langle X \rangle| \leq k|A|8k^6$. □

1.7 Balog-Szemerédi-Gowers

Recall the definition of additive energy.

Definition 1.7.1. $\mathrm{A} \leq F_2^\infty$. $E(A) = |\{(a_1, a_2, a_3, a_4) \in A \times A \times A \times A : a_1 + a_2 = a_3 + a_4\}|$.

Elementary observations:

Fact 1.7.2. $E(A) \leq |A|^3$ since a_1, a_2, a_3 fixed imply a_4.

Fact 1.7.3. Suppose $|A + A| \leq k|A|$, then $E(A) \geq |A|^3/k$.

Proof. Write $r(x) = |\{(a_1, a_2) \in AxA : a_1 + a_2 = x\}|$. Then $\sum r(x) = |A|^2$ and $\sum r(x)^2 = E(A)$. Note that $|\operatorname{supp}(r(x))| = |A + A| \leq k|A|$ where $\operatorname{supp}(r(x)) = \{x \ : \ r(x) \neq 0\}$. Thus

$$|A|^4 = \Big(\sum r(x)\Big)^2 \leq_{C-S} \sum_{x \in \operatorname{supp}(r)} 1 \sum_x r(x)^2 \leq k|A|E(A)\,.$$

(The first inequality follows by Cauchy-Schwarz.) □

Remark 1.7.4. The converse is not true: Take A= $B_1 \cup B_2$ where B_1 and B_2 are skew subgroups with large additive energy but a big doubling.

Theorem 1.7.5 (Balog-Szemerédi-Gowers)**.** *Let $k \geq 2$ and suppose $E(A) \geq \frac{|A|^3}{k}$. Then, $\exists A' \subseteq A$ with $|A'| \geq K^{-c}|A|$, and $|A' + A'| \leq K^c|A'|$.*

Definition 1.7.6 (Bipartite graph)**.** A graph whose vertices can be divided into two disjoint sets U and V such that every edge connects a vertex in U to one in V.

Proposition 1.7.7. *Let $0 < \alpha < 1/2$. Take a bipartite graph on vertex set $V \cup W$ with $|V| = |W| = n$ and αn^2 edges. Then there are sets $V' \subseteq V$ and $W' \subseteq W$ with cardinalities satisfying*

$$\frac{|V'|}{|V|}, \frac{|W'|}{|W|} \geq \alpha^c$$

such that between any vertices $x \in V'$ and $y \in W'$, there are $\geq \alpha^c n^2$ paths of length 3 between x and y.

We'll deduce it from:

Lemma 1.7.8 (Paths of length 2)**.** *Adopt the same assumptions as above and let $0 < \eta < 1$. Then there is a set $V' \subseteq V$ with $|V'| \geq \alpha n/2$ such that for at least a fraction $1 - \eta$ of pairs $x, y \in V'$, there are at least $\eta\alpha^2 n/2$ paths of length 2 between x and y.*

Remark 1.7.9. This not true when $\eta = 0$.

Proof. Let E denote the edges in G. Then

$$\mathbb{E}_{w \in W}\, \mathbb{E}_{v \in V}\, 1_{vw \in E} \geq \alpha \quad \text{and hence} \quad \mathbb{E}_{w \in W}\, \mathbb{E}_{v,v' \in V}\, 1_{vw \in E} 1_{v'w \in E} \geq \alpha^2$$

(squaring and using Cauchy–Schwarz). Let $N(v)$ denote the set of vertices in the neighberhood of v. Then

$$\mathbb{E}_{v,v' \in V}\, |N(v) \cap N(v')| \geq \alpha^2 n\,.$$

Here, as above, we write $\mathbb{E}_{w\in W}$ for $\frac{1}{n}\sum_{w\in W}$. Say that v, v' are *antisocial* if $|N(v)\cap N(v')| \le \eta\alpha^2 n/2$ and let S denote the set of antisocial pairs. Combining the last two inequalities,

$$\mathbb{E}_{v,v'\in V}\,(\eta - 1_{(v,v')\in S})|N(v)\cap N(v')| \ge \frac{\eta\alpha^2 * n}{2}$$

and hence

$$\mathbb{E}_{v,v'\in V}\,(\eta - 1_{(v,v')\in S})\sum_{w\in W} 1_{v\in N(w)}1_{v'\in N(w)} \ge \frac{\eta\alpha^2 n}{2}\,.$$

Pulling the sum over w to the outside and pigeonholing, there is at least one $w \in W$ such that

$$\mathbb{E}_{v,v'\in V}\,(\eta - 1_{(v,v')\in S}) * 1_{v,v'\in N(w)} \ge \frac{\eta * \alpha^2 n}{2}\,.$$

What does this mean? Take $V' = N(w)$; the fact that the last equation is ≥ 0 already says that at most η of the pairs $(v, v') \in$ V are antisocial:

$$(\eta - 1)\eta\ \text{(antisocial)} + \eta(1-\eta)\ \text{(not antisocial)} = 0\,.$$

That is to say for at least $1-\eta$ of the pairs $v, v' \in V'$, v and v' have at least $\frac{\eta\alpha^2 n}{2}$ common neighbors. The last equation (divided by η) implies

$$\mathbb{E}_{v,v'\in V}\,1_{v,v'\in V'} \ge \frac{\alpha^2}{2}\,.$$

This implies that $|V'| \ge \frac{\alpha|V|}{\sqrt{2}}$. □

Proof sketch of Proposition 1.7.7. Assume all vertices in V have degree at least $\alpha n/2$. Apply the lemma giving V' such that almost all pairs $x, y \in V'$ are social. Work a little more to get everyone in V'' sociable with almost everyone else. Find W' such that every $y \in W'$ is joined to many vertices in V''. □

1.7.1 Proof of the Balog-Szemerédi-Gowers theorem

We'll prove the following: If A, B are subsets of an abelian group with $|A| = |B| = n$ and $|\{a_1 + b_1 = a_2 + b_2\}| \ge \frac{n^3}{k}$ then $\exists A' \subseteq A, B' \subseteq B$ so that

$$\frac{|A'|}{|A|}, \frac{|B'|}{|B|} \ge k^{-c} \quad \text{and} \quad |A' - B'| \le k^c n\,.$$

Remark 1.7.10. Why does this imply the first version? We get A', A'' with $|A' + A''| \le k^c n$. By averaging, there is x: $|A' \cap (A'' + x)| \ge k^{-c} * n$. Take $A''' = A' \cap (A'' + x)$ and we have $|A''' + A'''| \le k^{c''} n$.

Proof. Idea: Apply proposition on paths of length 3 to the "popular sum graph" of A and B. Take a bipartite graph on vertex sets A, B. Join a to b (and say $a + b$ is *popular*) if

$$r(x) = |\{(a', b') \in AxB : a' + b' = x\}| \ge \frac{n}{2k}\,.$$

We showed earlier that this graph has many edges. Let A', B' be as in the path of length 3 proposition. Then if $a' \in A'$ and $b' \in B'$, there are many b_1, a_1such that $a' + b_1$, $b_1 + a_1$, and $a_1 + b'$ are all popular sums. But then $a' + b' = (a' + b_1) + (b_1 + a_1) + (a_1 + b')$. So one can write $a' + b'$ as a sum of 3 popular sums x_1, x_2, x_3 in $\ge k^{-c}n^2$ ways. But the number of popular sums is manifestly $\le 2kn$ (otherwise we would have $\ge \frac{2knn}{2k} = n^2$ elements in $A \times B$). Therefore $|A' + B'|k^{-c}n^2 \le (2kn)^3$ which implies $|A' + B'| \le k^{c'} n$. □

LECTURE 7

Lecturer: Ben Green *Scribe: Pierre McKenzie*

The following is known:

Theorem 1.7.11 (Green-Tao-Konyagin)**.** *If $A \subseteq \mathbb{F}_2^K$ and $|A+A| \leq K|A|$ then $A \subseteq H$, for a subspace H of size at most $2^{2K}K^c|A|$ and for some constant c.*

This motivates the following conjecture:

Conjecture 1.7.12 (Polynomial Freiman-Ruzsa conjecture)**.** Under the hypotheses of the Green-Tao-Konyagin theorem, $A \subseteq \cup_{i=1}^{M}(H+x_i)$ with H a subspace of size $\leq |A|$ and $M = K^{O(1)}$.

Today's lecture deals with approximate homomorphisms, which one might more appropriately call approximate affine functions. Today we will use the following one of many possible (often equivalent) notions of approximate homomorphisms:

Definition 1.7.13 (K-approximate homomorphism)**.** The function $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^{n'}$ is a K-approximate homomorphism if for every $x, y \in \mathbb{F}_2^n$,

$$\varphi(x+y) = \varphi(x) + \varphi(y) + s_{x,y},$$

where $s_{x,y} \in S$ and S is an "error set" with $|S| \leq K$.

Suppose that $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^{n'}$ is a K-approximate homomorphism with error set S. Can we express φ as $\psi + \varepsilon$ where $\psi : \mathbb{F}_2^n \to \mathbb{F}_2^{n'}$ is linear and $\varepsilon : \mathbb{F}_2^n \to \mathbb{F}_2^{n'}$ is somehow small? A trivial upper bound on such an $|\operatorname{im}\varepsilon|$ is 2^K. This trivial bound is obtained by setting $\psi(g) = \varphi(g)$ for each g in a generating set for $\mathbb{F}_2^n$ and extending ψ to a linear homomorphism. Since $\psi(x) - \varphi(x)$ belongs to the linear span of S for each $x \in \mathbb{F}_2^n$, an ε with $|\{\varepsilon(x) \ : \ x \in \mathbb{F}_2^n\}| \leq 2^{|S|} \leq 2^K$ does the job.

Conjecture 1.7.14. An ε exists with $|\operatorname{im}\varepsilon| \leq K^{O(1)}$.

This conjecture can be shown equivalent to the Polynomial Freiman-Ruzsa Conjecture 1.7.12. We will prove the following weaker form of it, obtained very recently by Schoer.

Theorem 1.7.15 (Schoer 2010)**.** *Given a K-approximate homomorphism $\varphi : \mathbb{F}_2^n \to \mathbb{F}_2^{n'}$, there is a linear homomorphism $\psi : \mathbb{F}_2^n \to \mathbb{F}_2^{n'}$ and a function $\varepsilon : \mathbb{F}_2^n \to \mathbb{F}_2^{n'}$ such that $\varphi(x) = \psi(x) + \varepsilon(x)$ for all $x \in \mathbb{F}_2^n$ and $|\operatorname{im}\varepsilon| \leq 2^{2^{c\sqrt{\log K}}}$ for some c (hence $|\operatorname{im}\varepsilon| < 2^{K^{o(1)}}$, significantly improving on the trivial 2^K bound).*

We will do most of the proof today, and finish it tomorrow. Here is an outline of the proof:

Step 0. By applying Cauchy-Schwartz (left as an exercise), we obtain that

$$\mathbb{P}[\varphi(x_1) + \varphi(x_2) = \varphi(x_3) + \varphi(x_4) \ : \ x_1 + x_2 = x_3 + x_4] \geq K^{-2}.$$

Step 1. (Ruzsa) There is a set $A \subseteq \mathbb{F}_2^n$, $|A| \geq K^{-c}N$, such that φ is a *Freiman* 16*-homomorphism* on A, that is to say: for any $a_i, a_i' \in A$, if

$$a_1 + \cdots + a_{16} = a_1' + \cdots a_{16}'$$

then

$$\varphi(a_1) + \cdots + \varphi(a_{16}) = \varphi(a'_1) + \cdots \varphi(a'_{16}) .$$

Here $N = 2^n$ and c is a new constant; note that constants such as c here will generally depend on each other, but we will often skim over these chains of dependencies.

Observe that φ gives a well-defined map $\tilde{\varphi}$ on $B = 8A$ by defining

$$\tilde{\varphi}(a_1 + \cdots + a_8) = \varphi(a_1) + \cdots + \varphi(a_8) .$$

(Just add $8a_1$ to both sides of an equality $a_1 + \cdots + a_8 = a'_1 + \cdots + a'_8$ and deduce from φ being a 16-homomorphism that $\varphi(a_1) + \cdots + \varphi(a_8) = \varphi(a'_1) + \cdots \varphi(a'_8)$). In fact, if $b_1, b_2, b'_1, b'_2 \in B$ and $b_1 + b_2 = b'_1 + b'_2$ then $\tilde{\varphi}(b_1) + \tilde{\varphi}(b_2) = \tilde{\varphi}(b'_1) + \tilde{\varphi}(b'_2)$. Schoer's ingredient to the proof is that $8A$ contains a subspace of co-dimension $2^{c\sqrt{\log K}} = K^{o(1)}$. (A 1937 theorem of Bogolyubov's shows that $4A$ contains a subspace of co-dimension $K^{c'}$.)

Step 2. From here it is "relatively easy," later upgraded to "not too hard," to conclude, since $\tilde{\varphi}$ is already known to be linear on a large set.

So let us spell out **Step 1**.

Note that

$$\mathbb{P}[\varphi(x_1) + \varphi(x_2) = \varphi(x_3) + \varphi(x_4) \,:\, x_1 + x_2 = x_3 + x_4] \geq K^{-c}$$

is equivalent to

$$E(\Gamma) \geq K^{-c}|\Gamma|^3$$

where $\Gamma = \{(x, \varphi(x)) \;:\; x \in \mathbb{F}_2^n\} \subseteq \mathbb{F}_2^{n+n'}$. We can thus apply the Balog-Szemerédi-Gowers theorem proved in the last lecture to Γ, to obtain $\Gamma' \subseteq \Gamma$ with $|\Gamma'| \geq K^{-c}|\Gamma|$ and $|2\Gamma'| \leq K^c|\Gamma'|$. Note that $(x_1, y_1) \neq (x_2, y_2)$ for $(x_1, y_1), (x_2, y_2) \in \Gamma'$ implies $x_1 \neq x_2$ by definition of Γ. So denote Γ' by Γ_A for the set $A \subseteq \mathbb{F}_2^n$, $|A| = |\Gamma_A| \geq K^{-c}N$, such that Γ' is the graph of the restriction $\varphi|_A$ of φ on A.

Now look at $32\Gamma_A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^{n'}$. By sumset estimates,

$$33\Gamma_A \leq K^{O(1)}N .$$

Note that $33\Gamma_A$ contains $\Gamma_A + (0, S)$ where $S = (32\Gamma_A)_0$ is the fibre of $32\Gamma_A$ above 0, that is, $S = \{s \;:\; (0, s) \in 32\Gamma_A\}$. (The fiber of a set of pairs above an element a is just the set of elements b such that (a, b) is a pair in the set.) Note that S is precisely the set of values taken by

$$\varphi(a_1) + \cdots + \varphi(a_{16}) - \varphi(a'_1) - \cdots - \varphi(a'_{16})$$

as $a_1, \ldots, a_{16}, a'_1, \ldots, a'_{16}$ range over A with

$$a_1 + \cdots + a_{16} = a'_1 + \cdots a'_{16} .$$

Now $|S| \leq K^c$ because the map $((x, \varphi(x)), s) \mapsto (x, s + \varphi(x))$ is an injection from $\Gamma_A \times S$ to $\Gamma_A + (0, S)$, so that $K^{-c}N|S| \leq |\Gamma_A||S| \leq |\Gamma_A + (0, S)| \leq |33\Gamma_A| \leq K^{c''}N$.

So now we have a set $A \subseteq \mathbb{F}_2^n$, $|A| \geq K^{-c}N$, and a set $S \subseteq \mathbb{F}_2^{n'}$, $|S| \leq K^c$, such that

$$\begin{array}{c} a_1, \dots, a_{16} = a'_1, \dots, a'_{16} \\ \Downarrow \\ \varphi(a_1) + \cdots + \varphi(a_{16}) - \varphi(a'_1) - \cdots - \varphi(a'_{16}) \in S. \end{array}$$

A discussion broke out at this point to the effect that the existence of such an S could have been deduced directly from the hypotheses of Theorem 1.7.15 without the need to first weaken the hypothesis (in step 0) and then appeal to the Balog-Szemerédi-Gowers theorem.

By the probabilistic method, let $v_1, \dots, v_m \in \mathbb{F}_2^{n'}$ with $m = 1 + \log_2 |S|$ be such that $S \cap v_1^\perp \cap \cdots \cap v_m^\perp = \{0\}$. Take A' to be any set of the form

$$\{x \in A \ : \ \langle \varphi(x), v_1 \rangle = b_1, \cdots, \langle \varphi(x), v_m \rangle = b_m\}.$$

By pigeonholing, there is a choice of $b_1, \dots, b_m$ such that

$$|A'| \geq 2^{-m}|A| \geq K^{-c'}N.$$

Note that $a_1, \dots, a_{16}, a'_1, \dots, a'_{16} \in A'$ and $a_1 + \cdots + a_{16} = a'_1 + \cdots + a'_{16}$ imply

$$\varphi(a_1) + \cdots + \varphi(a_{16}) - \varphi(a'_1) - \cdots - \varphi(a'_{16}) \in S \cap v_1^\perp \cap \cdots \cap v_m^\perp = \{0\}.$$

Schoer's contribution was to find $X, Y \subseteq 2A$ with a large value of γ, in the terminology of Proposition 1.7.16 below. This will be explained in the next lecture. The rest of this lecture is devoted to proving Proposition 1.7.16.

Proposition 1.7.16 (Bogoluykov, Chang, Ruzsa)**.** *Suppose $X, Y \subseteq \mathbb{F}_2^n$, $|X| = \alpha N$, $|Y| = \beta N$ and*

$$E(X, Y) \geq \gamma |X|^2 |Y|$$

where $E(X, Y) = |\{(x, x', y, y') \in X \times X \times Y \times Y \ : \ x + y = x' + y'\}|$. Then $2X + 2Y$ contains a subspace of size $\geq \alpha^{c/\gamma} N$.

Recall the Fourier transform, specialized to the setting $f : \mathbb{F}_2^n \to \mathbb{C}$:

$$\begin{aligned} \hat{f} \ : \ \mathbb{F}_2^n &\to \mathbb{C} \\ r &\mapsto \mathbb{E}_x \, f(x)(-1)^{r^T x}. \end{aligned}$$

Lemma 1.7.17 (Chang)**.** *Suppose $A \subseteq \mathbb{F}_2^n$, $|A| = \alpha N$. Let*

$$R = \{r \in \mathbb{F}_2^n \ : \ |\hat{\mathbf{1}}_A(r)| \geq \rho\alpha\}.$$

Then $\dim(\langle R \rangle) \leq 8\rho^{-2} \log(\frac{1}{\alpha})$.

Note that Parseval's identity applied to A taken from Chang's lemma yields

$$\Sigma_r |\hat{\mathbf{1}}_A(r)|^2 = \mathbb{E}_x \, \mathbf{1}_A(x)^2 = \alpha.$$

This yields $|R| \leq \rho^{-2}\alpha^{-1}$ which is much weaker than the bound claimed in Chang's lemma.

Proof of Lemma 1.7.17 (Chang's lemma). Take independent elements $r_1, \dots, r_d \in R$, $d = \dim(R)$. We know $|\hat{\mathbf{1}}_A(r_i)| \geq \rho\alpha$. By translating A by r_i if necessary, we can assume

$$\hat{\mathbf{1}}_A(r_i) \geq \rho\alpha. \tag{1.1}$$

For each $\omega = (\omega_1, \omega_2, \dots, \omega_d) \in \{0,1\}^d$ write α_ω for the average of $\mathbf{1}_A$ on the set

$$\{x \ : \ r_1^T x = \omega_1, \dots, r_d^T x = \omega_d\}.$$

Expanding the $\hat{\mathbf{1}}_A$ in (1.1) yields the following equivalent statement:

$$\mathbb{E}_\omega (-1)^{\omega_i} \alpha_\omega \geq \rho\alpha \qquad \text{for } 1 \leq i \leq d.$$

We will maximize α in the system of 2^d linear equations

$$\mathbb{E}_\omega(\alpha_\omega) = \alpha, 0 \leq \alpha_\omega \leq 1, \quad \text{for} \quad \omega = (\omega_1, \omega_2, \dots, \omega_d) \in \{0,1\}^d.$$

Consider

$$x_\omega = \max\left(0, \frac{2\langle\omega\rangle}{d\rho} - 1\right)$$

where $\langle\omega\rangle = (-1)^{\omega_1} + \cdots + (-1)^{\omega_d}$. Observe that

$$\mathbb{E}_\omega(1 - \frac{\langle\omega\rangle}{d\rho})\alpha_\omega = \frac{1}{d\rho}\sum_{i=1}^{d}(\rho - (-1)^{\omega_i})\alpha_\omega \tag{1.2}$$

$$= \frac{1}{d\rho}\sum_{i=1}^{d}(\rho\alpha - \hat{\mathbf{1}}_A(r_i)) \leq 0. \tag{1.3}$$

Then

$$\alpha = \mathbb{E}_\omega \alpha_\omega \leq \mathbb{E}_\omega(2 - \frac{2\langle\omega\rangle}{d\rho} + x_\omega)\alpha_w \tag{1.4}$$

$$\leq \mathbb{E}_\omega x_\omega \alpha_\omega \qquad \text{by (1.3)} \tag{1.5}$$

$$\leq \mathbb{E}_\omega x_\omega \tag{1.6}$$

$$\approx \mathbb{P}[\langle\omega\rangle \geq d\rho/2]. \tag{1.7}$$

But we can compute that $\mathbb{E}_\omega x_\omega \leq e^{-\rho^2 d/8}$ from (1.7) using Chernoff bounds. See Section 1.8 for an alternate proof of Chang's lemma due to Bourgain. □

Proof of Proposition 1.7.16. Write $f = \mathbf{1}_X * \mathbf{1}_X * \mathbf{1}_Y * \mathbf{1}_Y$. If $f(t) > 0$, then $t \in 2X + 2Y$. By Fourier inversion, $f(x) = \Sigma_r \hat{f}(r)(-1)^{r^T x}$. Since the Fourier transform of a convolution is the convolution of the Fourier transforms, $\hat{f} = \hat{\mathbf{1}}_X^2 \hat{\mathbf{1}}_Y^2$. Hence

$$f(t) = \sum_r |\hat{\mathbf{1}}_X(r)|^2 |\hat{\mathbf{1}}_Y(r)|^2 (-1)^{r^T t}.$$

Let $e = \gamma^{1/2}/2$ and define $R = \{r \ : \ |\hat{\mathbf{1}}_X(r)| \geq \rho\alpha\}$.
Claim: $\mathbf{1}_X * \mathbf{1}_X * \mathbf{1}_Y * \mathbf{1}_Y(t) > 0$ if $t \in R^\perp$, a subspace. At this point the result follows from Chang's lemma and a computation.

To see the claim, if $t \in \langle R \rangle^\perp$, we have $r^T t = 0$ for $r \in \langle R \rangle$, so

$$f(t) \geq \sum_{r \in \langle R \rangle} |\hat{\mathbf{1}}_X(r)|^2 |\hat{\mathbf{1}}_Y(r)|^2 - \sum_{r \notin \langle R \rangle} |\hat{\mathbf{1}}_X(r)|^2 |\hat{\mathbf{1}}_Y(r)|^2 \tag{1.8}$$

by assuming -1 for $r^T t$ as worst case in the negative term. Then by adding and subtracting $\Sigma_{r \notin \langle R \rangle}$,

$$\geq \sum_r |\hat{\mathbf{1}}_X(r)|^2 |\hat{\mathbf{1}}_Y(r)|^2 - 2 \sum_{r \notin \langle R \rangle} |\hat{\mathbf{1}}_X(r)|^2 |\hat{\mathbf{1}}_Y(r)|^2. \tag{1.9}$$

But the $\sum_r$ equals $||\mathbf{1}_X * \mathbf{1}_Y||^2$ and this is $\frac{E(X,Y)}{N^3}$ by Parseval.

Now

$$\sum_{r \notin \langle R \rangle} |\hat{\mathbf{1}}_X(r)|^2 |\hat{\mathbf{1}}_Y(r)|^2 \leq \rho^2 \alpha^2 \Sigma_r |\hat{\mathbf{1}}_Y(r)|^2 \tag{1.10}$$

$$= \rho^2 \alpha^2 \beta \qquad \text{by Parseval} \tag{1.11}$$

$$= \frac{\gamma}{4} \alpha^2 \beta \tag{1.12}$$

$$= \frac{\gamma}{4} \frac{|X|^2 |Y|}{N^3}. \tag{1.13}$$

Hence the right-hand side of (1.8) $\geq \gamma \frac{|X|^2|Y|}{N^3} - \frac{2}{4}\gamma \frac{|X|^2|Y|}{N^3} \geq \frac{1}{2}\gamma \frac{|X|^2|Y|}{N^3} \geq 0$, proving the claim. □

1.8 An alternate proof of Chang's lemma

Ryan O'Donnell recalled the following, alternate proof of Chang's lemma due to Bourgain.

Theorem 1.8.1 (Chang's Lemma, restated)**.** *Let $f : \mathbb{F}_2^n \to [0,1]$ satisfy $\mathbb{E}[f(x)] = \alpha < 1/2$ and $\hat{f}(r_i) \geq \rho\alpha$ for a family $\{r_1, \ldots, r_d\}$ of linearly independent characters. Then $d = O(\ln(\frac{1}{\alpha})/\rho^2)$.*

Proof. Let $g : \mathbb{F}_2^n \to \mathbb{R}$ be the function $g(x) = \sum_{i=1}^d (-1)^{r_i^T x}$. Then

$$\mathbb{E}[f(x)g(x)] = \sum_{i=1}^d \hat{f}(r_i) \geq \rho\alpha d\,.$$

Considering that $r_1, \ldots, r_d$ are linearly independent, the random variables $r_1(x), \ldots, r_d(x)$ determined by selecting x uniformly at random in $\mathbb{F}_2^n$ are independent. To upper bound the quantity above, choose a threshold $t > 0$ so that

$$\mathbb{P}[\sum_i r_i(x) \geq t] \geq \alpha$$

and let $A = \{x \ : \ \sum_i r_i(x) \geq t\}$. Since $\mathbb{E}[f] = \alpha$, we must have

$$\mathbb{E}[f(x)g(x)] \leq \mathbb{E}[g(x)\mathbf{1}_A]\,.$$

As $g(x)$ is nearly Gaussian, with expectation zero and variance d, we may take the threshold t above to be $c\sqrt{\ln 1/\alpha}\sqrt{d}$ and it follows that $d = O(\ln(\frac{1}{\alpha})/\rho)$. □

LECTURE 9

Lecturer: Ben Green *Scribe: Shachar Lovett*

The plan for the talk today is:

1. Finish Schoen's argument.
2. Approximate subgroups of $\mathsf{SL}_2(p)$.

1.9 Completion of Schoen's argument

Recall from previous talk: we have a function $\phi : \mathbb{F}_2^n \to \mathbb{F}_2^{n'}$ such that

$$\phi(x+y) = \phi(x) + \phi(y) + s_{x,y}\,,$$

where $s_{x,y} \in S$ and $|S| = K$. Schoen showed that that $\phi(x) = \ell(x) + \epsilon(x)$ where $\ell(x)$ is linear and

$$|\mathrm{Im}(\epsilon)| \le 2^{2^{c\sqrt{\log k}}}\,.$$

We showed that there exists $A \subset \mathbb{F}_2^n$ such that ϕ restricted to A is a Freiman 16-homomorphism, i.e. for any $a_1, \cdots, a_{16}, a'_1, \cdots, a'_{16} \in A$ such that $a_1 + \cdots + a_{16} = a'_1 + \cdots + a'_{16}$ we have $\phi(a_1) + \cdots + \phi(a_{16}) = \phi(a'_1) + \cdots + \phi(a'_{16})$. This implies that we can define ϕ uniquely on $8A$, and moreover, ϕ is a 2-homomorphism on $8A$. Hence our goal will to be to find a large linear subspace $V \subset 8A$. This will imply that ϕ is linear on V.

Let $N = 2^n$. We have seen the following proposition of Bogoliubov:

Proposition 1.9.1. *If $X, Y \subset \mathbb{F}_2^n$, $|X| = \alpha N$, $|Y| = \beta N$ and $E(X,Y) = \gamma |X|^2 |Y|$. Then $2X + 2Y$ contains a linear subspace of size at least $\alpha^{c/\gamma} N$.*

The task at hand is thus: given A such that $|A + A| \le K|A|$, find $X, Y \subset 2A$ with large $E(X,Y)$. We will now describe Schoen's new idea. It was also discovered (in other contexts) by Tom Sanders. We recall the definition of Sym.

Definition 1.9.2 (Sym-set)**.** Let $A \subset \mathbb{F}_2^n$. $\mathrm{Sym}_\delta(A)$ is defined as

$$\begin{aligned}\mathrm{Sym}_\delta(A) &= \{x \ : \ |A \cap (A+x)| \ge \delta |A|\} \\ &= \{x \ : \ x = a - a' \text{ in at least } \delta|A| \text{ ways}\}\,.\end{aligned}$$

We will need some basic facts about Sym.

Fact 1.9.3. Let Sym be defined as above.

(1) If $|A + A| \le K|A|$ then $\mathrm{Sym}_{1/2K}(A) \ge |A|/2$.

Proof. This is a simple averaging argument. Let $r(x) = |\{a, a' \in A \ : \ a + a' = x\}|$. Then $\mathrm{Sym}_{1/2K}(A) = \{x \ : \ r(x) \ge 1/2K\}$. We have $r(x) \le A$, $\sum_x r(x) = |A|^2$ and

$$\sum_{x \notin \mathrm{Sym}_{1/2K}(A)} r(x) \le |A + A| \cdot |A|/2K \le \frac{|A|^2}{2}\,.$$

Hence $\sum_{x \in \mathrm{Sym}_{1/2K}(A)} r(x) \ge |A|^2/2$ and $|\mathrm{Sym}_{1/2K}(A)| \ge |A|/2$. □

(2) $E(A, \mathrm{Sym}_\delta(A)) \geq \delta^2 |\mathrm{Sym}_\delta(A)|^2 |A|$.

Proof. Write $t(a)$ for the number of $s \in \mathrm{Sym}_\delta(A)$ such that $a + s \in A$. Then

$$\sum_{a \in A} t(a) \geq \delta |\mathrm{Sym}_\delta(A)||A| .$$

Hence

$$E(A, \mathrm{Sym}_\delta(A)) = \sum_a t(a)^2 \geq \sum_{a \in A} t(a)^2 \geq \frac{1}{|A|} \sum_{a \in A} t(a) = \delta^2 |\mathrm{Sym}_\delta(A)|^2 |A| ,$$

where the last inequality follows by Cauchy-Schwarz. □

Before proving Schoen's theorem, note that if A was a random subset of $\mathbb{F}_2^n$ of density $1/K$ then $A \cap (A + x) \approx K^{-2}N \approx K^{-1}|A|$, hence $\mathrm{Sym}_\delta(A) = \{0\}$ for $\delta \gg 1/K$. But, $2A = \mathbb{F}_2^n$ hence $\mathrm{Sym}_{0.99}(2A)$ is huge. This is the advantage of summing several copies of A.

Lemma 1.9.4 (Schoen)**.** *Suppose $A \subset \mathbb{F}_2^n$ such that $|A + A| \leq K|A|$. Then for any $\delta > 0$ there exists a set $A' \subset 2A$, $|A'| \geq |A|$ such that*

$$|\mathrm{Sym}_\delta(A')| \geq (2K)^{1 - 2^{\frac{\log K}{\log 1/\delta}}} |A| .$$

Note that if $\delta = K^{-0.01}$ that $\mathrm{Sym}_\delta(A') \geq K^{-c}|A|$ for $c = 2^{100}$, hence we get large symsets even for polynomial small δ.

Proof. Set $t = \lceil \log K / \log(1/\delta) \rceil$. We will construct a sequence of sets $B_0 \supseteq B_1 \supseteq \ldots \supseteq B_t$. Define $B_0 = A$. We will denote $|B_i| = \beta_i |A|$, and we will see that

$$\beta_i \geq (2K)^{1-2^i}. \tag{1.14}$$

Define B_{i+1} to be the set among the set of intersections $\{B_i \cap (B_i + x) \ : \ x \in \mathrm{Sym}_{\beta_i/2K}(B_i)\}$ for which $|A + B_{i+1}|$ is minimal. Note that the set of intersections is nonempty as always 0 is in the symset. We will in fact prove later that the symset is quite large. Note that (1.14) follows from the definition since for any $x \in \mathrm{Sym}_{\beta_i/2K}(B_i)$ we have $|B_i \cap (B_i + x)| \geq \beta_i/2K \cdot |B_i| = \beta_i^2/2K \cdot |A|$.

Consider the sequence of sets $A + B_0, A + B_1, \ldots, A + B_t$. For any $0 \leq i \leq t$ we have $A \subseteq A + B_i \subseteq 2A$, and by assumption $|2A| \leq K|A|$. Hence by the pigeonhole principle and the choice of t there must exist a pair of sets such that

$$|A + B_i| \geq \delta |A + B_{i+1}|.$$

Define $A' = A + B_i$. We now observe that for $x \in \mathrm{Sym}_{\beta_i/2K}(B_i)$ we have

$$\begin{aligned}
|A' \cap (A' + x)| &= |(A + B_i) \cap (A + B_i + x)| \\
&\geq |A + (B_i \cap (B_i + x))| \\
&\geq |A + B_{i+1}| && \text{(by minimality of } B_{i+1}) \\
&\geq \delta |A + B_i| = \delta |A'| && \text{(by the pigeonhole principle).}
\end{aligned}$$

Hence we conclude that $\mathrm{Sym}_{\beta_i/2K}(B_i) \subseteq \mathrm{Sym}_\delta(A')$, and hence $|\mathrm{Sym}_\delta(A')| \geq |B_i|/2 \geq \beta_i/2|A|$. □

1.10 Approximate homomorphisms which are far from genuine homomorphisms

We have seen thus far that approximate homomorphisms and subgroups of $\mathbb{F}_2^n$ are in fact close to genuine homomorphisms/subgroups. We will later see this is also the case in $\mathsf{SL}_2(p)$. One might wonder if this is the general case for all groups. The next example shows this is not true in general (it is true for $\mathbb{F}_2^n$ because it has high torsion, and for $\mathsf{SL}_2(p)$ because it is a quasi-random group, i.e. its doesn't have small irreducible representations).

Example 1.10.1. Let $G = \mathbb{Z}/N\mathbb{Z} = \{1, 2, \dots, N\}$. Define the following set

$$A = \{1 \leq n \leq N \ : \ -0.1 \leq [n\sqrt{2}] \leq 0.1\}$$

where $[x]$ for a real number x denotes its fractional part, which is between -0.5 and 0.5. Define a function $\phi : A \to \mathbb{R}/\mathbb{Z}$ by

$$\phi(n) = \sqrt{3}[n\sqrt{2}].$$

It is simple to see that for $a_1, a_2 \in A$ we have $\phi(a_1 + a_2) = \phi(a_1) + \phi(a_2)$, since if $x, y \in \mathbb{R}$ are such that $\{x\}, \{y\} \in [-0.1, 0.1]$ then $\{x + y\} = \{x\} + \{y\}$. On the other hand, if $\psi : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a genuine homomorphism one can verify that $\mathbb{P}[\phi(x) = \psi(x)] \lesssim N^{-1/2}$.

1.11 Approximate subgroups of $\mathsf{SL}_2(p)$

We recall that $\mathsf{SL}_2(p)$ is the group of 2×2 matrices over $\mathbb{F}_p$ with determinant 1. We denote $G = \mathsf{SL}_2(p)$ for the reminder of this section. We will prove the following theorem.

Theorem 1.11.1 (Helfgott)**.** *Suppose $A \subset G$ such that A generates G. Let $K \geq 2$. Then one of the following holds:*

(1) A is very small: $|A| \leq K^C$.

(2) A is very large: $|A| \geq K^{-C}|G|$.

(3) A is not an approximate subgroup: $|A \cdot A \cdot A| \geq K|A|$.

Helfgott's theorem can be interpreted as follows: approximate subgroups which generate $\mathsf{SL}_2(p)$ are basically $\{\mathbb{1}\}$ or $\mathsf{SL}_2(p)$. There are generalizations of Helfgott's theorem to $\mathsf{SL}_n(p)$ and to more general lie groups by Breuillard, Green, and Tao [5] and by Pyber and Szabo [16].

One can deduce from Helfgott's theorem the following corollary. Recall that a function $\nu : G \to \mathbb{R}$ is a probability measure if $\nu \geq 0$ and $\sum_x \nu(x) = 1$. The definition of the ℓ_2 norm is $\|\nu\| = (\mathbb{E}_{x \in G}\, \nu(x)^2)^{1/2}$. Note that for delta functions $\|\delta_g\| = \sqrt{|G|}$.

Corollary 1.11.2. *Suppose $\nu : G \to \mathbb{R}$ is a probability measure with $\|\nu\| < |G|^{1/2-\delta}$. Then one of the following holds (for appropriate constants $c(\delta)$ and δ')*

(1) $\| \underbrace{\nu * \dots * \nu}_{c(\delta)\ times} \| \leq |G|^{-\delta}$.

(2) $\nu(H) \geq |G|^{-\delta'}$ for some proper subgroup $H < G$.

Note that (1) and Babai-Nikolov-Pyber theorem implies that $2c(\delta)$ convolutions of ν yields an almost uniform distribution over $\mathsf{SL}_2(p)$, since the minimal representation of $\mathsf{SL}_2(p)$ has size $p = |G|^{1/3}$. We now sketch the proof of the corollary.

Proof sketch. Suppose that $\nu(H) < |G|^{-\delta'}$ for all $H \lneq G$.

(1) It is enough to show that $\|\nu * \nu\| < |G|^{-\epsilon}\|\nu\|$; this can be iterated.

(2) If $\|\nu * \nu\| \geq |G|^{-\epsilon}\|\nu\|$, then by a variant of the Balog-Szemerédi-Gowers theorem for measures we get that ν is close to the uniform measure over H, where H is a subgroup of G. But by Helfgott's theorem H cannot generate G unless ν was already close to uniform, hence ν must have mass on a proper subgroup.

□

We now proceed to proof Helfgott's theorem.

Proof sketch of Theorem 1.11.1. Fix K and suppose that $|A \cdot A \cdot A| \leq K|A|$. Suppose that A generates $\mathsf{SL}_2(p)$. We will show that either $|A| \leq K^C$ or $|A| \geq K^{-C}|G|$. We may assume w.l.o.g that $A = A^{-1}$, otherwise we take $A = A \cup A^{-1}$. We will in fact work in the algebraic closure $\mathsf{SL}_2(\overline{\mathbb{F}_p})$ in order to take eigenvalues, etc.

A *torus* is a subgroup of $\mathsf{SL}_2(\overline{\mathbb{F}_p})$ which is conjugate to the group of diagonal matrices. That is, let

$$D = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} : \lambda \in \overline{\mathbb{F}_p} \right\}$$

be the group of diagonal matrices. Then $T = xDx^{-1}$ for any $x \in \mathsf{SL}_2(\overline{\mathbb{F}_p})$ is a torus. We note that all tori are conjugate.

The following fact can be easily verified by simple linear algebra for 2×2 matrices: if $x \in T$ for some torus T, then either $x = \pm \mathrm{id}$ or x has distinct eigenvalues. An element which is not $\pm \mathrm{id}$ and has distinct eigenvalues is called *regular semi-simple*.

The normalizer of a group $H < G$ is defined as $N_G(H) = \{x \in G \ : \ xHx^{-1} = H\}$. Let T be a torus. The *Weyl group* is defined as $W = N_G(T)/T$. A key fact in the case of tori in $\mathsf{SL}_2(p)$ is that $|W| = 2$. This can be seen by first changing basis so that $T = D$ and then noting that

$$N_G(D) = D \cup \left\{ \begin{pmatrix} 0 & -\lambda \\ \lambda^{-1} & 0 \end{pmatrix} : \lambda \in \overline{\mathbb{F}_p}^* \right\}.$$

Assume now that $A \subset G$ and $|A \cdot A \cdot A| \leq K|A|$. We first fix notation. Denote $X \lesssim Y$ if $X \leq K^c Y$ for some fixed constant c, and $X \approx Y$ if $X \lesssim Y$ and $Y \lesssim X$; that is, we ignore fixed powers of K. The following is a key fact: suppose T is a torus which contains at least one regular semi-simple element of A. then $|A^2 \cap T| \approx |A|^{1/3}$.

This was proved by Helfgott, but it can also be derived from a more general theorem due to Larson and Pink [12]. We quote it here for $\mathsf{SL}_n(\overline{\mathbb{F}_p})$ but it holds for any semi-simple lie group.

Theorem 1.11.3 (Larsen-Pink)**.** *Assume $|A \cdot A \cdot A| \leq K|A|$. Let V by any subvariety of $G = \mathsf{SL}_n(\overline{\mathbb{F}_p})$. Then*

$$|A \cap V| \leq |A|^{\frac{\dim(V)}{\dim(G)}}.$$

Note that the Larsen-Pink theorem implies immediately that $|A^2 \cap T| \lesssim |A|^{1/3}$. This is true since $\dim(G) = 3$ and $\dim(T) = \dim(D) = 1$ (note that D, hence also T, is a subvariety, since it is characterized by the off-diagonal elements being 0 and the determinant being 1, which are both algebraic relations). We now sketch the proof that $|A^2 \cap T| \gtrsim |A|^{1/3}$.

Let $x \in A \cap T$. Look on $C(x) = \{g^{-1}xg \ : \ g \in G\}$. Note that if $x \neq \pm \mathrm{id}$ then $C(x)$ is a subvariety of G of dimension 2, since $y \in C(x)$ iff it shares the same eigenvalues as x, which can be characterized by $\mathrm{Det}(y) = 1$ and $\mathrm{tr}(y) = \mathrm{tr}(x)$ (we note that $C(x)$ is a subvariety also in $\mathsf{SL}_n(\overline{\mathbb{F}_p})$, as the eigenvalues can be uniquely described by polynomial equations on the elements of the matrix). Hence by the Larsen-Pink theorem (applied to $A^3 = A \cdot A \cdot A$) we get

$$|A^3 \cap C(x)| \lesssim |A|^{2/3}.$$

Consider the set of elements $\{a^{-1}xa \ : \ a \in A\}$. They are all contained in $A^3 \cap C(x)$, hence there must be distinct $m \gtrsim |A|^{1/3}$ elements $a_1, \ldots, a_m \in A$ such that

$$a_1^{-1}xa_1 = \cdots = a_m^{-1}xa_m.$$

Thus we get that all $a_i a_j^{-1}$ commute with x. From this we have that $a_i a_j^{-1} \in T$ for all $1 \leq i, j \leq m$, hence in particular $a_i a_1^{-1} \in A^2 \cap T$, i.e. $|A^2 \cap T| \gtrsim |A|^{1/3}$.

Let T be a torus. We say T is *involved* with A if $A^2 \cap T$ contains a regular semi-simple element (i.e. some $x \neq \pm \mathrm{id}$). We will prove the following proposition: if $|A| > K^{100}$ then the set of involved tori is invariant under conjugation by elements of A. That is, if T is involved then so is $a^{-1}Ta$ for any $a \in A$. This will conclude the proof: we first verify that A^2 intersects at least one torus. This could be guaranteed by the "Escape from subvarieties lemma" of Elkin, Mozes and Oh [9], which states in our case that A^c for a large enough constant c must intersect some torus. Thus, we will carry the entire argument to A^c, from which we will deduce that $|A^c| \approx |G|$, but from Ruzsa's theorem this will imply that $|A| \approx |G|$. Thus, we may assume without loss of generality that there exists at least one involved torus. Since A generates G this implies that in fact all tori must be involved. That is, A^2 intersects all tori. Moreover we already proved that the size of any such intersection is $\approx |A|^{1/3}$. As the number of tori is about $p^2 = |G|^{2/3}$, and distinct tori intersect only at $\{\mathrm{id}\}$, we get that

$$|A| \approx |A^2| \approx |G|^{2/3}|A|^{1/3}$$

hence $|A| \approx |G|$ as we aimed to prove.

Thus, in order to conclude, we need to prove that the set of involved tori is closed under conjugation by $a \in A$. Let T be an involved torus, and let $\tilde{T} = a^{-1}Ta$ be any conjugate torus where $a \in A$. We will prove $\tilde{T}$ is also involved. Consider the set $\{x\tilde{T}y \ : \ x, y \in A^2\}$. Assume $x', y' \in A$ such that $x' = xt'$ and $y' = t''y$. Then $x\tilde{T}y = x'\tilde{T}y'$. We have proved already that $|A^2 \cap T| \approx |A|^{1/3}$. Hence the number of distinct $x\tilde{T}y$ for $x, y \in A^2$ is at most $|A|^2/|A^2 \cap T|^2 \lesssim |A|^{4/3}$. Hence by the pigeonhole principle there exist $b_1, \ldots, b_m, c_1, \ldots, c_m \in A^2$ for $m \approx |A|^{2/3}$ such that

$$b_1\tilde{T}c_1 = \cdots = b_m\tilde{T}c_m\,.$$

We now observe that there must be at least $\ell = \sqrt{m} \approx |A|^{1/3}$ distinct b_i's or ℓ distinct c_i's. Assume without loss of generality that $b_1, \ldots, b_\ell$ are distinct. Hence we get that

$$(b_1\tilde{T}b_1^{-1})(b_1c_1) = \cdots = (b_\ell\tilde{T}b_\ell^{-1})(b_\ell c_\ell)\,.$$

Note that each $b_i \tilde{T} b_i^{-1}$ is a subgroup of G, hence each $(b_i \tilde{T} b_i^{-1})(b_i c_i)$ is a coset of subgroup of G. This implies that

$$b_1 \tilde{T} b_1^{-1} = \cdots = b_\ell \tilde{T} b_\ell^{-1} .$$

The reason is simple: in general, if H, K are subgroups of G, and $Hx = Ky$ are equal cosets of H, K, then we must have $H = K$. Thus, we can continue to deduce that all $b_i^{-1} b_j \in N_G(\tilde{T})$. Since in particular all the elements $b_1^{-1} b_1, \ldots, b_1^{-1} b_\ell$ are distinct, we get that

$$|A^4 \cap N_G(\tilde{T})| \gtrsim |A|^{1/3} .$$

Since we have already shown that $[N_G(T) : T] = |W| = 2$, this implies that also

$$|A^4 \cap \tilde{T}| \gtrsim |A|^{1/3}.$$

Hence A^4 intersects all tori (and this is enough to conclude the argument from before, as this will imply that $|A^4| \approx |G|$ hence also $|A| \approx |G|$). □

1.12 Sum-product theorem

The sum-product theorem of Bourgain, Katz and Tao [4].

Theorem 1.12.1 (Sum-product theorem)**.** *Let $A \subset \mathbb{F}_p$ such that $|A| \leq p^{0.9}$. Then* $\max(|A + A|, |A \cdot A|) \geq |A|^{1.01}$. *Equivalently, if $|A + A|, |A \cdot A| \leq K|A|$ then either $|A| \leq K^c$ or $|A| \geq K^{-c}p$ (i.e. there are no non-trivial approximate sub-rings of $\mathbb{F}_p$).*

The following Lemma states that if A doesn't increase much under both additions and multiplications, then it contains a subset which doesn't increase much under polynomials.

Lemma 1.12.2 (Katz-Tao lemma)**.** *Assume $|A + A|, |A \cdot A| \leq K|A|$. Then there is $A' \subset A$, $|A'| \geq K^{-c}|A|$ such that $|A' \cdot A' + A' \cdot A'| \leq K^c|A'|$.*

Chapter 2

Representation theory of finite groups, and applications

A series of 5 lectures by Avi Wigderson.

LECTURE 2

Lecturer: Avi Wigderson *Scribe: Anil Ada & Laszlo Egri*

2.1 Some applications of representation theory

This lecture will introduce formally the notion of group representations. Before we do that, we start by mentioning some of the applications of representation theory that will be discussed in subsequent lectures.

One major application of representation theory is to understanding random walks on groups, or expansion in groups. Let G be a group and S a subset of G. There is a natural graph, the Cayley graph, that one can define with respect to G and S. We denote this graph by $\mathrm{Cay}(G, S)$. The vertices are the elements of the group G and there is an edge from g to g' if there is $s \in S$ such that $gs = g'$. Note that if S is closed under taking inverses, then the Cayley graph becomes undirected. We will always deal with such S.

The main question we will ask about these graphs is whether they are expanding (we will define expanders formally later). Informally we can say that expansion controls the convergence of random walks on the graph. Also, expanders are graphs in which the diameter is smallest possible.

If $\mathrm{Cay}(G, S)$ is an expander then we say that S is an *expanding generating set*. We will be interested in the following questions: Which groups have "small" expanding generating set? When is "small" $O(1)$? Does expansion depend only on the group G or does it also depend on the choice of S? More specifically, we will cover the following applications.

- How long does it take to reach a near perfect random deck of cards if we shuffle the deck by taking two cards at random and swap them? Diaconis and Shahshahani [8] show that $O(n \log n)$ swaps lead to a nearly perfect random deck. (If one uses standard techniques, one obtains $O(n^2 \log n)$ swaps.)
- Alon and Roichman [1] show that for every group G, most generating sets S of size $O(\log |G|)$ is expanding.
- Lubotzky and Weiss [13], and Meshulam and Wigderson [15] answer the question about the size of expanding generating sets in t-step solvable groups.
- Cohn and Umans [6] present a representation theoretic approach to fast matrix multiplication. This approach can potentially achieve the optimum exponent for matrix multiplication.
- A dimension expander is a useful linear algebraic generalization of a standard expander. Lubotzky and Zelmanov [14] explicitly construct such objects for fields of characteristic 0.

2.2 Representation theory of finite groups

2.2.1 Group actions and representations

In all our lectures, we will be dealing with finite groups and algebraically closed fields where the characteristic of the field does not divide the size of the group. There is a theory for more general groups and fields but we will not need it for our purposes.

Let G be a finite group with $|G| = n$. We say that a group G *acts* on a finite set Ω if there is a homomorphism $\rho : G \to S_{|\Omega|}$. In other words, for all $x \in G$, we have a permutation $\rho(x) : \Omega \to \Omega$ and these permutations

satisfy $\rho(x) \cdot \rho(y) = \rho(xy)$ for all $x, y \in G$, i.e. the actions of the group elements respect the group operation. The main objective is to understand every possible action of every possible group.

Here are some examples:

- Let G be any finite group and let $\Omega = G$. Every element $x \in G$ defines a permutation on G by left multiplication. Furthermore, $x(yz) = (xy)z$ and so G acts on itself by left multiplication.
- Let $G = S_k$ be the symmetric group. Then G trivially acts on $[k]$ by permuting the elements. Also, G acts on $\left[\binom{k}{2}\right]$, e.g. G acts on graphs on k vertices.
- Let
$$G = \mathsf{SL}_2(p) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \ : \ a, b, c, d \in \mathbb{F}_p,\ ad - bc = 1 \right\}.$$
Then G acts on $\mathbb{F}_p^2$ by viewing the elements of G as linear transformations. One can also define the Möbius action as follows. G acts on $\{0, 1, \ldots, p-1\} \cup \{\infty\}$ by
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \alpha = \frac{a\alpha + b}{c\alpha + d}.$$

We will see how one can describe all possible actions of a given group. Furthermore, we will see that this theory generalizes the Fourier transform over Abelian groups to non-Abelian groups.

In order to understand an object very well in mathematics, a common trick is to turn it into a linear object. We will now see a canonical way of doing this in the case of group actions.

Let $\pi : \Omega \to \Omega$ be a permutation and $\mathbb{F}$ any field. We can extend π naturally to a linear operator $\tilde{\pi} : \mathbb{F}^\Omega \to \mathbb{F}^\Omega$ that permutes the coordinates of the vectors in $\mathbb{F}^\Omega$ according to π (the elements of $\mathbb{F}^\Omega$ can be viewed as functions $f : \Omega \to \mathbb{F}$ or equivalently as vectors over $\mathbb{F}$ whose coordinates are indexed by Ω). Observe that $\tilde{\pi}$ is a linear operator over a vector space since $\tilde{\pi}(f + g) = \tilde{\pi}(f) + \tilde{\pi}(g)$ and $\tilde{\pi}(cf) = c\tilde{\pi}(f)$.

Suppose we have an action of G characterized by ρ. Using the above method, we obtain linear operators $\tilde{\rho}(x)$ on $\mathbb{F}^\Omega$. The homomorphism $\tilde{\rho}$ now defines an action of G on $\mathbb{F}^\Omega$. The idea is to study the linear objects $\tilde{\rho}(x)$ rather than the non-linear objects $\rho(x)$.

We can think of $\tilde{\rho}(x)$ as a matrix in $\mathsf{GL}_{|\Omega|}(\mathbb{F})$. So we can identify the linear action $\tilde{\rho}(x)$, $x \in G$, with the homomorphism $\tilde{\rho} : G \to \mathsf{GL}_{|\Omega|}(\mathbb{F})$. Note that in this setting $\tilde{\rho}$'s image is a collection of permutation matrices but we will study more generally all homomorphisms from G to $\mathsf{GL}_{|\Omega|}(\mathbb{F})$. (We'll drop the tilde from now on.)

Definition 2.2.1. We say that $\rho : G \to \mathsf{GL}_d(\mathbb{F})$ is a *G-representation* if ρ is a group homomorphism.

2.2.2 Maschke's theorem and irreducible representations

We will view a G-representation ρ as follows.

$$\begin{array}{ccccc} g_1 & & g_2 & \cdots & g_n \\ \downarrow & & \downarrow & & \downarrow \\ \left(\begin{array}{c} \\ \rho(g_1) \\ \\ \end{array} \right) & \left(\begin{array}{c} \\ \rho(g_2) \\ \\ \end{array} \right) & \cdots & \left(\begin{array}{c} \\ \rho(g_n) \\ \\ \end{array} \right) \end{array}$$

Here $g_1, g_2, \ldots, g_n$ represent the group elements and $\rho(g_1), \ldots, \rho(g_n)$ the corresponding matrices. We do not distinguish between two representations that are the same up to a change of basis. Therefore we say that two representations τ and σ are equivalent, denoted $\tau \cong \sigma$, if there is an invertible Z such that for all $x \in G$, $\sigma(x) = Z\tau(x)Z^{-1}$.

To understand G-representations, we will identify the building blocks of representations, i.e. the *irreducible* representations. There is a nice analogy between representations and integers. Every integer has a unique decomposition into its prime factors where a prime appears a certain number of times in the decomposition. Primes are the building blocks of integers which cannot be further decomposed. We will see that representations have a very similar structure in the sense that they can be decomposed into *irreducible* representations where each irreducible representation appears a certain number of times.

Suppose W is a non-trivial subspace of V such that $\rho(x)W \subseteq W$ for all $x \in G$. In this case, we call W ρ-invariant. It is clear that $\text{span}\{\rho(x)W\} = W$ since $\rho(\text{id}) = I_d$. We can now hope to change basis so that we separate the action of the matrices $\rho(x)$ on W and on the complement of W. This hope is realized using Maschke's Theorem.

Theorem 2.2.2 (Maschke's Theorem)**.** *If W satisfies the above condition, then there exists a subspace U of V so that $V = W \oplus U$ and $\rho(x)U \subseteq U$ for all $x \in G$, i.e. U is ρ-invariant.*

Given Maschke's Theorem, we can apply an invertible transformation Z, $\rho(x) \mapsto Z\rho(x)Z^{-1}$, to do a change of basis to obtain block diagonal matrices:

$$\begin{array}{ccccc} g_1 & & g_2 & \cdots & g_n \\ \downarrow & & \downarrow & & \downarrow \\ \left(\begin{array}{c|c} \tau(g_1) & 0 \\ \hline 0 & \sigma(g_1) \end{array}\right) & \left(\begin{array}{c|c} \tau(g_2) & 0 \\ \hline 0 & \sigma(g_2) \end{array}\right) & \cdots & \left(\begin{array}{c|c} \tau(g_n) & 0 \\ \hline 0 & \sigma(g_n) \end{array}\right) \end{array}$$

In this case, we can write $\rho = \tau \oplus \sigma$ where τ and σ are G-representations. We call ρ *irreducible* if it cannot be decomposed as above.

Clearly, we can keep applying Maschke's Theorem recursively to a representation (in an arbitrary order) until we end up with a collection of irreducible representations. Later we will show that this procedure gives a unique decomposition up to equivalence of representations, into irreducible representations. Furthermore we will show that G has finitely many irreducible representations.[1] There are many nice facts about irreducible representations which we will prove in the next lecture. For now let's state two of them. If we denote all the irreducible representations by $\rho_1, \ldots, \rho_t$ with dimensions $d_1, \ldots, d_t$ respectively, then

(1) $\sum_{i=1}^{t} d_i^2 = n$,

(2) t is equal to the number of conjugacy classes of G.[2]

Note that these statements are independent of the field, so long as it is algebraically closed and does not have characteristic that divides $|G|$. If G is Abelian then $t = n$ with $d_i = 1$ for all i. In this case, the

[1] A note on the computational complexity of computing all the irreducible representations of a given group: Babai and Rónyai [3] give a polynomial time algorithm that given G's multiplication table as input, produces all the irreducible representations of G.

[2] For $x \in G$, the conjugacy class of x is $\{yxy^{-1} \ : \ y \in G\}$.

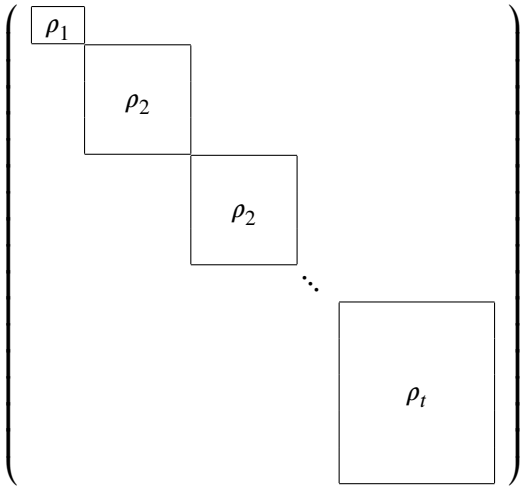
homomorphisms $\rho : G \to \mathbb{C}$, which are also known as the characters of the group, are exactly the irreducible representations. This gives us an idea on how group representations generalize the notion of Fourier transform over Abelian groups.[3]

Define the functions $\chi_i : G \to \mathbb{F}$ by $\chi_i(x) = \mathrm{Tr}(\rho_i(x))$. These are called the characters (generalizing the Abelian case) and provide very useful information about the representations. Information about the characters along with the dimensions $d_1, \ldots, d_t$ will be utilized in applications.

2.2.3 Finding all irreducible representations

Our task now is to find all the irreducible representations of a group. The most natural representation that one can try to decompose is the action of G on itself. We give this representation a special name and call it the *regular representation*. So in the regular representation $R : G \to \mathsf{GL}_n(\mathbb{F})$, the matrix for $R(x)$ is a permutation matrix which contains a 1 in the the coordinate (y, z) if and only if $x = yz^{-1}$, i.e. $y = xz$.

We can now recursively apply Maschke's Theorem to obtain the decomposition $R = \bigoplus_{i=1}^{t} m_i \rho_i$, where $\rho_1, \ldots, \rho_t$ are the distinct (non-equivalent) set of irreducible representations and $m_i \rho_i$ denotes $\underbrace{\rho_i \oplus \cdots \oplus \rho_i}_{m_i}$.



It turns out that $\{\rho_1, \ldots, \rho_t\}$ is the set of *all* irreducible representations of G. Furthermore, $m_i = d_i = \dim(\rho_i)$ for all i. That is,

Theorem 2.2.3. *In the decomposition of the regular representation R, each irreducible representation appears the number of times equal to its dimension.*

This theorem, the facts stated earlier and much more follows elegantly from one basic and fundamental lemma called Schur's Lemma. The proof of these statements via Schur's Lemma will be presented in the second lecture. We end this lecture with the statement and proof of Schur's Lemma.

Lemma 2.2.4 (Schur's Lemma)**.** *Let σ_1 and σ_2 be irreducible representations of G with dimensions d_1 and d_2 respectively. Suppose A is a $d_1 \times d_2$ dimensional matrix such that*

$$\forall x \in G : \quad A = \sigma_1(x) A \sigma_2(x)^{-1}. \tag{2.1}$$

[3]The homomorphisms (or characters) $\rho : G \to \mathbb{C}$, when G is Abelian, form an orthonormal basis for the vector space of functions $\{f : G \to \mathbb{C}\}$. This is the Fourier basis and the Fourier expansion of f is the expression for f as a linear combination of the characters.

If $\sigma_1 \ncong \sigma_2$, then $A = 0$. If $\sigma_1 = \sigma_2$, then $A = \lambda \mathbb{1}$.

Proof. We claim that $\mathrm{Ker}(A)$ is σ_2-invariant and $\mathrm{Im}(A)$ is σ_1-invariant. To see this, let $v \in \mathrm{Ker}(A)$. Then by (2.1), we know $0 = \sigma_1(x)Av = A\sigma_2(x)v$, which implies $\sigma_2(x)v \in \mathrm{Ker}(A)$. Hence $\mathrm{Ker}(A)$ is σ_2-invariant. Similar argument shows $\mathrm{Im}(A)$ is σ_1-invariant. Since σ_1 and σ_2 are irreducible, we know $\mathrm{Ker}(A)$ and $\mathrm{Im}(A)$ are either 0 or the whole space.

For the first case, suppose $A \neq 0$. We will show $\sigma_1 \cong \sigma_2$. If $\mathrm{Ker}(A)$ is the whole space or $\mathrm{Im}(A)$ is 0 then $A = 0$ so $\mathrm{Ker}(A) = 0$ and $\mathrm{Im}(A) = \mathbb{F}^{d_1}$. Since $d_2 = \dim(\mathrm{Im}(A)) + \dim(\mathrm{Ker}(A))$, we get $d_1 = d_2$. Furthermore, since $\mathrm{rank}(A) = \dim(\mathrm{Im}(A))$ we know that A is invertible. Then rewriting (2.1), we obtain $A^{-1}\sigma_1(x)A = \sigma_2(x)$, i.e. $\sigma_1 \cong \sigma_2$.

For the second case, since $\sigma_1(x) = \sigma_2(x)$, we know $d_1 = d_2$. If $A = 0$ then we are done so suppose $A \neq 0$. Then we know $\mathrm{Ker}(A)$ is not the whole space and hence $\mathrm{Ker}(A) = 0$, i.e. A is invertible. This means that A has a non-zero eigenvalue λ. Define A' to be $A - \lambda\mathbb{1}$. Then it is easy to verify that A' satisfies (2.1). Thus, as done with A, we can conclude that $\mathrm{Ker}(A')$ is σ_2-invariant and therefore is either 0 or the whole space. It cannot be 0 since the eigenvector corresponding to λ is in the kernel. So we conclude that $A' = 0$, i.e. $A = \lambda\mathbb{1}$. □

LECTURE 4

Lecturer: Avi Wigderson *Scribe: Phuong Nguyen*

We will discuss some properties of the regular representation of a group, which will be useful for understanding the group algebra. These are important in showing that certain Cayley graphs are expanders. In fact the first examples of expander graphs are Cayley graphs.

2.3 The regular representation

In this section we will show that the regular representation of a group G contains all information about the irreducible representations of G.

Two representations τ and σ are said to be *isomorphic* if one can be obtained from the other by change of basis, i.e., there is some invertible matrix Z so that for all x:

$$\tau(x) = Z\sigma(x)Z^{-1}\,.$$

Recall that the *character* χ_ρ of a representation ρ is

$$\chi_\rho(x) = \mathrm{tr}(\rho(x))\,.$$

i.e., $\chi_\rho(x)$ is the trace of the matrix $\rho(x)$.

The main theorem of today's lecture is:

Theorem 2.3.1. *Let R be the regular representation of G, and suppose that by applying (in arbitrary order) the procedure given by Maschke's Theorem R is decomposed into*

$$R = \bigoplus_{i=1}^{t} m_i\rho_i$$

where the ρ_i are irreducible and distinct up to isomorphism. Then

1. *The set $\{\rho_1, \rho_2, \ldots, \rho_t\}$ does not depend on the process by which R is decomposed. Furthermore, $\{\rho_1, \rho_2, \ldots, \rho_t\}$ are all irreducible representations of G.*
2. *Let d_i be the dimension of ρ_i, then $m_i = d_i$ for $1 \le i \le t$.*
3. *$\sum_{i=1}^{t} d_i^2 = n$.*
4. *t is the number of conjugacy classes of G.*
5. *The characters of $\{\rho_1, \rho_2, \ldots, \rho_t\}$ are orthonormal. When G is Abelian, the characters form an unitary matrix.*
6. *In general, the set of character functions $\{\chi_{\rho_i} \;:\; 1 \le i \le t\}$ spans the space of all class functions of G.*

By 1 the set $\{\rho_1, \rho_2, \dots, \rho_t\}$ is well defined, and it will be denoted by $\mathrm{Irrep}(G)$, the set of all irreducible representations of G.

Before proving this theorem, let's have a look at an example. Let G be the permutation group S_3:

$$G = \{\mathrm{id}, (123), (132), (12), (13), (23)\}.$$

Note that S_3 is non-commutative. Consider the action ρ of S_3 on the set $\{1, 2, 3\}$. Here each element $x \in G$ is a permutation on the indices $\{1, 2, 3\}$, and $\rho_x = \rho(x)$ is the linear map (say on $\mathbb{C}^3$)

$$\rho_x \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} v_{x(1)} \\ v_{x(2)} \\ v_{x(3)} \end{pmatrix}.$$

In particular, the values of ρ_x are:

x	id	(123)	(132)	(12)	(13)	(23)
ρ_x	$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$	$\begin{pmatrix} & & 1 \\ 1 & & \\ & 1 & \end{pmatrix}$	$\begin{pmatrix} & 1 & \\ & & 1 \\ 1 & & \end{pmatrix}$	$\begin{pmatrix} & 1 & \\ 1 & & \\ & & 1 \end{pmatrix}$	$\begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ & & 1 \\ & 1 & \end{pmatrix}$

Notice that

$$\rho_x \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

for all x, so the subspace spanned by $(1, 1, 1)$ is invariant for ρ. By Maschke's Theorem the orthogonal subspace $U = \langle (1, 1, 1) \rangle^{\perp}$ is also invariant for ρ. Consider the following basis for U:

$$\left\{ \begin{pmatrix} 1 \\ \omega \\ \omega^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \omega^2 \\ \omega \end{pmatrix} \right\}$$

where $\omega = e^{2\pi i/3}$ is a 3rd-root of unity. Using this basis we can describe the decomposition of ρ easily. For example, consider $\rho_{(123)}$:

$$\begin{pmatrix} 1 \\ \omega \\ \omega^2 \end{pmatrix} \xrightarrow{(123)} \begin{pmatrix} \omega^2 \\ 1 \\ \omega \end{pmatrix} = \omega^2 \begin{pmatrix} 1 \\ \omega \\ \omega^2 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ \omega^2 \\ \omega \end{pmatrix} \xrightarrow{(123)} \begin{pmatrix} \omega \\ 1 \\ \omega^2 \end{pmatrix} = \omega \begin{pmatrix} 1 \\ \omega^2 \\ \omega \end{pmatrix}.$$

So $\rho_{(123)}$ becomes

$$\begin{pmatrix} 1 & & \\ & \omega^2 & \\ & & \omega \end{pmatrix}.$$

Similarly for ρ_x for others $x \in G$. Thus over the basis

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ \omega \\ \omega^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \omega^2 \\ \omega \end{pmatrix} \right\}$$

ρ is decomposed into the trivial representation and another representation which we call σ:

x	id	(123)	(132)
1	$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ & \omega^2 & \\ & & \omega \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ & \omega & \\ & & \omega^2 \end{pmatrix}$
σ_x			

x	(12)	(13)	(23)
1	$\begin{pmatrix} 1 & & \\ & & \omega \\ & \omega^2 & \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ & & \omega^2 \\ & \omega & \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ & & 1 \\ & 1 & \end{pmatrix}$
σ_x			

Now we have found here two irreducible representations for G, namely the trivial (of dimension 1) and σ (of dimension 2). Because $1^2 + 2^2 = 5$, using the theorem above we know that we have not found all irreducible representations of G; in fact one representation of dimension 1 is missing. A good guess gives us the homomorphism from G to $\{\pm 1\}$ that maps each permutation in G to its sign:

x	id	(123)	(132)	(12)	(13)	(23)
sign	1	1	1	-1	-1	-1

Again by the theorem we know that the regular representation of G is decomposed into

$$1 \oplus \text{sign} \oplus \sigma \oplus \sigma\,.$$

Finally, here are the characters of G:

x	id	(123)	(132)	(12)	(13)	(23)
χ_1	1	1	1	1	1	1
χ_{sign}	1	1	1	-1	-1	-1
χ_σ	2	-1	-1	0	0	0

For another example application of the theorem, suppose that G is an Abelian group. Then each element of G is itself a conjugacy class, and G has exactly n conjugacy classes. So $t = n$, and hence it follows from 3 that all $d_i = 1$, i.e., all irreducible representations of G have dimension 1. Note also that when G is Abelian, the character functions of G form an unitary matrix (see 5).

To prove the theorem we need Lemma 2.3.3 below which follows from Schur's Lemma. In general, given two representations σ and τ, in order to apply Schur's Lemma, we wish to find a matrix A so that

$$\forall x \in G : \qquad A = \sigma(x) A \tau(x)^{-1}\,. \tag{2.2}$$

Such a matrix A can in fact be obtained from an arbitrary matrix B by *symmetrization*:

Lemma 2.3.2. *Suppose that σ and τ are two representations of dimensions d_σ and d_τ respectively, and B is a $d_\sigma \times d_\tau$ matrix. Let*

$$A = \mathbb{E}_{y \in G}\left[\sigma(y) B \tau(y)^{-1}\right]$$

Then A satisfies (2.2).

The proof of this lemma is straightforward and is left as an exercise.

Thus by Schur's lemma, for any matrix B, we have

$$A = \mathbb{E}_{y\in G}\left[\sigma(y)B\tau(y)^{-1}\right] = \begin{cases} 0 & \text{if } \sigma \ncong \tau\,, \\ \lambda\mathbb{1} & \text{if } \sigma = \tau\,. \end{cases}$$

To state the next lemma we introduce the following notation. For each irreducible representation ρ of G let d_ρ denote the dimension of ρ. (In Theorem 2.3.1 we use d_i for d_{ρ_i}.) For such a ρ and i, j, $1 \le i, j \le d_\rho$, define the function $f_{\rho,i,j} : G \to \mathbb{C}$ so that

$$f_{\rho,i,j} : x \mapsto \rho(x)_{i,j}\,.$$

(In other words, $f_{\rho,i,j}(x)$ is the (i, j) entry of the matrix $\rho(x)$.)

For two functions f and g on G, define their inner product $\langle f, g\rangle$ by

$$\langle f, g\rangle = \mathbb{E}_{x\in G}\left[f(x)\overline{g(x)}\right]$$

where $\mathbb{E}$ denotes expectation, and $\overline{g(x)}$ is the complex conjugate of $g(x)$. Note that we will be using this inner product for group homomorphisms, so $\overline{g(x)} = g(x^{-1})$.

Lemma 2.3.3. *For any two representations σ, τ of G and $1 \le i, j \le d_\sigma$, $1 \le k, \ell \le d_\tau$ we have:*

$$\langle f_{\sigma,i,j}, f_{\tau,k,\ell}\rangle = \begin{cases} \frac{1}{d_\sigma} & \text{if } \sigma = \tau \ \wedge\ (i,j) = (k,\ell)\,, \\ 0 & \text{if } \sigma \ncong \tau \text{ or } (\sigma = \tau \text{ and } (i,j) \ne (k,\ell))\,. \end{cases}$$

Proof. We will use Schur's Lemma. Consider $B = B^{j,\ell}$ where $(B^{j,\ell})_{j,\ell} = 1$ and all other entries of $B^{j,\ell}$ are 0. Let

$$A^{j,\ell} = \mathbb{E}_{y\in G}\left[\sigma(y)B\tau(y)^{-1}\right] \quad \text{so that} \quad A^{j,\ell}_{i,k} = \mathbb{E}_{y\in G}\left[\sigma(y)_{i,j}(\tau(y)^{-1})_{\ell,k}\right].$$

Now $\tau(y)$ is equivalent to a permutation matrix, so $\tau(y)$ is unitary, i.e., $(\tau(y))^{-1}$ is just the conjugate of the transpose of $\tau(y)$. As a result, $(\tau(y)^{-1})_{\ell,k} = \overline{\tau(y)_{k,\ell}}$. Consequently,

$$A^{j,\ell}_{i,k} = \mathbb{E}_{y\in G}\left[\sigma(y)_{i,j}\overline{\tau(y)_{k,\ell}}\right]$$

so, by definition,

$$A^{j,\ell}_{i,k} = \langle f_{\sigma,i,j}, f_{\tau,k,\ell}\rangle\,.$$

Now, by Lemma 2.3.2, $A^{j,\ell}$ satisfies the hypothesis (2.2) of Schur's Lemma. Therefore (by Schur's Lemma):

$$A^{j,\ell} = \begin{cases} 0 & \text{if } \sigma \ncong \tau\,, \\ \lambda\mathbb{1} & \text{if } \sigma = \tau\,. \end{cases}$$

Now if $\sigma \ncong \tau$ then $A^{j,\ell} = 0$, i.e., $\langle f_{\sigma,i,j}, f_{\tau,k,\ell}\rangle = 0$ for all i, j, k, ℓ.

On the other hand, suppose that $\sigma = \tau$. Then $A^{j,\ell} = \lambda\mathbb{1}$ for some λ. So if $i \ne k$ we also have $A^{j,\ell}_{i,k} = 0$, i.e., $\langle f_{\sigma,i,j}, f_{\tau,k,\ell}\rangle = 0$.

Finally, suppose that $\sigma = \tau$ and $i = k$. Then $A^{j,\ell}_{i,i} = \lambda\mathbb{1}$ for some λ. In particular,

$$A^{j,\ell}_{i,i} = \frac{\operatorname{tr}(A^{j,\ell})}{d_\sigma}$$

(because $A^{j,\ell}$ is a $d_\sigma \times d_\sigma$ matrix). We calculate $\mathrm{tr}(A^{j,\ell})$:

$$\begin{aligned}
\mathrm{tr}(A^{j,\ell}) &= \mathrm{tr}(\mathbb{E}_{y\in G}[\sigma(y)B^{j,\ell}\sigma(y)^{-1}]) \\
&= \mathbb{E}_{y\in G}\big[\mathrm{tr}((\sigma(y)B^{j,\ell}\sigma(y)^{-1}))\big] \\
&= \mathbb{E}_{y\in G}\big[\mathrm{tr}(B^{j,\ell})\big] \qquad (\text{because } \mathrm{tr}(ZAZ^{-1}) = \mathrm{tr}(A)) \\
&= \begin{cases} 0 & \text{if } j \neq \ell\,, \\ 1 & \text{if } j = \ell\,. \end{cases}
\end{aligned}$$

Thus

$$A^{j,\ell}_{i,i} = \begin{cases} 0 & \text{if } j \neq \ell\,, \\ \frac{1}{d_\sigma} & \text{if } j = \ell\,. \end{cases}$$

This completes the proof of Lemma 2.3.3. □

Now we return to the proof of the main theorem.

Proof of Theorem 2.3.1. We prove the items listed in Theorem 2.3.1 in the order 5, 2, 3, 4, 6, 1.

First let $\rho \in \{\rho_1, \rho_2, \ldots, \rho_t\}$. Recall the characteristic function $\chi_\rho = \mathrm{tr}(\rho)$. So

$$\chi_\rho = \sum_{i=1}^{d_\rho} f_{\rho,i,i}\,.$$

For two different $\sigma, \tau \in \{\rho_1, \rho_2, \ldots, \rho_t\}$ we have $\sigma \ncong \tau$ so that, by Lemma 2.3.3, the functions $f_{\sigma,i,i}$ and $f_{\tau,j,j}$ are pairwise orthogonal. Therefore $\langle \chi_\sigma, \chi_\tau \rangle = 0$. On the other hand, also by Lemma 2.3.3 one can derive:

$$\langle \chi_\sigma, \chi_\sigma \rangle = \left\langle \sum_{i=1}^{d_\sigma} f_{\sigma,i,i}, \sum_{j=1}^{d_\sigma} f_{\sigma,j,j} \right\rangle = 1\,.$$

Thus we have shown that $\{\chi_{\sigma_i} \ : \ 1 \leq i \leq t\}$ are orthonormal. This establishes the first sentence of 5. When G is Abelian, from 3 (or also 4) we have $t = n$, i.e., we have n vectors χ_{σ_i}. Hence they form a unitary matrix. This proves the last sentence of 5.

Next, by assumption

$$R = \bigoplus_{i=1}^{t} m_i \rho_i$$

so the character of the regular representation χ_R is

$$\chi_R = \sum_{i=1}^{t} m_i \chi_i$$

(where $\chi_i = \chi_{\rho_i}$). As a result,

$$\langle \chi_j, \chi_R \rangle = \sum_{i=1}^{t} m_i \langle \chi_j, \chi_i \rangle = m_j\,.$$

We will compute $\langle \chi_j, \chi_R \rangle$ in another way and show that it is d_j; it will follow that $d_j = m_j$. By definition, $R(x)$ is simply the permutation by x, so for $x = \mathrm{id}$ (the identity element of G) we have $R(\mathrm{id}) = \mathbb{1}$, hence

$\chi_R(\mathrm{id}) = \mathrm{tr}(\mathbb{1}) = n$. On the other hand, for $x \neq \mathrm{id}$ the elements on the diagonal of $R(x)$ are all 0, so $\chi_R(x) = \mathrm{tr}(R(x)) = 0$. Now by definition,

$$\langle \chi_j, \chi_R \rangle = \mathbb{E}_{y \in G}\left[\mathrm{tr}(\rho_j(y)) \, \mathrm{tr}(R(y)) \right].$$

Using the values of $R(y)$ calculated above we get

$$\langle \chi_j, \chi_R \rangle = \frac{1}{n} \mathrm{tr}(\rho_j(\mathrm{id}))n = \mathrm{tr}(\rho_j(\mathrm{id})).$$

As $\rho_j(\mathrm{id})$ is the $d_j \times d_j$ identity matrix we have $\mathrm{tr}(\rho_j(\mathrm{id})) = d_j$. Therefore $\langle \chi_j, \chi_R \rangle = d_j$, and this proves 2.

Next we prove 3. From the above we have

$$\chi_R = \sum_{i=1}^{t} d_i \chi_i ;$$

hence

$$\chi_R(\mathrm{id}) = \sum_{i=1}^{t} d_i \chi_i(\mathrm{id}).$$

The LHS is simply n because $R(\mathrm{id})$ is the identity matrix of size $n \times n$, and the RHS is $\sum_{i=1}^{t} d_i^2$ because $\rho_i(\mathrm{id})$ is the identity matrix of size $d_i \times d_i$. This proves 3.

The proof of 4 is left as an exercise.

Note that

$$\mathrm{tr}(A) = \mathrm{tr}(ZAZ^{-1}) \quad \text{and so} \quad \mathrm{tr}(\rho(x)) = \mathrm{tr}(\rho(z)\rho(x)\rho(z^{-1})).$$

Therefore

$$\mathrm{tr}(\rho(x)) = \mathrm{tr}(\rho(zxz^{-1})).$$

In other words, the character functions are class functions. Therefore 6 follows immediately from 4 and the fact (established above) that the character functions are orthonormal.

Now we prove 1. It suffices to show that the set $\{\rho_1, \rho_2, \ldots, \rho_t\}$ contains all irreducible representations of G up to isomorphism. We prove this by contradiction. By Lemma 2.3.3 the set

$$\{ f_{\rho_s, i, j} \ : \ 1 \leq s \leq t, 1 \leq i, j \leq d_s \}$$

is an orthogonal subset of the functions on G, and by 3 this set consists of exactly n functions. As a result, this set spans the whole vector space of all functions on G. Now suppose for a contradiction that some irreducible representation ρ of G is not isomorphic to any in $\{\rho_1, \rho_2, \ldots, \rho_t\}$. Then the functions $f_{\rho, k, \ell}$ where $1 \leq k, \ell \leq d_\rho$ are also orthogonal to all functions $f_{\rho_s, i, j}$, a contradiction. □

2.4 Group algebras and Cayley graphs

Given a field $\mathbb{F}$ and a group G, consider the algebra, called the *group algebra* for G:

$$\mathbb{F}[G] = \{ f \ : \ f \text{ is a function from } G \text{ to } \mathbb{F} \}.$$

Here addition is defined by

$$(f + g)(x) = f(x) + g(x)$$

and multiplication is defined by convolution, as follows. Write each $f \in \mathbb{F}[G]$ as

$$\sum_{x \in G} f(x)\,\mathbf{x}$$

where the boldface $\mathbf{x}$ is a formal object which you may treat as the function taking the value 1 on x and zero elsewhere. If we define the product of two such formal objects so that $\mathbf{x} \cdot \mathbf{y} = \mathbf{xy}$, this yields the definition of $h = f \star g$ by linearity:

$$(\sum_{x \in G} f(x)\,\mathbf{x})(\sum_{y \in G} g(y)\,\mathbf{y}) = \sum_{z \in G} h(z)\,\mathbf{z}$$

where h is the function

$$h(z) = \sum_{w \in G} f(w) g(w^{-1} z)\,.$$

It can be verified that the vector space $\mathbb{F}[G]$ together with the convolution operation form an algebra of dimension $n = |G|$.

Using the regular representation R for G (where each element $x \in G$ is represented as an $n \times n$ matrix $R(x)$) we can represent each function $f \in \mathbb{F}[G]$ by a matrix $R(f)$:

$$R(f) = \sum_{x \in G} f(x) R(x)\,.$$

Note that

$$(R(f))_{y,z} = f(yz^{-1})\,.$$

In particular, the first column of $R(f)$ consists of the elements

$$(R(f))_{y,\mathrm{id}} = f(y)\,,$$

i.e., it gives us the graph of f.

It is easy to see that

$$R(f) + R(g) = R(f + g)$$

and it can be verified that

$$R(f)R(g) = R(f \star g)\,. \tag{2.3}$$

The latter can be seen as follows: For $y, z \in G$:

$$\begin{aligned}(R(f \star g))_{y,z} &= (f \star g)(yz^{-1}) \\ &= \sum_{w \in G} f(w) g(w^{-1} yz^{-1})\,.\end{aligned}$$

On the other hand, by definition of convolution:

$$\begin{aligned}(R(f)R(g))_{y,z} &= \sum_{u \in G} (R(f))_{y,u} (R(g))_{u,z} \\ &= \sum_{u \in G} f(yu^{-1}) g(uz^{-1}) \\ &= \sum_{w \in G} f(w) g(w^{-1} yz^{-1})\end{aligned}$$

where the last equality is obtained by letting $w = yu^{-1}$ (so $u = w^{-1}y$). This proves (2.3).

Notice that $\mathbb{F}[G]$ has dimension n (the size of G), while the space of all $n \times n$ matrices has dimension n^2. So the matrices that represent $\mathbb{F}[G]$ form a proper subspace of $M_n(\mathbb{F})$. The next theorem tells us what this subspace looks like.

Theorem 2.4.1 (Wederburn decomposition)**.** *Let $\mathbb{F}$ be an algebraically closed field, G a group of size n so that* $\mathrm{char}(\mathbb{F}) \nmid n$*. Let* $\mathrm{Irrep}(G) = \{\rho_1, \rho_2, \ldots, \rho_t\}$ *and let $d_i \times d_i$ be the dimension of ρ_i, for $1 \le i \le t$. Then*

$$\mathbb{F}[G] \cong \bigoplus_{i=1}^{t} M_{d_i}(\mathbb{F}) .$$

Here the isomorphism $\cong$ is under the Fourier Transform (i.e., a sequence of basis changes), and $M_{d_i}(F)$ denotes the space of all $d_i \times d_i$ matrices over $\mathbb{F}$.

The proof of this theorem is left as an exercise.

Definition 2.4.2 (Cayley graph)**.** Let G be a group and $S \subseteq G$. The Cayley graph $C(G, S)$ is defined as follows. The vertices of $C(G, S)$ are elements of G, and the edges of $C(G, S)$ are of the form (x, sx) for $s \in S$.

In most application S will be closed under taking inverse, i.e., $x^{-1} \in S$ for all $x \in S$. In such cases, $C(G, S)$ is really an undirected graph.

Now consider taking a random walk on the graph $C(G, S)$ by starting at the identity element and at each step we go from a vertex x to vertex sx with uniform probability over all $s \in S$. The probability transition matrix for such a random walk is the normalized adjacency matrix of $C(G, S)$, i.e., the $|G| \times |G|$ matrix whose (x, sx) entry is $1/|S|$ for all $s \in S$, and all other entries are 0. So the transition matrix is the regular representation $R(p_S)$ of the function

$$p_S = \frac{1}{|S|} \sum_{s \in S} \mathbf{s} .$$

We are interested in the eigenvalues of the above probability transition matrix. By diagonalizing (i.e., proper change of basis) this leads us to the eigenvalues of different irreducible components of $R(f)$. For example, the second largest eigenvalue of the probability transition matrix is the largest eigenvalue of all nontrivial irreducible components of G. Another useful piece of information is that, suppose that the nontrivial irreducible representations of G all have "high" dimension, i.e., d_i are large. Then in the decomposition of R each ρ_i has many (i.e., at least d_i) repetitions, because each ρ_i occurs d_i times. So except for the largest eigenvalue, all eigenvalues of $R(p_S)$ occur with many repetitions.

The following claim can be proved using Schur's Lemma.

Claim 2.4.3. *Suppose that S is a conjugacy class (or a union of conjugacy classes) of G. Then after the Fourier Transform, the regular representation $R(p_S)$ of p_S is a diagonal matrix.*

Another useful result is Parseval's identity for Fourier Transform.

Lemma 2.4.4 (Parseval's identity)**.** *When $\mathbb{F} = \mathbb{C}$, then*

$$n\|f\|^2 = \mathrm{tr}(R(f)^t R(f)) .$$

N. b. Both [17, 11] contain elementary introductions to the representation theory of finite groups.

LECTURE 6

Lecturer: Avi Wigderson *Scribe: Arkadev Chattopadhyay*

2.5 Introduction

We are going to consider expansion in groups, using the machinery of linear representations of groups. We shall assume throughout that our generating sets are closed under inverse, though we may not mention it explicitly. In particular, we will discuss the following results:

1. **Alon-Roichman:** Let G be a finite group and $S \subseteq G$ be a random subset, elements chosen independently and uniformly from G. Then there is a universal constant c so that if $|S| = c \log |G|$, the Cayley graph $\mathrm{Cay}(G, S)$ is an expander with high probability.

2. **Bourgain-Gamburd:** Recall that for any prime p, $\mathsf{SL}_2(p)$ is the group of 2×2 matrices over the field $\mathbb{F}_p$ with determinant 1. The result of Bourgain and Gamburd asserts that if we pick two random elements of this group, then the corresponding Cayley graph expands with high probability.

3. **Diaconis-Shahshahani:** Consider $\mathfrak{S}_k$, the symmetric group of permutations on k objects. Let $S = \{(1,2),(1,3),\ldots,(k-1,k)\}$ be the set of all transpositions. Then the random walk on $\mathrm{Cay}(\mathfrak{S}_k, S)$ converges to within ϵ (in L_1) of the uniform distribution in $\frac{1}{2}k\log k + c'\epsilon$ steps, for some constant c'.

2.6 Review of the group algebra

Recall that an algebra over a field $\mathbb{F}$ is a vector space over the field equipped with an additional operation of multiplication that is bilinear with respect to vector addition. Given a group G, the group algebra $\mathbb{F}[G]$ is the set of all functions $f : G \to \mathbb{F}$, each of which is expressed as the following formal sum:

$$\sum_{x\in G} f(x)\cdot x\,.$$

Multiplication of two functions is defined as follows:

$$\Big(\sum_{x\in G} f(x)\cdot x\Big)\Big(\sum_{y\in G} g(y)\cdot y\Big) \;=\; \sum_{z\in G} h(z)\cdot z\,,$$

where

$$h(z) = \sum_{w\in G} f(w)g(w^{-1}z)\,.$$

In other words, h is the convolution of f and g, denoted by $f * g$. This defines an algebra of dimension n, the order of G. Another convenient way to view this algebra is as the following matrix algebra: to each function f we associate the matrix

$$R(f) = \sum_{x\in G} f(x)R(x)\,,$$

where $R(x)$ is the matrix representation of the linear operator associated with the element $x \in G$ for the regular representation of G. Note that $R(f)$ by definition is a $|G| \times |G|$ matrix, such that $R(f)_{y,z} = f(yz^{-1})$. In other words,

$$f = R(f) \cdot \begin{pmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} \tag{2.4}$$

Armed with this observation, it is easy to verify the following two properties:

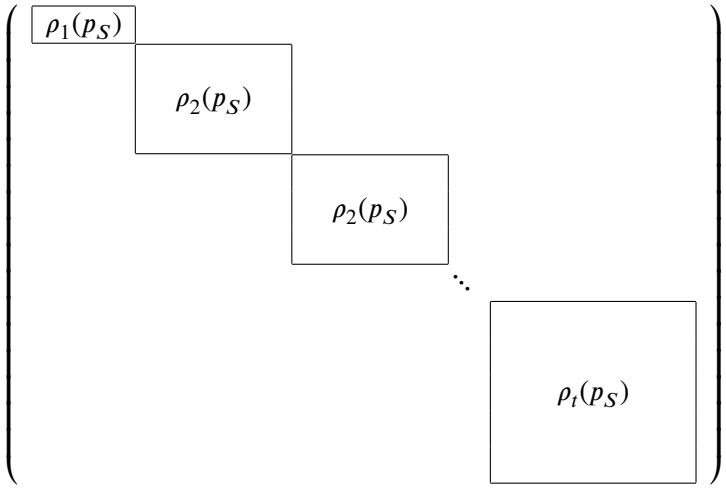
$$R(f+g) = R(f) + R(g); \quad R(f * g) = R(f)R(g)\,. \tag{2.5}$$

Thus, the matrix algebra generated by the matrices corresponding to the regular representation of G is isomorphic to the complex group algebra $\mathbb{C}[G]$.

2.7 Random walks

Let p be any probability distribution on elements of group G, i.e. $p : G \to \mathbb{R}$ is a non-negative function that satisfies $\sum_{x \in G} p(x) = 1$. Observe that p is an element of the group algebra $\mathbb{C}[G]$ and, in particular, $R(p)$ denotes the corresponding matrix in the associated matrix algebra. For each $S \subseteq G$, we define the probability distribution given by the following: $p_S(x) = 1/|S|$ if $x \in S$ and otherwise $p_S(x) = 0$. The simple but key observation is that $R(p_S)$ is the probability transition matrix of the random walk on $\mathrm{Cay}(G, S)$.

Recalling the machinery developed in previous lectures, let $\rho_1, \ldots, \rho_t$ be the irreducible representations of G. Let d_i be the dimension of ρ_i. Then we know that for any element $x \in G$, we can apply a set of unitary transformations to $R(x)$, called the Fourier transform (w.r.t a predetermined orthonormal basis) to transform it into the following block diagonal matrix: $R(x) = \bigoplus_{i=1}^{t} d_i \rho_i$. Thus, under this Fourier transform, $R(p_S)$ gives rise to the following block diagonal matrix:



where $\rho_i(p_S) = \sum_{x \in G} p_S(x)\rho_i(x)$. As always, we assume that our set S is closed under taking inverses. Thus, $R(p_S)$ is symmetric and has real eigenvalues. Further, note that $R(p_S)$ is a doubly stochastic matrix. Hence

its maximal eigenvalue is 1. Let, the eigenvalues be denoted as follows:

$$1 = \lambda_1(p_S) \geq \lambda_2(p_S) \geq \ldots \geq \lambda_n(p_S)\,.$$

Let

$$\lambda(p_S) \equiv_{\text{def}} \max_{i\geq 2}\left|\lambda_i(p_S)\right|$$

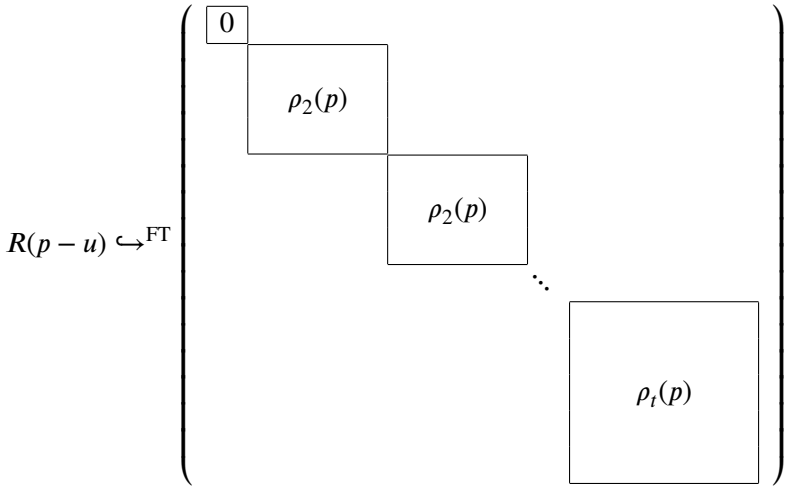
be the second largest eigenvalue. Then, $\text{Cay}(G, p_S)$ is called an ϵ-expander if $\lambda(p_S) \leq \epsilon$.

Remark 2.7.1. For the Bourgain-Gamburd result about the group $\mathsf{SL}_2(p)$, they get a 0.999 expander with high probability when a random set of two elements is chosen.

The quantity $1 - \lambda(p)$ is called the spectral gap and it is closely related to the notion of combinatorial expansion in graphs.

2.7.1 Convergence to the uniform distribution

For any probability distribution p, let us look at the function $p - u$, where u denotes the uniform distribution over the group G. We claim that the block diagonal matrix corresponding to $R(p-u)$, written in the Fourier basis, looks as follows:



In other words, the only difference between the block diagonal forms of $R(p-u)$ and that of $R(p)$ is in the top left hand corner, where in the former it is zero and in the latter we had 1. This follows from the fact that $R(p-u) = R(p) - R(u)$ and the following:

Proposition 2.7.2. *For any non-trivial irreducible representation ρ of G, we have*

$$\sum_{x\in G} \rho(x) = 0\,.$$

Proof. Let $h = \sum_{x\in G}\rho(x)$. Then, $\rho(y)h = h\rho(y) = h$ for all $y \in G$. Applying Schur's Lemma, we see that $h = \lambda\mathbb{1}$, where λ is some scalar and $\mathbb{1}$ is the identity matrix. Since ρ is non-trivial, there exists some $z \in G$ for which $\rho(z) \neq \mathbb{1}$. However, $h = \rho(z)h$ and, hence, $\lambda\mathbb{1} = \lambda\rho(z)$. Thus $\lambda = 0$ implying $h = 0$. □

This immediately shows $R(u)$ in its block diagonal form has a single non-zero entry at its top left hand corner and this is equal to 1. Thus, $R(p-u)$ has the block diagonal form claimed in the figure above.

Recall that the operator norm of a matrix A, denoted by $\|A\|$ is defined as $\max_{||v||=1} \|Av\|$. For symmetric matrices, this is equal to the maximum eigenvalue ignoring sign, i.e. $\max_i |\lambda_i(A)|$. Hence, noting the block diagonal form of $R(p-u)$ from above,

$$\|R(p-u)\| = \lambda(p)$$

where $\lambda(p)$ is the spectral gap of $R(p)$.

We define

$$p^{\ell} \equiv \underbrace{p * p * \cdots * p}_{\ell}.$$

Recall, for any function f,

$$R(f^{\ell}) \;=\; R(f)^{\ell}.$$

It is easy also to verify, using the fact $u * p \;=\; p * u \;=u$, that

$$(p^{\ell} - u) \;=\; (p-u)^{\ell}.$$

Hence, using (2.4) and (2.5), we obtain

$$p^{\ell} - u = R(p-u)^{\ell} \cdot \begin{pmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}.$$

Thus,

$$||p^{\ell} - u|| \leq ||R(p-u)^{\ell}|| = \lambda(p)^{\ell}. \tag{2.6}$$

2.8 Expanders

Theorem 2.8.1 (Alon-Roichman)**.** *Let $S = \{X_1, \ldots, X_k\}$, where each X_i is chosen independently and uniformly at random from the group G. Then,*

$$\mathbb{P}\left[\lambda(p_S) \;>\; \frac{1}{2}\right] \;\leq\; |G| \cdot \exp(-k).$$

Proof. For every element $x \in G$, define

$$R'(x) \equiv R\left(\frac{x + x^{-1}}{2} - u\right)$$

Then,

$$R(p_S - u) \;=\; \frac{1}{|S|} \sum_{x \in S} R'(x).$$

□

We note the following properties of $R'(x)$:

1. $||R'(x)|| \leq 1$. This is because $R(x)$ is a permutation matrix, by definition.
2. When we average, $\mathbb{E}_{x \in G} R'(x) = 0$. This is because, $\sum_{x \in G} R(x) = J$, where J is the all 1 matrix. Thus, $\mathbb{E}_{x \in G} R(x) = J/|G| = R(u)$ which implies the claim.
3. We are sampling $|S| = k$ such random matrices. Their average is zero and each of their norm is bounded by 1. We would like to argue that with very high probability, the norm of the average of k such random matrices is greater than half.
4. If these matrices were of dimension 1, then the above is true via the classical Chernoff-Hoeffding bound. What we are therefore looking for is a generalization of this inequality for matrix valued random variables.

Luckily, such a generalization does exist:

Theorem 2.8.2 (Alswhede-Winter Inequality [**?**])**.** *Let R be any real symmetric matrix-valued random variable satisfying the following: $||R|| \leq 1$ and $\mathbb{E}\left[R\right] = 0$. If $R_1, \ldots, R_k$ are independent random variables each identical to R, then*

$$\mathbb{P}\left[\left|\left|\frac{1}{k}\sum_{i=1}^{k} R_i\right|\right| \geq \gamma\right] \leq n \cdot \exp(-\gamma^2 k/2)\,.$$

The proof of the Alon-Roichman Theorem now immediately follows from the Alswhede-Winter Inequality.

Remark 2.8.3. The above can be derandomized via the method of conditional expectations: One can find in polynomial time $O(\log |G|)$ generators with respect to which the given group G expands.

Note that, unlike the Chernoff bound, there is an extra factor of n in the Alswhede-Winter Inequality. This is unavoidable in general as can be seen by considering R that is restricted to take values from diagonal matrices. The best one can do, in this case, is to take a union bound for n different sums of scalar valued random variables and bound each sum by the Chernoff bound. However, Alswhede-Winter can be perhaps strengthened in some interesting situations. For instance, Avi conjectures the following:

Conjecture 2.8.4. Let G be a group and ρ be any one of its irreducible non-trivial representations. Let $X_1, \ldots, X_k$ be k random elements chosen independently and uniformly from G. Then,

$$\mathbb{P}\left[\left|\left|\frac{1}{k}\sum_{i=1}^{k} \frac{\rho(X_i) + \rho(X_i^{-1})}{2}\right|\right| > \frac{1}{2}\right] \leq \exp(-\Omega(k)).$$

Indeed, it is not known if the above holds for the more general case, when the random matrices are selected uniformly at random from the set of unitary matrices. If such a strengthened inequality does exist, then one can apply it to each irreducible representation of G individually. This gives rise to a gain when the dimensions of the irreducible representation are large. However, when G is abelian this does not result in any improvement and in fact Avi gives the following exercise:

Exercise: If G is abelian and $\mathrm{Cay}(G, S)$ is an expander show that $|S| = \Omega(\log |G|)$.

2.8.1 Solvable groups

Recall that the commutator subgroup of G, denoted by $\left[G \ : \ G\right]$, is defined to be the subgroup generated by the commutators of G, i.e. $\left\langle \{aba^{-1}b^{-1} \ : \ a, b \in G\} \right\rangle$. If G is abelian then $\left[G \ : \ G\right] = 1$. It is not hard to verify that $\left[G \ : \ G\right]$ is a normal subgroup of G and that the quotient group $G/\left[G \ : \ G\right]$ is abelian. Furthermore, every normal subgroup having this property contains the commutator subgroup. Consider the following series:

$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_t$$

where $G_{i+1} = \left[G_i \ : \ G_i\right]$. We say G is ℓ-step solvable if $G_\ell = 1$. G is called solvable if there exists an ℓ such that it is ℓ-step solvable. In contrast, G is called simple if $\left[G \ : \ G\right] = G$, i.e. the group does not move at all. Hence, finite simple groups are finite and very non-abelian.

Theorem 2.8.5 (Lubotzky-Weiss)**.** *If G is ℓ-step solvable and* $\mathrm{Cay}(G, S)$ *is expanding then*

$$|S| \geq \underbrace{\log \log \cdots \cdots \log}_{\ell} |G| \, .$$

Theorem 2.8.6 (Meshulam-Wigderson)**.** *For every ℓ, there exists an ℓ-step solvable group G with an expanding generating set S such that*

$$|S| \leq \underbrace{\log \log \cdots \cdots \log}_{\ell} |G| \, .$$

The above shows that the lower bound of Lubotzky and Weiss on the size of the generating set can be tight for some solvable groups. However, it is known to be not tight in general. For example, consider the affine group below:

$$A_p \equiv_{\text{def}} \left\{ f \ : \ x \mapsto ax + b \ \ : \ \ a \neq 0, \, a, b \in \mathbb{F}_p \right\} .$$

The group operation is function composition. Then, $A_p = \mathbb{F}_p \rtimes \mathbb{F}_p^*$ is a semi-direct product of $\mathbb{F}_p$ and $\mathbb{F}_p^*$ and is 2-step solvable. However, it requires $\Omega(\log p)$ generators to expand.

2.8.2 Stories

Lubotzky shows the following corollary to Selberg's $\frac{3}{16}$th Theorem: consider the following set S of generators for $\mathsf{SL}_2(p)$:

$$S = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\}.$$

The set S also generates $\mathsf{SL}_2(\mathbb{Z})$. Selberg shows that $\mathsf{SL}_2(\mathbb{Z})$is expanding with respect to S. Consequently, the expansion of all Cayley graphs $\mathrm{Cay}(\mathsf{SL}_2(p), S)$ for all p, follows from the expansion of one "mother group."

The first group known to be expanding was $\mathsf{SL}_n(p)$ for $n \geq 3$. In a recent, capstone result, all nonabelian finite simple groups were shown to admit a constant number of generators that yield a expander Cayley graph.

Theorem 2.8.7 (Nikolov and Kassabov)**.** *Every finite simple group is expanding with a constant number of generators.*

Although we have talked about expansion in Cayley graphs, the arguments extend easily to a more general class of graphs called *Schreier* graphs. A group G acts on a set Ω if every element $x \in G$ is associated with a map $\Pi_x : \Omega \to \Omega$ so that for each $x, y \in G$, $\Pi_x \circ \Pi_y = \Pi_{xy}$ and, furthermore, Π_e is the identity map when e is the group identity. It is easy to verify then that each Π_x is a bijection as it is invertible. Let $\Pi = \{\Pi_x \mid x \in G\}$. Note that G acts on itself in an obvious way. Just as this action of G gives rise to the regular representation, any action of G on Ω gives rise to a $|\Omega|$-dimensional representation of G. Abusing notation, let us denote this representation Π. Given any subset $S \subseteq G$, we denote by $\mathrm{Sch}(G, \Pi, \Omega, S)$ the Schreier graph whose set of vertices is Ω and whose edges are all pairs $u, v \in \Omega$ for which there is an $s \in S$ so that $\Pi_s u = v$. As before, this graph is not directed as S is closed under inverse.

It is simple to verify that the transition matrix of the random walk on $\mathrm{Sch}(G, \Pi, \Omega, S)$ is equal to $\sum_{x \in G} \Pi(x) p_S(x)$. Just as before for Cayley graphs, one can apply the Fourier transform to convert this into block diagonal form where each block belongs to a copy of an irreducible representation ρ_i of G, i.e. $\sum_{x \in G} \rho_i(x) p_S(x)$. Hence, one gets the following immediately:

Theorem 2.8.8. *If* $\mathrm{Cay}(G, S)$ *is an expander, then* $\mathrm{Sch}(G, \Pi, \Omega, S)$ *is also an expander.*

Though the above is a simple result, it is quite useful. For example, using it we prove next that the following graph is an expander: the set of vertices of the graph is $\mathbb{F}_p$, where p is any prime. Every $x \in \mathbb{F}_p^*$ is connected to $x + 1$, $x - 1$ and $-1/x$. Finally, the vertex 0 is connected to 1 and -1.

We do this in two steps. First, consider the natural action of $\mathsf{SL}_2(\mathbb{F}_p)$ on the affine space $\mathbb{F}_p^2$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

This induces an action on the set of projective lines $\mathbb{P}^2(\mathbb{F}_p)$. Recall that this set has p finite lines that we identify with $\mathbb{F}_p$ and a single point at infinity. The action on this set of finite lines by elements of the generating set S can therefore be written as follows:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} x \to x + 1; \quad \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} x \to x - 1; \quad \text{and} \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} x \to \frac{-1}{x}.$$

Thus, the element 0, under this action, maps to 1, -1 and the point at infinity. It is easy to verify that the point at infinity maps to itself by two of the elements of S and to the point 0 by the other element of S. Thus, the Schreier graph of this action on projective lines by S is a graph on $p + 1$ vertices, where we view the first p vertices as points from $\mathbb{F}_p$ and the $p + 1$th vertex is the point at infinity. This Schreier graph is an expander as $\mathrm{Cay}(\mathsf{SL}_2(p), S)$ is an expander. If we remove the $p + 1$th vertex from this graph and delete the three edges incident to it, we get exactly our desired graph on $\mathbb{F}_p$. Thus, the desired graph is also an expander.

LECTURE 8

Lecturer: Avi Wigderson *Scribe: Valentine Kabanets and Antonina Kolokolova*

First, let us clarify some things from the previous lecture.

- **Mixing time in terms of the second largest eigenvalue:** We want to argue that $\|p^\ell - u\| \leq \lambda(p)^\ell$. To see this, observe that for any function f, the matrix $R(f)$ contains (the truth table of) f as its first column. Hence,

$$p^\ell - u = R(p^\ell - u) \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

  Taking the norms on both sides, we have $\|p^\ell - u\| \leq \|R(p^\ell - u)\| \leq \lambda(p)^\ell$.

- **Schreier graph:** Let G be a group and $S \subseteq G$. Let π be the G-action on some set Ω, i.e., $\pi : G \to \mathfrak{S}_\Omega$ is a homomorphism of G in to the group of permutations of the set Ω. By treating each permutation as its associated permutation matrix, we may extend this to a homomorphism $\pi : G \to \mathsf{GL}_{|\Omega|}(\mathbb{C})$. Let $\mathrm{Irrep}(G) = \{\rho_i, \ldots, \rho_t\}$ be the set of all irreducible representations of G. Then $\pi = \bigoplus_{i=1}^t m_i \rho_i$. Thus, $\mathrm{spec}\{\mathrm{Sch}(G, \pi, \Omega; S)\} \leq \mathrm{spec}\{\mathrm{Cay}(G; S)\}$ (where spec denotes the spectrum of the matrix); we will see the same matrices $\rho_i(P_s)$ in both cases, but possibly with different multiplicities, thus the same eigenvalues appear, perhaps with different multiplicities. In particular, if $\mathrm{Cay}(G; S)$ is an expander, then so is $\mathrm{Sch}(G, \pi, \Omega; S)$. Note: If S is a generating set, the trivial representation occurs once and the graph is connected; otherwise, we get a disconnected graph.

2.9 Fast matrix multiplication

Let A, B be $k \times k$ matrices over $\mathbb{C}$. We want to compute $A \cdot B = C$. Let ω be the least real number such that there exists a matrix multiplication algorithm that uses only $n^{\omega+\epsilon}$ multiplications for every $\epsilon > 0$. (Note: It is always possible to make the number of additions no more than the number of multiplications plus a constant; thus it's sufficient to bound the number of multiplications only.)

Trivially, $\omega \leq 3$. Strassen [**?**] showed that $\omega \leq \log_2 7$. The currently best bound is due to Coppersmith and Winograd [**?**]: $\omega \leq 2.38...$. In 2003, Cohn and Umans [6] suggested a new approach, based on group representation theory. The initial algorithm wasn't even better than k^3 but, in later work, Cohn, Kleinberg, Szegedy, and Umans [**?**] improved it to $\omega \leq 2.38...$, matching the Coppersmith-Winograd [**?**] bound.

The Cohn-Umans approach relies on finding groups with specific properties. Let G be a group, and let $H_1, H_2, H_3 \leq G$ be three subgroups. The group G and subgroups H_1, H_2, H_3 satisfy the *triple-product property* if the following holds:

$$\forall h_1 \in H_1, h_2 \in H_2, h_3 \in H_3 \qquad h_1 h_2 h_3 = 1 \Rightarrow h_1 = h_2 = h_3 = 1\,.$$

Assume, without loss of generality, that $|H_i| = k$. We treat the $k \times k$ matrices A and B as being indexed by the elements in $H_1 \times H_2$ and $H_2 \times H_3$, respectively. The idea is to think of A, B as elements a, b of the group algebra. Then the convolution $a * b$ will correspond to the matrix multiplication $A \cdot B$ if we have the triple property.

More precisely, let $\mathbb{C}[G]$ be the group algebra over $\mathbb{C}$. Associate with the matrix $A_{H_1 \times H_2}$ the element $a = \sum A_{h_1 h_2}(h_1 h_2^{-1})$ of $\mathbb{C}[G]$. Similarly, associate $B_{H_2 \times H_3}$ with $b = \sum B_{h_2 h_3}(h_2 h_3)^{-1}$. Let $C = A \cdot B$. We will show that the convolution $a * b(h_1 h_3^{-1}) = C_{h_1 h_3}$. We have:

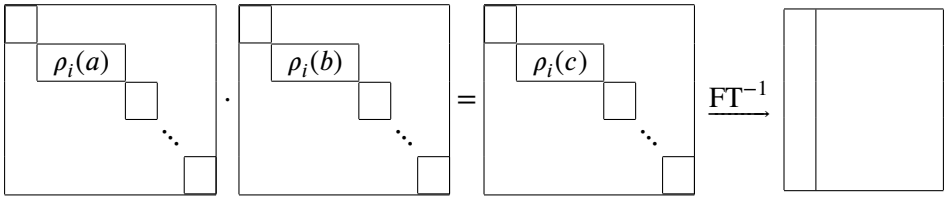
$$a * b(h_1 h_3^{-1}) = \Sigma_{\hat{h}_1, h_2, \hat{h}_2, \hat{h}_3} A_{\hat{h}_1 h_2} B_{\hat{h}_2 \hat{h}_3}\,, \tag{2.7}$$

where the summation is over those indices satisfying the condition $\hat{h}_1 h_2^{-1} \hat{h}_2 \hat{h}_3 = h_1 h_3^{-1}$. The latter condition can be equivalently written as

$$\underbrace{(h_1^{-1}\hat{h}_1)}_{H_1}\underbrace{(h_2^{-1}\hat{h}_2)}_{H_2}\underbrace{(\hat{h_3}^{-1} h_3)}_{H_3} = 1\,,$$

which implies, by the triple property, that $h_1 = \hat{h}_1$, $h_2 = \hat{h}_2$, $h_3 = \hat{h}_3$. Thus, the right-hand side of equation (2.7) is equal to $\sum_{h_2} A_{h_1 h_2} B_{h_2 h_3} = C_{h_1 h_3}$, as required.

We know that $R(a * b) = R(a)R(b)$. To compute the matrix product $R(a)R(b)$, we first apply the Fourier transform to both $R(a)$ and $R(b)$, resulting in block-diagonal matrices; then we multiply these block-diagonal matrices; finally, we apply the inverse Fourier transform to recover $R(a * b)$:



We may carry out the Fourier transform once, in advance, for a constant size (depending on k) matrix. Then this strategy can be applied recursively.

Cost of this algorithm: To compute the matrix products $\rho_i(a) \cdot \rho_i(b) = \rho_i(c)$ over all t irreducible representation $\rho_1, \ldots, \rho_t$, we need $\Sigma_{i=1}^t d_i^\omega \le n \cdot d_{\max}^{\omega-2}$ (where $n = |G|$ and we used the fact that $\sum_{i=1}^t d_i^2 = n$). Note that $k^\omega \le \sum_{i=1}^t d_i^\omega$. If there is some α such that $k^\alpha > \Sigma d_i^\alpha$ then $\omega \le \alpha$. If there were a subgroup of size $k > \sqrt{n}$, then we would get $\omega = 2$, but getting subgroups of that size is impossible. Thus we wish to find a group with subgroups satisfying the triple-product property that are as large as possible. In the original paper, there are interesting examples with k close to $\sqrt{n}$. In their examples, including $k = 2.38...$, the dimensions d_i are small: 1 or 2. If a group is abelian, this approach cannot beat $\omega = 3$.

2.10 The Fourier transform over general groups

Suppose we have a group G, with $\mathrm{Irrep}(G) = \{\rho_1, \ldots, \rho_t\}$ of dimensions $d_1, \ldots, d_t$. Given a function $f : G \to \mathbb{F}$, we define its *Fourier transform* at ρ_i as

$$\hat{f}(\rho_i) = \mathbb{E}_x\, f(x)\rho_i(x)\,;$$

observe that this is matrix-valued. The *inverse Fourier transform* is given by the following formula:

$$f(y) = \sum_{i=1}^{t} d_i \,\mathrm{tr}[\hat{f}(\rho_i)\rho_i(y)^{-1}]\,.$$

Let us prove this formula.

Proof. Expanding the right-hand side and using the linearity of expectation, we have

$$\sum_{i=1}^{t} d_i \operatorname{tr}[\mathbb{E}_x f(x)\rho_i(xy^{-1})] = \mathbb{E}_x \sum_{i=1}^{t} d_i \operatorname{tr}[f(x)\rho_i(xy^{-1})] = \mathbb{E}_x f(x) \sum_{i=1}^{t} d_i \operatorname{tr}[\rho_i(xy^{-1})]\,. \tag{2.8}$$

Observe that

$$\sum_{i=1}^{t} d_i \operatorname{tr}[\rho_i(xy^{-1})] = \operatorname{tr}[R(xy^{-1})] = \begin{cases} n & xy^{-1} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the right hand-side of equation (2.8) is equal to $(1/n)nf(y) = f(y)$. □

We also have the *Parseval identity*:

$$n\|f\|_2^2 = \operatorname{tr}[R(f)^t R(f)]\,,$$

where $\|f\|^2 = \sum_x |f(x)|^2$, and t denotes the transpose.

2.11 Using the multiplicity of eigenvalues: Gowers' trick

For a group G with irreps ρ_i's of dimensions d_i's, let $m(G) = \min d_i$, for $i \neq 1$. Note that the second largest eigenvalue of the regular representation of G has multiplicity at least $m(G)$. This can be used to get nontrivial upper bounds on the size of subgroups of G, to bound mixing time in G, etc.

The first use of this idea is due to Sarnak and Xue [**?**] (see also the book by Davidoff, Sarnak, and Valette [**?**]). It was also used by Bourgain and Gamburd [**?**, **?**] and other papers with different motivations, in particular, by Gowers [**?**] and by Babai, Nikolov, and Pyber [**?**].

We will prove the following.

Theorem 2.11.1. *Let p and q be arbitrary probability distributions on G, and let u be the uniform distribution. Then*

$$\|p * q - u\|^2 \leq \frac{n}{m(G)} \|p - u\|^2 \cdot \|q - u\|^2\,.$$

Remark 2.11.2. In abelian groups, $m(G) = 1$. In this case, can use Cauchy-Schwartz for the proof. (Exercise!)

Gowers [**?**] was interested in the following question for $\mathsf{SL}_2(p)$. Suppose X, Y, Z are subsets of the group. Can one infer the existence of solutions to the equation $xy = z$, for $x \in X$, $y \in Y$, and $z \in Z$, from the sizes of the sets? The answer is yes: If $|X| = |Y| = |Z| = n^{.99}$, then $|XY| > n - n^{.99}$, and hence a solution exists.

Here are some examples of groups G and the corresponding $m(G)$.

	$m(G)$	$\lvert G\rvert$
G abelian	$m(G) = 1$	
$G = \mathrm{Alt}_k$ (alternating)	$m(G) = k - 1$	$k!$
$G = \mathsf{SL}_2(p)$	p	p^3

Figure 2.1: Examples of groups G and values $m(G)$.

Quasirandom groups are those groups where $m(G)$ is large. One can use Theorem 2.11.1 to get upper bounds on the size of subgroups of a given group. Suppose $H \leq G$, take $X = Y = H$. Then take p and q to be uniform on H. Using

$$\|p * q - u\|^2 \leq \frac{n}{m(G)} \|p - u\|^2 \cdot \|q - u\|^2 ,$$

one easily gets

$$\frac{1}{|H|} \leq \frac{n}{m(G)} \cdot \frac{1}{|H|} \cdot \frac{1}{|H|}$$

so $|H| \leq \frac{n}{m(G)}$. In particular, this shows that quasirandom groups cannot have large subgroups.

Proof of Theorem 2.11.1. What we want to prove can equivalently be written as follows:

$$n\|p * q - u\|^2 \leq \frac{1}{m(G)} (n\|p - u\|^2) \cdot (n\|q - u\|^2) . \tag{2.9}$$

Denote $p' = p - u$ and $q' = q - u$. Observe that $(p - u) * (q - u) = p * q - u$. Using Parseval's identity (see the previous section) and the fact that $R(f * g) = R(f)R(g)$, we can write the left-hand side of equation (2.9) as follows:

$$\begin{aligned} \operatorname{tr}[(R(p')R(q'))^t (R(p')R(q'))] &= \operatorname{tr}[R(q')^t R(p')^t R(p')R(q')] \\ &= \operatorname{tr}[(R(p')^t R(p'))(R(q')^t R(q'))] , \end{aligned}$$

where for the second equality we used the fact that $\operatorname{tr}[AB] = \operatorname{tr}[BA]$ for any matrices A and B. We will also need the following

Fact 2.11.3. If A and B are positive semidefinite, then $\operatorname{tr}(AB) \leq \|A\| \cdot \operatorname{tr}(B)$.

In our case, we have matrices of the form $A^t A$, which are positive semidefinite, and so, applying the above fact, we can continue

$$\begin{aligned} \operatorname{tr}[R(p')^t R(p')R(q')^t R(q')] &\leq \|R(p')^t R(p')\| \cdot \operatorname{tr}[R(q')^t R(q')] \\ &\leq \frac{1}{m(G)} \operatorname{tr}[R(p')^t R(p')] \cdot \operatorname{tr}[R(q')^t R(q')] , \end{aligned}$$

where for the last inequality we used the observation that each nonzero eigenvalue of $R(p')$ has multiplicity at least $m(G)$ (by the definition of $R(p')$ and $m(G)$), and hence,

$$\operatorname{tr}[R(p')^t R(p')] \geq m(G) \cdot \|R(p')^t R(p')\| .$$

Applying Parseval's identity, we get the right-hand side of equation (2.9), as required. □

Exercise: If p is a class function, this multiplicity can be improved to $m(G)^2$.

LECTURE 10

Lecturer: Avi Wigderson *Scribe: Pavel Pudlák*

2.12 Lubotzky's 1-2-3 question

In 1979 Margulis gave the first explicit definition of expanders. The motivation came from Pinsker's research into error-correcting codes. A few years before Kazhdan had defined *Property (T)*. Lubotzky then observed that a deep result of Selberg concerning $\mathsf{SL}_2(\mathbb{Z})$ implies that the infinite group $\mathsf{SL}_2(\mathbb{Z})$ is expanding: it follows that if $\mathsf{SL}_2(\mathbb{Z})$ is expanding with a set S, then, for all p, $\mathrm{Cay}(\mathsf{SL}_2(p), S)$ is expanding.

Question 2.12.1 (Lubotzky's 1-2-3 question)**.** Denote by S_a the set

$$\left\{ \begin{pmatrix} 1 & \pm a \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \pm a & 1 \end{pmatrix} \right\}.$$

1. Then $\langle S_1 \rangle = \mathsf{SL}_2(\mathbb{Z})$ and hence can be used to generate expanders.
2. The set $\langle S_2 \rangle$ generates a subgroup of *finite index* in $\mathsf{SL}_2(\mathbb{Z})$ and hence also suffices for expansion.
3. $\langle S_3 \rangle$ generates a subgroup of *infinite index*. Does it give expanders on $\mathsf{SL}_2(p)$?[4]

Theorem 2.12.2 (Bourgain-Gamburd [**?**])**.** *If $x, y \in SL_2(p)$ such that the girth of* $\mathrm{Cay}(SL_2(p), \{x, y\})$ *is no more than $\epsilon \log |SL_2(p)|$, then* $\mathrm{Cay}(SL_2(p), \{x, y\})$ *is an expander.*

Remark 2.12.3. The set S_a generates a free subgroup of $\mathsf{SL}_2(\mathbb{Z})$ for $a \geq 2$; it follows that the graph $\mathrm{Cay}(\mathsf{SL}_2(p), S_a)$ satisfies the condition of the theorem. In particular, the theorem answers Lubotzky's question. Moreover, random elements $\{x, y\}$ suffice!

To prove that S_a generates a free group is not difficult.To see that the resulting graph has large girth, observe that a nontrivial product of matrices from S_a may be 0 only if it has large enough elements to be divisible by p. It follows that the product has to be of length $\Omega(\log p)$. The result for random elements uses the same argument, except that one has to combine it with the Schwartz-Zippel lemma. We outline the proof below.

Proof sketch of the Bourgain-Gamburd theorem. Let P be the distribution given by the set $\{x, x^{-1}, y, y^{-1}\}$. By Parseval's equality (Lemma 2.4.4)

$$n\|P^\ell - U\|_2^2 = \mathrm{tr}[R(P^\ell - U)^t R(P^\ell - U)] = \sum_{j=2}^{n} \lambda_j(P)^{2\ell}.$$

Thus an upper bound on $n\|P^\ell - U\|_2^2$ yields an upper bound on $\lambda_2(P)$, the second eigenvalue of the Caley graph. Namely, if we could upper-bound it by $1/n^\epsilon$, $\epsilon > 0$ for $\ell = O(\log n)$, we would get $\lambda(P) < c < 1$. For this we would need $\|P^\ell\|_2^2 \leq 1/n + 1/n^{1+\epsilon}$. But this seems too difficult. Therefore Bourgain and Gamburd

[4]Note that $\begin{pmatrix} 1 & -a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & +a \\ 0 & 1 \end{pmatrix}^{-1}$.

use an observation of Sarnak and Xue [**?**] (appearing also in Davidoff-Sarnak-Valette [**?**]) that $\lambda(P)$ occurs with multiplicity m, roughly $n^{1/3}$. Then we have

$$m\lambda(P)^{2\ell} \le \sum_{j=2}^{n} \lambda_j(P)^{2\ell}\,.$$

Thus it suffices to prove a bound

$$\|P^\ell - U\|_2^2 \le \frac{1}{n^{1-\epsilon}}\,, \tag{2.10}$$

for ϵ such that $m \ge n^{2\epsilon}$ (so $\epsilon \approx 1/6$). This will give us

$$\lambda(P)^{2\ell} \le \frac{n}{mn^{1-\epsilon}} = \frac{1}{n^\epsilon}\,,$$

whence $\lambda(P) < c < 1$.

First we explain, using group representation theory, why the multiplicity is high.

$$n\|P^\ell - U\|_2^2 = \mathrm{tr}[R(P^\ell - U)^t R(P^\ell - U)] = \sum_{i=1}^{t} d_i\,\mathrm{tr}[\rho(P^\ell - U)^t \rho(P^\ell - U)]\,,$$

where the sum is over all irreducible representations ρ_i of the group and d_i are their multiplicities. The last equality follows from the fact that $n\|P^\ell - U\|_2^2$ is invariant under the group actions. If we know that $d_i \ge m$ for all $i > 1$, then the multiplicity of $\lambda(P)$ is at least m.

To compute the upper bound on $\|P^\ell - U\|_2^2$, we first use the assumption about the girth of the graph $\mathrm{Cay}(\mathsf{SL}_2(p), \{x, y\})$. This gives (computation omitted)

$$\|P^\ell\|_2^2 \le \frac{1}{n^\delta}\,,$$

for some $\delta > 0$ and $\ell_0 = O(\log|\mathsf{SL}_2(p)|)$. To get the bound needed in (2.10), we apply Corollary 1.11.2 of Helfgott's theorem. According to that result, for some $\ell = O(\ell_0)$, we have either

$$\|P^\ell\|_2^2 \le \frac{1}{n^{1-\delta/2}}$$

or P is concentrated on some proper subgroup H ($P(H) \ge 1/n^{1-\delta'}$). Proving that the second possibility cannot take place requires an argument that is omitted. □

Example 2.12.4. Here is another example where one needs a better analysis of the convergence to uniform distribution. Diaconis and Shashiani [8] studied the problem of "shuffling cards." Given a deck of k cards, we randomly pick two consecutive cards and switch them. This can be represented as a random walk on $\mathrm{Cay}(\mathfrak{S}_k, \{(1,2),(2,3),\dots,(k-1,k)\})$. Observe that

$$\lambda(Q) = 1 - \frac{1}{k}\,,$$

where Q denotes the uniform distribution on the permutations $\{(1,2),(2,3),\dots,(k-1,k)\}$. Using a trivial analysis based on the leading eigenvalue, one can show that $O(k^2 \log k)$ suffice to get close to the uniform distribution on all $k!$ permutations. Diaconis and Shashiani proved that $O(k \log k)$ steps in fact suffice to get uniform mixing.

2.13 Kazhdan's constant

Definition 2.13.1. Let S be a generating subset of a group G. Let ρ be a unitary representation of G that does not contain the trivial representation.[5] If for $\epsilon > 0$,

$$\forall \vec{v} \neq \vec{0}, \exists x \in S \quad \|\rho(x)\vec{v} - \vec{v}\| \geq \epsilon \|\vec{v}\| ,$$

then the *Kazhdan constant* $K_G(S) \geq \epsilon$.

Suppose $K_G(S) \geq \epsilon > 0$. Then $\mathrm{Cay}(G, S)$ is a combinatorial expander. More precisely

Proposition 2.13.2. *If $K_G(S) \geq \epsilon$, then for all $T \subseteq G$ of size at least $n/2$,*

$$\exists x \in S \quad |Tx \triangle T| \geq \frac{\epsilon}{\sqrt{2}} |T| .$$

Sketch of proof. The inequality is equivalent to

$$\|R(x)\vec{1}_T - \vec{1}_T\| \geq \frac{\epsilon}{\sqrt{2}} \|\vec{1}_T\| .$$

We would be done if R did not contain the trivial representation. If we restricted our attention to vectors $\vec{v} \perp \vec{1}$, the stement would similarly follow. Write $\vec{1}_T = \alpha\vec{1} + \vec{v}$, with $\vec{v} \perp \vec{1}$. Thus $\vec{v}$ is the part of $\vec{1}_T$ that projects on the subspace of the nontrivial irreducible representations. Since $T \leq n/2$, $\|\vec{v}\| \geq \|\vec{1}_T\|/\sqrt{2}$. We can replace $\vec{1}_T$ by $\vec{v}$ (loosing the factor of $1/\sqrt{2}$), because $R(x)\vec{1} = \vec{1}$. □

Note that for $|S| = O(1)$, $K_G(S)$ and $1 - \lambda(P_S)$ are related upto squares.

2.14 Dimension expanders

Definition 2.14.1. A dimension expander on $\mathbb{F}^d$ is a family of linear operators $T_1, \ldots, T_k : \mathbb{F}^d \to \mathbb{F}^d$ such that for every subspace $V \subseteq \mathbb{F}^d$ of dimension less than $d/2$,

$$\exists i \ \dim(T_i V \cap V) \leq (1 - \epsilon) \dim V .$$

Note that a random set of constant size of linear operators is a dimension expander.

Conjecture 2.14.2 (Wigderson)**.** If $\mathrm{Cay}(G; S)$ is expanding and ρ is an irreducible representation of G, then $\{\rho(x) \ : \ x \in S\}$ is a dimension expander.

Theorem 2.14.3 (Lubotzky-Zelmanov [14])**.** *The conjecture is true for* $\mathbb{C}$.

Explicit dimension expanders of constant degree over every field have been constructed by Bourgain (see the Dvir-Wigderson paper [**?**]).

[5] This means that the decomposition of ρ into irreducible representations does not contain the trivial representation. Equivalently, no nozero vector is fixed by all the group actions.

2.15 More on expanders

Theorem 2.15.1 (Lubotzky-Weiss [13]). *If G is ℓ-step solvable and* $\mathrm{Cay}(G;S)$ *is an expander, then* $|S| \geq \log^{(\ell)}|G|$.

Theorem 2.15.2 (Meshulam-Wigderson [15]). *There exists G and S such that G is ℓ-step solvable and* $\mathrm{Cay}(G;S)$ *is an expander and* $|S| \leq \log^{(\ell/2)}|G|$.

Lubotzky and Weiss asked if expansion is a group property. Namely: does the the fact that $\mathrm{Cay}(G;S)$ is an expander depend only on G, not on the choice of S?[6]

This question was answered negatively by Alon, Lubotzky and Wigderson. They used the semidirect product of groups and the fact that the Caley graphs of semidirect products are essentially the zig-zag products of the Caley graphs of the components. (The zig-zag product was introduced by Reingold, Vadhan and Wigderson [**?**].)

Outline of the proof of the Meshulam-Wigderson Theorem. Construct groups

$$G_1, \ldots, G_\ell, \ldots \quad \text{and sets} \quad S_1, \ldots, S_\ell, \ldots$$

as follows:

$$G_{\ell+1} = G_\ell \ltimes \mathbb{F}_{p_\ell}[G_\ell]\,.$$

I.e., take the semidirect product with the additive group of the group algebra $\mathbb{F}_{P_\ell}[G_\ell]$. (The semidirect product is defined by $(g,v)(h,u) = (gh, v^h + u)$, where v^h is v with the coordinates permuted by h.) Here p_ℓ are distinct primes. (We need to ensure that p_ℓ does not divide $|G_\ell|$.) The orders of the groups $n_\ell = |G_\ell|$ satisfy the recursion $n_{l+1} = n_\ell p_\ell^{n_\ell}$; for the sets we will have $|S_{\ell+1}| \leq |S_\ell|^{10}$.

The set $S_{\ell+1}$ is constructed by taking the union of the orbits (with respect to the action of the group G_ℓ on $\mathbb{F}_{p_\ell}[G_\ell]$) of a certain number randomly chosen set of elements of $\mathbb{F}_{p_\ell}[G_\ell]$.

A crucial step in the analysis of this construction is an estimate on the number of irreducible representations of small dimensions of $\mathbb{F}[G]$. Let

$$\mathbb{F}[G] = \bigoplus_{i=1}^{t} M_{d_i}(\mathbb{F})$$

and define

$$\beta(d) = |\{i \ : \ d_i \leq d\}|\,.$$

What we need is an exponential upper bound on $\beta(d)$. A general upper bound was proved by de la Harpe, Robertson and Vallete:

Theorem 2.15.3. *If $k_G(S) \geq \epsilon$, then for all d, $\beta(d) \leq \exp(d^2)$.*

Because this is not enough they prove a better bound for a class of groups:

Theorem 2.15.4. *If G is a monomial group and $k_G(S) \geq \epsilon$, then for all d, $\beta(d) \leq \exp(d)$.*

A group is monomial if all irreducible representation consist of permutation matrices. They conjecture, however, that their bound holds true in general. Concerning monomial groups they prove:

Theorem 2.15.5. *If G is monomial, then $G \ltimes \mathbb{F}[G]$ is monomial too.*

Thus they can start with any abelian group G_1. □

[6] for sets of a given size

Bibliography

[1] Noga Alon and Yuval Roichman. Random Cayley graphs and expanders. *Random Structures Algorithms*, 5:271–284, 1997. 31

[2] L. Babai and P. Erdős. Representation of group elements as short products. *Annals of Discrete Math.*, 12:27–30, 1982. 13

[3] László Babai and Lajos Rónyai. Computing irreducible representations of finite groups. *Mathematics of Computation*, 55(192):705–722, 1990. 33

[4] Jean Bourgain, Nets Katz, and Terence Tao. A sum-product estimate in finite fields, and applications. *Geometric and Functional Analysis*, 14(1):27–57, February 2004. 29

[5] Emmanuel Breuillard, Ben Green, and Terence Tao. Linear approximate groups. Technical Report arxiv:1001.4570, arXiv electronic preprint, 2010. preprint. 26

[6] Henry Cohn and Christopher Umans. A group-theoretic approach to fast matrix multiplication. In *FOCS '03: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, page 438, Washington, DC, USA, 2003. IEEE Computer Society. 31, 52

[7] G. Cooperman. Towards a practical, theoretically sound algorithm for random generation in a finite group. Technical Report arXiv:math/0205203, arXiv electronic preprint, 2002. 11

[8] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Probability Theory and Related Fields*, 57(2):159–179, 1981. 31, 57

[9] Alex Eskin, Shahar Mozes, and Hee Oh. On uniform exponential growth for linear groups. *Inventiones Mathematicae*, 160(1):1–30, April 2005. 28

[10] John J. F. Fournier. Sharpness in Young's inequality for convolution. *Pacific Journal of Mathematics*, 72(2):383–397, 1977. 8, 13

[11] William Fulton and Joe Harris. *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer, 1991. 44

[12] M. Larsen and R. Pink. Finite subgroups of algebraic groups. preprint, 1995. 27

[13] Alexander Lubotzky and Barak Weiss. *Expanding Graphs – Groups and expanders*. DIMACS: Series in Discrete Mathematics and Theoretical Computer Science, 1993. 31, 59

[14] Alexander Lubotzky and Efim Zelmanov. Dimension expanders. *Journal of Algebra*, 319:730–738, 2008. 31, 58

[15] Roy Meshulam and Avi Wigderson. Expanders in group algebras. *Combinatorica*, 24(4):659–680, 2004. 31, 59

[16] László Pyber and Endre Szabó. Growth in finite simple groups of Lie type. Technical Report arxiv:1001.4556, arXiv electronic preprint, 2010. 26

[17] Jean-Pierre Serre. *Linear Representations of Finite Groups*. Number 42 in Graduate Texts in Mathematics. Springer, 1977. 44