QUASIRANDOMNESS AND GOWERS' THEOREM

QIAN ZHANG

August 16, 2007

Abstract. "Quasirandomness" will be described and the Quasirandomness Theorem will be used to prove Gowers' Theorem. This article assumes some familiarity with linear algebra and elementary probability theory.

Contents

1. Lindsey's Lemma: An Illustration of quasirandomness

Definition 1.1. A is a Hadamard matrix of size n if it is an $n \times n$ matrix with each entry ($\mathrm{a_{ij}}$) either +1 or -1. Moreover, its rows are orthogonal, i.e. any two row vectors have inner product = 0.

Remark 1.2. If A is a $n \times n$ matrix with orthogonal rows, then $AA^T = nI$, so $(\frac{1}{\sqrt{n}}A)(\frac{1}{\sqrt{n}}A)^T = I \Leftrightarrow (\frac{1}{\sqrt{n}}A)^T = (\frac{1}{\sqrt{n}}A)^{-1} \Leftrightarrow A^TA = nI$, so the colums of A are also orthogonal.

Notation 1.3. $\vec{1}$ denotes the column vector that has 1 as every component and $\vec{1}^T$ denotes the row vector with 1 as every component.

Remark 1.4. For any matrix A, $\vec{1}^T A \vec{1}$ is the sum of the entries of A. [1]

Definition 1.5. Given a matrix A and a submatrix T of A, let X be the set of rows of T and let Y be the set of columns of T. Let x_i be any component of $\vec{x}$ and let y_i be any component of $\vec{y}$. $\vec{x}$ is an **incidence vector** of X when $x_i = 1$ if the ith row vector of A is a row vector of T and $x_i = 0$ otherwise. $\vec{y}$ is an incidence

[1] $\vec{1}$ selects columns of A and sums their corresponding components. $\vec{1}^T$ selects rows of A, selecting and adding together certain component sums. Replacing the ith component of $\vec{1}$ with a 0 would deselect the ith column of A and replacing the ith component of $\vec{1}^T$ with 0 would deselect the ith row of A.

vector of Y when $y_i = 1$ if the ith column vector of A is a column vector of T and $y_i = 0$ otherwise.

Lemma 1.6. *(Lindsey's Lemma) If $A = (a_{ij})$ is an $n\times n$ Hadamard matrix and T is a $k\times l$-submatrix, then $|\sum_{(i,j)\in T} a_{ij}| \leq \sqrt{kln}$*

Proof. Let X be the set of rows of T and let Y be the set of columns of T. Note that $|X| = k$ and $|Y| = l$. Let $\overrightarrow{x}$ be the incidence vector of X and let $\vec{y}$ be the incidence vector of Y.

$\vec{x}^T A\vec{y}$ is $\sum_{(i,j)\epsilon T} a_{ij}$, which is the sum of all entries of T, so $\left|\vec{x}^T A\vec{y}\right| = \left|\sum_{(i,j)\epsilon T} a_{ij}\right|$. By the Cauchy-Schwarz inequality, $\left|\vec{x}^T A\vec{y}\right| \leq \|\vec{x}\| \, \|A\vec{y}\|$.

$$\left\|\tfrac{1}{\sqrt{n}}A\vec{y}\right\| = \vec{y}^T(\tfrac{1}{\sqrt{n}}A)^T(\tfrac{1}{\sqrt{n}}A)\vec{y} = \vec{y}^T\vec{y} = \|\vec{y}\|$$
$$\|A\vec{y}\| = \left\|\sqrt{n}(\tfrac{1}{\sqrt{n}}A\vec{y})\right\| = \sqrt{n}\left\|\tfrac{1}{\sqrt{n}}A\vec{y}\right\| = \sqrt{n}\,\|\vec{y}\|$$

Substituting for $\|A\vec{y}\|$ in the Cauchy-Schwarz inequality and noting that $\|\vec{x}\| = \sqrt{k}$ and $\|\vec{y}\| = \sqrt{l}$, $\left|\vec{x}^T A\vec{y}\right| \leq \|\vec{x}\|\,(\sqrt{n}\,\|\vec{y}\|) = \sqrt{kln}$. □

1.1. **How Lindsey's Lemma is a Quasirandomness result.** The following corollary illustrates how Lindsey's Lemma is a "quasirandomness" result. It says that if T is a sufficiently large submatrix, then the number of +1's and the number of -1's in T are about equal.

Corollary 1.7. *Let T be a $k\times l$ submatrix of an $n\times n$ Hadamard matrix A. If $kl \geq 100n$, then the number of +1's and the number of -1's each occupy at least 45% and at most 55% of the cells of T.*

Proof. Let x be the number of +1's in T and let y be the number of -1's in T. Suppose $kl \geq 100n$. We want to show that $(0.45)kl \leq x \leq (0.55)kl$ and $(0.45)kl \leq y \leq (0.55)kl$.

By Lindsey's Lemma, $|\sum_{(i,j)\epsilon T} a_{ij}| \leq \sqrt{kln}$. Note that $x - y = \sum_{(i,j)\epsilon T} a_{ij}$, so $\left|\sum_{(i,j)\epsilon T} a_{ij}\right| = |x-y| \leq \sqrt{kln}$. We know that $k > 0$ and $l > 0$, so $kl > 0$.

$$\frac{|x-y|}{kl} \leq \sqrt{\frac{kln}{(kl)^2}} = \sqrt{\frac{n}{kl}} \leq \sqrt{\frac{n}{100n}} = \frac{1}{10}$$

where the last inequality holds because $kl \geq 100n$. Since all entries of T are either +1 or -1, the sum of the number of +1's and the number of -1's is the number of entries in T, so $x + y = kl$, hence $y = kl - x$. Substituting in for y,

$$
\begin{aligned}
\frac{|x-(kl-x)|}{kl} &\leq \frac{1}{10} \\
|2x-kl| &\leq \frac{kl}{10} \\
\frac{-kl}{10} &\leq 2x-kl \leq \frac{kl}{10} \\
\frac{9kl}{20} &\leq x \leq \frac{11kl}{20} \\
\frac{9kl}{20} &\leq kl-x \leq \frac{11kl}{20} \\
\frac{9kl}{20} &\leq y \leq \frac{11kl}{20}
\end{aligned}
$$

□

Definition 1.8. A **random matrix** is a matrix whose entries are randomly assigned values. Entries' assignments are independent of each other.

To see how Corollary 1.7 shows Hadamard matrix A to be like a random matrix but not a random matrix, consider a random $n \times n$ matrix B whose entries are assigned either +1 or -1 with probability p and $1-p$ respectively. Consider U, a $k \times l$ submatrix of B. U has kl entries, and w, the number of entries of the kl entries that are +1, would be a random variable.

Considering U's entries' assignments independent trials that result in either success or failure and calling the occurrence of +1 a "success," $P(w = s)$ is the probability of s successes in kl independent trials, which is the product of the probability of a particular sequence of s successes, $p^s(1-p)^{kl-s}$, and the number of such sequences, $\binom{kl}{s}$, so $P(w = s) = \binom{kl}{s} p^s(1-p)^{kl-s}$. In other words, w has a binomial probability distribution. Hence, w has expected value klp. If each entry has equal probability of being assigned +1 or -1, $p = \frac{1}{2}$ so $\mathrm{E}(w) = kl(\frac{1}{2})$. Note that w can take values far from $\mathrm{E}(w)$, since $P(w = s)$ shows w has nonzero probability of being any integer s where $0 \leq s \leq kl$.

Now consider $n \times n$ Hadamard matrix A, its $k \times l$ submatrix T, and x, the number of +1's in T. Corollary 1.7 shows that x must take values close to $\mathrm{E}(w)$. More precisely, if $kl > 100n$, x must be within 5% of $\mathrm{E}(w)$. x is like random w in that we can expect x to take values close to the expected value of w. However, x is not random because it *must* be within 5% of $\mathrm{E}(w)$, while random w can take values farther from $\mathrm{E}(w)$, any value ranging from 0 to kl. [2]

The above argument is symmetrical: It can be used to compare y, the number of -1's in T, and z, the number of -1's in U. In deriving $P(w = s)$, we called the occurrence of +1 a "success." We could have arbitrarily called the occurrence of -1 a "success." Then $P(z = s) = \binom{kl}{s} p^s(1-p)$, $\mathrm{E}(z) = kl(\frac{1}{2})$ if $p = \frac{1}{2}$, and y would be like random z, but not random, in the same way that x would be like random w, but not random.

[2] If $kl \geq 100n$, x must be within 5% of $\mathrm{E}(w)$. 100 was used in the hypothesis of 1.7 for the sake of concreteness. Any arbitrary constant c could have replaced 100, so that $kl \geq cn$. So long as $c > 1$, x is more limited than w in the values it can take.

In short, $n \times n$ Hadamard matrix A is "quasirandom" because it is like a random matrix B, but not itself a random matrix. Characteristics (x and y) of $k \times l$ T, a sufficiently large [3] submatrix of A, are similar to characteristics (w and z) of $k \times l$ U, a submatrix of B. A is like, but not, a random matrix B because submatrices of A have properties similar to, but not the same as, submatrices of B.

2. The Quasirandomness Theorem

Definition 2.1. A **graph** $G = (V, E)$ is a pair of sets. Elements of V are called vertices and elements of E are called edges. E consists of unordered pairs of vertices such that no vertex forms an edge with itself: $\forall v \in V, E \subset V \times V \setminus \{v, v\}$. $v_1, v_2 \in V$ are **adjacent** when $\{v_1, v_2\} \in E$, denoted $v_1 \sim v_2$. The **degree** of a vertex is the number of vertices with which it forms an edge.

Notation 2.2. If x is a vertex, deg(x) denotes its degree.

Remark 2.3. Vertices can be visualized as points and an edge can be visualized as a line segment connecting two points.

Definition 2.4. Consider a graph $G = (V, E)$ and let n denote G's maximum number of possible edges, i.e. the number of edges there would be if every vertex were connected with every other vertex, so that $n = \begin{pmatrix} |V| \\ 2 \end{pmatrix}$. $|E|$ is the number of edges in the graph. The **density** p of G is $\frac{|E|}{n}$.

Definition 2.5. A **bipartite graph** $\Gamma(L, R, E)$ is a graph consisting of two sets of vertices L and R such that an edge can only exist between a vertex in L and a vertex in R. Call L the "left set" and R the "right set."

Notation 2.6. Given two sets of vertices V_1 and V_2, $E(V_1, V_2)$ denotes the set of edges between vertices in V_1 and vertices in V_2. $|E(V_1, V_2)|$ denotes the number of elements in $E(V_1, V_2)$.

Definition 2.7. A **bipartite adjacency matrix** of a bipartite graph that has k vertices in the left set and l vertices in the right set is a $k \times l$ matrix such that

$$a_{ij} = \begin{cases} 1 & \text{if } i \sim j, \text{where } i \in L \text{ and } j \in R \\ 0 & \text{otherwise} \end{cases}$$

Remark 2.8. Let A be a k x l bipartite adjacency matrix. $(A^T A)^T = A^T (A^T)^T = A^T A$. Since $A^T A$ is symmetric, it has l real eigenvalues, denoted λ_1, ..., λ_l in decreasing order. $A^T A$ is **positive semidefinite** because $\forall x \in \mathbb{R}^l, x^T A^T A x = \|Ax\|^2 \geq 0$. Since $A^T A$ is positive semidefinite, its eigenvalues are nonnegative.

Definition 2.9. A **biregular bipartite graph** $\Gamma(L, R, E)$ is a bipartite graph where every vertex in L has the same degree s_r and every vertex in R has the same degree s_c.

Remark 2.10. $|E| = |L|\, s_r = |R|\, s_c$.

Fact 2.11. (Rayleigh Principle) Let $n \times n$ symmetric matrix A have eigenvalues $\lambda_1, ..., \lambda_n$ in decreasing order. Define the Rayleigh quotient $R_A(x) = \frac{\vec{x}^T A \vec{x}}{\vec{x}^T \vec{x}}$. Then $\lambda_1 = \underset{\vec{x} \in \mathbb{R}^n, \vec{x} \neq \vec{0}}{max} R_A(x)$.

[3] $kl > n$

Notation 2.12. Subscripts of the form $m \times n$ on matrices and vectors give their dimensions: m rows and n columns. $(x_1, .., x_n)_{1\times n}$, denotes a $1 \times n$ row vector where the x_i are components of $\vec{x}$. $\vec{1}$ denotes a vector with 1 for every component.

Lemma 2.13. *Let $\Gamma(L, R, E)$ be a biregular bipartite graph with $|L| = k$ and $|R| = l$. Let each vertex in L have degree s_r and let each vertex in R have degree s_c. Let A be the $k \times l$ adjacency matrix of Γ, and let λ_1 be the largest eigenvalue of $A^T A$. Then $\lambda_1 = s_r s_c$.*

Proof. Let $\vec{r_1}, ..., \vec{r_k}$ be the row vectors of A. Recall that A has only 1 or 0 for entries and that each $\vec{r_i}$ contains s_r 1's, so dotting $\vec{r_i}$ with some vector adds together s_r components of that vector.

$$\frac{\left\|A\vec{1}_{l\times 1}\right\|^2}{\left\|\vec{1}_{l\times 1}\right\|^2} = \frac{\left\|(\vec{r}_1 \cdot \vec{1}_{l\times 1}, ..., \vec{r}_k \cdot \vec{1}_{l\times 1})\right\|^2}{l} = \frac{\|(s_r, ..., s_r)_{1\times k}\|^2}{l} = \frac{k s_r^2}{l} = (\frac{k s_r}{l}) s_r = s_c s_r$$

where the last equality follows from $s_r k = s_c l$ 2.10.

We have that $\frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} = s_c s_r$ when $\vec{x} = \vec{1}_{l\times l}$. If we could show that $\forall \vec{x} \in \mathbb{R}^l$, $\frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} \leq s_c s_r$, then we would have that $\frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2}$ reaches its upper bound $s_c s_r$, so its max must be $s_c s_r$, and by 2.11,

$$\lambda_1 = \max_{\vec{x}\in\mathbb{R}^l, \vec{x}\neq\vec{0}} \frac{\vec{x}^T A^T A \vec{x}}{\vec{x}^T \vec{x}} = \max_{\vec{x}\in\mathbb{R}^l, \vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} = s_c s_r$$

It remains to show that $\forall \vec{x} \in \mathbb{R}^l$, $\frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} \leq s_c s_r$.

Let $x_1, ...x_l$ denote the components of $\vec{x}$. $A\vec{x} = (\vec{r}_1 \cdot \vec{x}, ..., \vec{r}_k \cdot \vec{x})^T$, so

$$\|A\vec{x}\|^2 = \sum_{i=1}^{k} (\vec{r}_i \cdot \vec{x})^2 \tag{2.14}$$

$\vec{r}_i \cdot \vec{x}$ is the sum of s_r components of $\vec{x}$. Let $x_{i1}, ..., x_{is_r}$ be the s_r components of $\vec{x}$ that $\vec{r}_i$ selects to sum. Then $\vec{r}_i \cdot \vec{x} = \sum_{j=1}^{s_r} x_{ij}$.

$$\vec{r}_i \cdot \vec{x} = \sum_{j=1}^{s_r} x_{ij} = (x_{i1}, ..., x_{is_r}) \cdot \vec{1}_{s_r\times 1} \leq \left\|\vec{1}_{s_r\times 1}\right\| \|(x_{i1}, ..., x_{is_r})\| = \sqrt{s_r} \sqrt{\sum_{j=1}^{s_r} (x_{ij})^2}$$

where the inequality follows from the Cauchy-Schwarz Inequality, so we have that $(\vec{r}_i \cdot \vec{x})^2 \leq s_r \sum_{j=1}^{s_r} (x_{ij})^2$. Substituting into 2.14,

$$\|A\vec{x}\|^2 \leq s_r \sum_{i=1}^{k} \sum_{j=1}^{s_r} (x_{ij})^2 \tag{2.15}$$

Observe that the first summation cycles through all the row vectors and, for each row vector $\vec{r}_i$, the second summation cycles through the components of $\vec{x}$ chosen by $\vec{r}_i$. Recall that A has s_c 1's in every column, so in multiplying A and $\vec{x}$, every component of $\vec{x}$ is selected by exactly s_c row vectors. Hence,

$$\sum_{i=1}^{k}\sum_{j=1}^{s_r}(x_{ij})^2 = s_c\sum_{i=1}^{l}(x_i)^2 = s_c \|\vec{x}\|^2$$

Substituting into 2.15, $\|A\vec{x}\|^2 \leq s_r s_c \|\vec{x}\|^2$, so $\forall \vec{x} \in \mathbb{R}^l$, $\frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} \leq s_c s_r$. □

Lemma 2.16. *Under the assumptions of 2.13, $\vec{1}_{l\times l}$ is an eigenvector of $A^T A$ corresponding to eigenvalue λ_1.*

Proof. Each entry of $A\vec{1}_{l\times 1}$ is the sum of a row of A, which is s_r, so $A\vec{1}_{l\times 1} = s_r\vec{1}_{k\times 1}$. Similarly, $A^T\vec{1}_{k\times 1} = s_c\vec{1}_{l\times 1}$. Hence, $A^T A\vec{1}_{l\times 1} = A^T(s_r\vec{1}_{k\times 1}) = s_r(A^T\vec{1}_{k\times 1}) = s_r s_c\vec{1}_{l\times 1} = \lambda_1\vec{1}_l$, where the last equality follows by 2.13. We have that $A^T A\vec{1}_{l\times 1} = \lambda_1\vec{1}_{l\times 1}$, so $\vec{1}_{l\times 1}$ is an eigenvector of $A^T A$ corresponding to eigenvalue λ_1. □

Notation 2.17. J denotes a matrix with 1 for every entry.

Theorem 2.18. *(Quasirandomness Theorem) Suppose $\Gamma(L, R, E)$ is a biregular bipartite graph with $|L| = k$ and $|R| = l$. Let the degree of every vertex in L be s_r and the degree of every vertex in R be s_c. Let $X \subseteq L$ and $Z \subseteq R$, let p be the density of Γ, let A be the $k\times l$ adjacency matrix of Γ, and let λ_i be the i^{th} eigenvalue of $A^T A$ in decreasing order. Then*

$$||E(X,Z)| - p\,|X|\,|Z|| \leq \sqrt{\lambda_2\,|X|\,|Z|}$$

Proof. Let $\vec{x}$ be the incidence vector of X and let $\vec{z}$ be the incidence vector of Z. $|E(X,Z)| = \vec{x}^T A\vec{z}$. Consider the subgraph $\Gamma(X, Z, E(X,Z))$. If all vertices in X were connected with all vertices in Z, the number of edges in the subgraph would be $|X|\,|Z| = \vec{x}^T J_{k\times l}\vec{z}$.

$$\begin{aligned} ||E(X,Z)| - p\,|X|\,|Z|| &= \left|\vec{x}^T A\vec{z} - p(\vec{x}^T J_{k\times l}\vec{z})\right| = \left|\vec{x}^T(A - pJ_{k\times l})\vec{z}\right| \\ &\leq \left\|\vec{x}^T\right\| \|(A - pJ_{k\times l})\vec{z}\| = \sqrt{|X|}\,\|(A - pJ_{k\times l})\vec{z}\| \end{aligned}$$

where the inequality follows by the Cauchy-Schwarz inequality. It remains to show that $\|(A - pJ_{k\times l})\vec{z}\| \leq \sqrt{\lambda_2\,|Z|}$ i.e. $\|(A - pJ_{k\times l})\vec{z}\|^2 \leq \lambda_2\,|Z| = \lambda_2\,\|\vec{z}\|^2$.

$$\begin{aligned} \|(A - pJ_{k\times l})\vec{z}\|^2 &= \vec{z}^T(A - pJ_{k\times l})^T(A - pJ_{k\times l})\vec{z} \\ &= \vec{z}^T(A^T - pJ_{k\times l}^T)(A - pJ_{k\times l})\vec{z} \\ &= \vec{z}^T(A^T A - pA^T J_{k\times l} - pJ_{k\times l}^T A + p^2 J_{k\times l}^T J_{k\times l})\vec{z} \end{aligned}$$

We will simplify $A^T A - pA^T J_{k\times l} - pJ_{k\times l}^T A + p^2 J_{k\times l}^T J_{k\times l}$ term-by-term.

(Simplifying $J_{k\times l}^T A$) Γ is biregular: Every vertex in R is connected to s_c vertices in L, so $s_c = \frac{|E|}{l}$, and every vertex in L is connected to s_r vertices in R, so $s_r = \frac{|E|}{k}$. Put another way, the entries of each column of A sum to s_c and the entries of each row of A sum to s_r. $p = \frac{|E|}{kl}$, so:

$$s_c = \frac{|E|}{l} = \frac{\frac{|E|}{kl}(kl)}{l} = \frac{pkl}{l} = pk$$

$$s_r = \frac{|E|}{k} = \frac{\frac{|E|}{kl}(kl)}{k} = \frac{pkl}{k} = pl$$

Notice that each entry of $J_{k\times l}^T A$ is s_c, which is pk, so $J_{k\times l}^T A = pkJ_{l\times l}$.

(Simplifying $A^T J_{k\times l}$) $A^T J_{k\times l} = (J_{k\times l}^T A)^T = (pkJ_{l\times l})^T = pkJ_{l\times l}$, where the last equality holds because $J_{l\times l}$ is symmetric.

(Simplifying $J_{k\times l}^T J_{k\times l}$) Each entry of $J_{k\times l}^T J_{k\times l}$ is the sum of a column of $J_{k\times l}$, which is k, so $J_{k\times l}^T J_{k\times l} = kJ_{l\times l}$.

Substituting in for $J_{k\times l}^T A$, $A^T J_{k\times l}$, and $J_{k\times l}^T J_{k\times l}$:

$$\begin{aligned} A^T A - pA^T J_{k\times l} - pJ_{k\times l}^T A + p^2 J_{k\times l}^T J_{k\times l} &= A^T A - p(pkJ_{l\times l}) - p(pkJ_{l\times l}) + p^2(kJ_{l\times l}) \\ &= A^T A - p^2 kJ_{l\times l} \equiv M \end{aligned}$$

By 2.16, $\vec{1}$ is an eigenvector of $A^T A$ to eigenvalue $\lambda_1 = s_r s_c = (pk)(pl) = p^2 kl$. Since $J_{l\times l}\vec{1} = l\vec{1}$, $(p^2 kJ_{l\times l})\vec{1} = p^2 k(J_{l\times l}\vec{1}) = p^2 k(l\vec{1}) = (p^2 kl)\vec{1} = \lambda_1\vec{1}$. Now consider $M = A^T A - p^2 kJ_{l\times l}$.

$$M\vec{1} = A^T A\vec{1} - p^2 kJ_{l\times l}\vec{1} = \lambda_1\vec{1} - \lambda_1\vec{1} = \vec{0} = 0\vec{1}$$

so $\vec{1}$ is an eigenvector of M corresponding to eigenvalue 0. Also, $M = A^T A - p^2 kJ_{l\times l} = (A^T A)^T - (p^2 kJ_{l\times l})^T = (A^T A - p^2 kJ_{l\times l})^T = M^T$. Since M is a symmetric matrix, by the Spectral Theorem, there exists an orthogonal eigenbasis to M. Let $\vec{e}_i$ be a vector in this orthogonal eigenbasis, so $M\vec{e}_i = u_i\vec{e}_i$, where $u_i \in \mathbb{R}$ is an eigenvalue of M. Let $\vec{e}_1 \equiv \vec{1}_l$, so $u_1 = 0$. Since the $\vec{e}_i$ are orthogonal, $\vec{1}$ is orthogonal to $\vec{e}_i$, $i \geq 2$. Notice that for $i \geq 2$, each entry of $J_{l\times l}\vec{e}_i$ is $\vec{1}\cdot\vec{e}_i = 0$, so $J_{l\times l}\vec{e}_i = \vec{0}$. Hence, for $i \geq 2$, $M\vec{e}_i = (A^T A - p^2 kJ_{l\times l})\vec{e}_i = A^T A\vec{e}_i - p^2 k(J_{l\times l}\vec{e}_i) = A^T A\vec{e}_i$. For $i \geq 2$, $u_i\vec{e}_i = M\vec{e}_i = A^T A\vec{e}_i = \lambda_i\vec{e}_i$ so $u_i = \lambda_i$ for $i \geq 2$.

This implies that the largest eigenvalue of M is λ_2, NOT λ_1: Since λ_i's are ordered by size and no $u_i = \lambda_1$ for $i \geq 2$ and $u_1 = 0$, which is not generally equal to $\lambda_1 = s_r s_c \geq 0$, no u_i ever is λ_1. The next largest value that a u_i can be is λ_2. (In particular, the largest eigenvalue of M is $u_2 = \lambda_2$.)

By 2.11, the largest eigenvalue of M is $max_{\vec{z}}\frac{\vec{z}^T M\vec{z}}{\vec{z}^T\vec{z}}$. $\frac{\vec{z}^T M\vec{z}}{\vec{z}^T\vec{z}} \leq max_{\vec{z}}\frac{\vec{z}^T M\vec{z}}{\vec{z}^T\vec{z}} = \lambda_2 \Rightarrow \vec{z}^T M\vec{z} \leq \lambda_2\vec{z}^T\vec{z}$, and $\vec{z}^T\vec{z} = \vec{z}\vec{z} = \|z\|^2$, so $\vec{z}^T M\vec{z} \leq \lambda_2\|z\|^2$. Recall,

$$\begin{aligned} \|(A - pJ)\vec{z}\|^2 &= \vec{z}^T(A - pJ)^T(A - pJ)\vec{z} \\ &= \vec{z}^T(A^T A - pA^T J_{k\times l} - pJ_{k\times l}^T A + p^2 J_{k\times l}^T J_{k\times l})\vec{z} \\ &= \vec{z}^T M\vec{z} \\ &\leq \lambda_2\|z\|^2 \end{aligned}$$

which is what we needed to finish the proof. □

The smaller λ_2 is, the closer $|E(X,Z)|$ is to $p|X||Z|$, so the closer $\frac{|E(X,Z)|}{|X||Z|}$ is to $\frac{p|X||Z|}{|X||Z|} = p$. Notice that $\frac{|E(X,Z)|}{|X||Z|}$ is the density of the bipartite subgraph formed by X and Z, $\Gamma(X \subseteq L, Z \subseteq R, E(X,Z))$. Hence, the Quasirandomness Theorem says that the density of $\Gamma(X,Z,E(X,Z))$ is approximately the density of the larger graph $\Gamma(L,R,E)$.

Corollary 2.19. *Under the same hypotheses as Theorem 2.18, if* $p^2|X||Z| > \lambda_2$, *then* $|E(X,Z)| > 0$.

Proof.

$$\begin{aligned} p^2\,|X|\,|Z| > \lambda_2 &\Leftrightarrow p^2\,(|X|\,|Z|)^2 > \lambda_2\,|X|\,|Z| \\ &\Leftrightarrow p\,|X|\,|Z| > \sqrt{\lambda_2\,|X|\,|Z|} \\ &\Leftrightarrow p\,|X|\,|Z| - \sqrt{\lambda_2\,|X|\,|Z|} > 0 \end{aligned}$$

By 2.18,

$$\begin{aligned} ||E(X,Z)| - p\,|X|\,|Z|| \le \sqrt{\lambda_2\,|X|\,|Z|} &\Rightarrow -\sqrt{\lambda_2\,|X|\,|Z|} \le |E(X,Z)| - p\,|X|\,|Z| \\ &\Leftrightarrow p\,|X|\,|Z| - \sqrt{\lambda_2\,|X|\,|Z|} \le |E(X,Z)| \end{aligned}$$

Combining the above results,
$0 < p\,|X|\,|Z| - \sqrt{\lambda_2\,|X|\,|Z|} \le |E(X,Z)| \Leftrightarrow 0 < |E(X,Z)|$ □

2.1. **How the Quasirandomness Theorem is a quasirandomness result.**

Definition 2.20. A **random graph** is a graph whose every pair of vertices is randomly assigned an edge. Pairs' assignments are independent of each other.

Remark 2.21. A **random bipartite graph** is a random graph such that any two vertices in the same set have 0 probability of forming an edge.

Consider a random situation. Let $G(L', R', E')$ be a random bipartite graph, and let each pair $\{l, r\}$, $l \in L'$ and $r \in R'$, have probability p of being an edge. Let $X' \subseteq L'$ and let $Z' \subseteq R'$. Consider the subgraph $g(X', Z', E(X', Z'))$. The number of pairs of vertices of g that can form edges is $|X'|\,|Z'|$.

Considering the designation of edge a "success," $|E(X', Z')|$, the number of "successes" in $|X'|\,|Z'|$ independent trials, would follow a binomial distribution: $P(|E(X', Z')| = s) = \binom{|X'||Z'|}{s} p^s(1-p)^{|X'||Z'|-s}$. $|E(X', Z')|$ would have expected value $p\,|X'|\,|Z'|$, so the density of g, $\frac{E(X',Z')}{|X'||\ Z'|}$, would have expected value $\frac{p|X'||Z'|}{|X'||Z'|} = p$. By the same argument, $P(|E'| = s) = \binom{|L'||R'|}{s} p^s(1-p)^{|L'||R'|-s}$, the expected value of $|E'|$ would be $p\,|L'|\,|R'|$, so the density of G, $\frac{E(L',R')}{|L'||\ R'|}$, would have expected value p. The density of G and the density of g have the same expected value, but there is no guarantee that the densities be within some range of each other. The probability that the densities are wildly different, say a density of 0 and a density of 1, is nonzero.

Now consider biregular bipartite graph $\Gamma(L, R, E)$ described in the hypotheses of 2.18. The Quasirandomness Theorem says that the density of subgraph $\Gamma(X \subseteq L, Z \subseteq R, E(X, Z))$ must be within some range [4] of the density of $\Gamma(L, R, E)$, so in this sense one can expect the density of $\Gamma(X, Z, E(X, Z))$ to be approximately the density of $\Gamma(L, R, E)$. Similarly, one can expect the density of G and the density of g to be close to each other (in the sense that their expected values are the same), but unlike the density of $\Gamma(L, R, E)$ and the density of $\Gamma(X, Z, E(X, Z))$, the density of G and the density of g are not necessarily within some range (other than 1) of each other.

[4]The range is controlled by λ_2 and the sizes of X and Z, and could be less than 1. The larger X and Z are and the smaller λ_2 is, the closer the density of the subgraph is to the density of the larger graph.

$\Gamma(L, R, E)$ is a quasirandom graph because it is like a random graph $G(L', R', E')$. One can expect sufficiently large subgraphs of $\Gamma(L, R, E)$ to have characteristics (namely densities) similar to characteristics of subgraphs of a random graph.

3. Gowers theorem

Theorem 3.1. *(Gowers' Theorem - GT) Let G be a group of order $|G|$ and let m be the minimum degree of nontrivial representations of G over the reals. If $X, Y, Z \subseteq G$ and $|X|\,|Y|\,|Z| \geq \frac{|G|^3}{m}$, then $\exists\, x \in X,\ y \in Y,\ z \in Z\, s.t.\, xy = z$.*

Corollary 3.2. *3.1 would still be true if its conclusion were replaced by $XYZ = G$*

Proof. Take $X, Y, Z \subseteq G$ such that $|X|\,|Y|\,|Z| \geq \frac{|G|^3}{m}$.
$XYZ = G$ means $\forall x \in X,\ y \in Y,\ z \in Z,\ \exists g \in G\, s.t.\, xyz = g$ and $\forall g \in G,\ \exists x \in X,\ y \in Y,\ z \in Z,\ s.t.\, xyz = g$. The first statement holds by closure of G, so it remains to show the second statement. Take $g \in G$. Let $Z' = gZ^{-1}$. By closure of G, $Z' \in G$. Since $|Z'| = |Z|$, $|X|\,|Y|\,|Z'| \geq \frac{|G|^3}{m}$. By 3.1, $\exists\, x \in X,\ y \in Y,\ z' \in Z'\, s.t.\, xy = z' \Leftrightarrow xy(z'^{-1}) = z'(z'^{-1}) = 1 \Leftrightarrow xy(z'^{-1}g) = g \Leftrightarrow xyz = g$. ☐

3.1. Translating Gowers Theorem: Proving $m_2 \geq m$ Proves Gowers' Theorem.

Variables in this subsection refer to those defined in the context of $\Gamma(G_2, G_2, E)$:

To prove 3.1, we take a graph theoretic view of it. Let G be a group. Let $\Gamma(G_1, G_2, E)$ be a bipartite graph with two sets of vertices G_1 and G_2, which are copies of G. Let there be an edge between $g_1 \in G_1$ and $g_2 \in G_2$ only if $\exists y \in Y \subseteq G\, s.t.\, g_1 y = g_2$, let A be the $|G| \times |G|$ adjacency matrix of Γ, let λ_2 be the second largest eigenvalue of $A^T A$, let p be the density of Γ, let $X \subseteq G_1$, and let $Z \subseteq G_2$.

3.1 says that, for sufficiently large X and Z, there is at least one edge between a member of X and a member of Z, i.e. $|E(X, Z)| > 0$. Curiously, which particular vertices are chosen to constitute X and Z is irrelevant to guaranteeing an edge between them. Rather, the sizes of X and Z are all that matter.

In this graph theoretic view of Gowers' Theorem, the hypotheses of the Quasirandomness Thrm hold. If, in addition, $p^2\,|X|\,|Z| > \lambda_2$ were to hold, then by 2.19, $|E(X, Z)| > 0$, proving Gowers' Theorem. To translate proving GT into proving some other statement, we use the following results:

Notation 3.3. g_1 denotes any vertex in G_1 and g_2 denotes any vertex in G_2.

Lemma 3.4. *The degree of every vertex of $\Gamma(G_1, G_2, E)$ is $|Y|$*

Proof. We will show that every vertex in G_1 has degree $|Y|$ and every vertex in G_2 has degree $|Y|$, so every vertex of Γ has degree $|Y|$.

Claim: Every $g_1 \in G_1$ has degree $|Y|$. Since G is a group, $\forall g, y \in G$, $gy \in G$ so $\forall g_1 \in G_1 = G$ and $y \in Y \subseteq G$, $g_1 y \in G = G_2$ so $g_1 y = g_2 \in G_2$. Every g_1 can be multiplied by every element in Y to get a g_2.

$\forall g_1$, multiplying g_1 by different y leads to distinct products. Take distinct $y_1, y_2 \in Y$ and suppose, for a contradiction, that $g_1 y_1 = h$ and $g_1 y_2 = h$. Then

$y_1 = g_1^{-1}h$ and $y_2 = g_1^{-1}h$, so $y_1 = y_2$, contradicting the assumption that y_1 and y_2 are distinct, so $g_1 y_1 \neq g_1 y_2$.

Hence, for each g_1, multiplying by every y yields $|Y|$ distinct products in G_2. Since $\{g_1, g_2\} \in E$ iff $\exists y \in Y \; s.t. \; g_1 y = g_2$, g_1 can form no other edges, so the degree of every g_1 is $|Y|$.

Claim: Every g_2 has degree $|Y|$. Every g_2 has $|Y|$ preimages in G_1: $\forall y \in Y, \exists$ unique $g_1 \in G_1$ s.t. $g_1 y = g_2$. Take $y \in Y \subseteq G$ so $y \in G$. Since G is a group, $y^{-1} \in G$. Take $g_2 \in G_2 = G$. By closure, $g_2 y^{-1} \in G = G_1$ so $g_1 = g_2 y^{-1}$.

To count the number of g_1's that form an edge with a g_2, it suffices to count the number of y's, which is $|Y|$. □

Corollary 3.5. $|E| = |G|\,|Y|$

Proof. Every $g_1 \in G_1$ forms $|Y|$ edges, and there are $|G|$ g_1's, so $|E| = |G|\,|Y|$ □

Fact 3.6. If A is an $n \times n$ real matrix with eigenvalues $\lambda_1, ..., \lambda_n$, then $Tr(A) = \sum_{i=1}^{n} \lambda_i$

Notation 3.7. λ_i denotes one of the $|G|$ eigenvalues of $A^T A$: $\{\lambda_1, ..., \lambda_{|G|}\}$, listed in decreasing order. m_i denotes the multiplicity of λ_i.

Corollary 3.8. $\lambda_2 < \frac{Tr(A^T A)}{m_2}$

Proof. By 3.6, $Tr(A^T A) = \sum_{i=1}^{|G|} \lambda_i = m_1 \lambda_1 + m_2 \lambda_2 + ... > m_2 \lambda_2$, where the last inequality follows from $A^T A$ having nonnegative eigenvalues (by 2.8). □

Lemma 3.9. $Tr(A^T A) = |E(X, Z)|$

Proof. Let $\vec{c_1}, ..., \vec{c_{|G|}}$ be the column vectors of A.

$$Tr(A^T A) = \sum_{j=1}^{|G|} \vec{c_j} \cdot \vec{c_j} = \sum_{j=1}^{|G|} (\sum_{i=1}^{|G|} c_{ij})$$

This double summation adds all the entries of A, hence counts the number of edges of $\Gamma(G_1, G_2, E)$.

An alternative view: The second summation gives the degree of a particular g_2. The first summation cycles through all vertices in G_2. Hence, the double summation counts all the edges that vertices in G_2 are members of, so it counts all the edges of Γ. □

Corollary 3.10. $\lambda_2 < \frac{|G||Y|}{m_2}$

Proof. $\lambda_2 < \frac{Tr(A^T A)}{m_2} = \frac{|E(X,Z)|}{m_2} = \frac{|G||Y|}{m_2}$. The first inequality holds by 3.8, the second equality holds by 3.9, and the third equality holds by 3.5. □

Remark 3.11. $p = \frac{|G||Y|}{|G||G|} = \frac{|Y|}{|G|}$, where the first equality follows from 3.5 and 2.4.

Proposition 3.12. *To prove Gowers' Theorem, it remains to show that* $m_2 \geq m$.

Proof. From 3.10, we have that $\lambda_2 < \frac{|G||Y|}{m_2}$. If we could show that $\frac{|G||Y|}{m_2} \leq p^2 |X|\,|Z|$, then $\lambda_2 < p^2 |X|\,|Z|$, fulfilling the hypothesis of 2.19 and reaching the conclusion of Gowers' Theorem. In other words, to prove GT, it remains to prove $\frac{|G||Y|}{m_2} \leq p^2 |X|\,|Z|$.

$\frac{|G||Y|}{m_2} \le p^2 |X| |Z| \Leftrightarrow \frac{|G||Y|}{m_2} \le \left(\frac{|Y|}{|G|}\right)^2 |X| |Z| \Leftrightarrow \frac{|G|^3}{m_2} \le |X| |Y| |Z|$, where the first iff follows from 3.11. To prove GT it remains to prove $\frac{|G|^3}{m_2} \le |X| |Y| |Z|$.

Given GT's hypothesis $|X| |Y| |Z| \ge \frac{|G|^3}{m}$, if we could show $m_2 \ge m$, then $|X| |Y| |Z| \ge \frac{|G|^3}{m_2}$. Hence, all we need to prove GT is $m_2 \ge m$. □

3.2. Proving $m_2 \ge m$.

Recall that m_2 is the multiplicity of λ_2 and m is the minimum dimension of nontrivial representations of G over $\mathbb{R}$ i.e. the smallest dimension of a real vector space in which G has nontrivial representation. To show that $m_2 \ge m$, we will need some preliminary definitions and results.

Definition 3.13. For a group G and an integer $d \ge 1$, a **d-dimensional representation of G** is a homomorphic map $\varphi : G \to GL(V)$, where V is a d-dimensional vector space, so $V \cong F^d$, where F is a field. $GL(V) \cong GL_d(F)$, which is the **general linear group**, the set of d x d invertible matrices whose entries are elements of F; the set forms a group under matrix multiplication. Since $GL(V) \cong GL_d(F)$, φ is a mapping $G \to GL_d(F)$, so we say φ is a **representation of G over F**. d is the **dimension** of φ.

Remark 3.14. A **representation of** G **over** $\mathbb{R}$ is a representation of G, $\varphi : G \to GL_d(\mathbb{R})$. To clarify, such a φ maps elements of G to $d \times d$ invertible matrices with entries from $\mathbb{R}$. Such matrices correspond to invertible mappings from $\mathbb{R}^d$ to $\mathbb{R}^d$.

Definition 3.15. Let V be a d-dimensional vector space. $U \subseteq V$ is **invariant** under $\varphi : G \to GL(V)$ if $\forall g \in G$, U is invariant under $\varphi(g)$, i.e. $\forall u \in U$, $g \in G$, $\varphi(g)u \in U$. In other words, every mapping that φ associates with an element of G maps U to U. The **trivial invariant subspaces** are the zero subspace (whose only element is $\vec{0} \in \mathbb{R}^d$) and V.

Definition 3.16. $\varphi : G \to GL_d(\mathbb{R})$ is a **trivial representation** if it maps every element of G to the identity transformation.

Definition 3.17. If $\lambda \in F$ and A is an n x n matrix over F, then the **eigenspace to eigenvalue** λ is $U_\lambda = \{\vec{x} \in F^n \, s.t. \, A\vec{x} = \lambda\vec{x}\}$. A member of the eigenspace is called an **eigenvector** corresponding to λ.

Lemma 3.18. *If $AB = BA$, then every eigenspace of A is invariant under B.*

Proof. Let U_λ be an eigenspace of A. We want to show that $\forall \vec{x} \in U_\lambda$, $B\vec{x} \in U_\lambda$. Since $\vec{x} \in U_\lambda$, $A\vec{x} = \lambda\vec{x}$, so $AB\vec{x} = BA\vec{x} = B(\lambda\vec{x}) = \lambda B\vec{x}$. □

Definition 3.19. An **eigenbasis** of a matrix A is a set of eigenvectors of A that forms a basis for the domain of the linear transformation corresponding to A.

Theorem 3.20. *(Spectral Theorem) Every real symmetric matrix has an orthogonal eigenbasis.*

Notation 3.21. Given mapping $f : A \to B$ and $C \subseteq A$, $f|_C$ denotes the mapping that is the same as f, except with domain restricted to C. Hom(A,B) denotes the set of homomorphisms from A to B.

Proposition 3.22. *Let $A = A^T$ be a real $d \times d$ matrix, and G a group. Let $m = \min\{s : \exists \phi \in$ nontrivial $Hom(G, GL_s(\mathbb{R}))\}$, i.e. m is the minimum dimension of nontrivial representations of G over the reals. Let $\varphi \in Hom(G, GL_d(\mathbb{R}))$ be nontrivial. Suppose that A commutes with all matrices in $GL_d(\mathbb{R})$. Then there is an eigenvalue of A with multiplicity at least m.*

Proof. By 3.20, we can choose a particular eigenbasis of A. Call this basis $\mathcal{B}_A = \{\vec{e_1}, ..., \vec{e_d}\}$. Pick $g_0 \in G$, such that $\varphi(g_0)$ is not the identity matrix. Let $\psi : \mathbb{R}^d \to \mathbb{R}^d$ be the unique linear map whose transformation matrix with respect to $\mathcal{B}_A$ is $\varphi(g_0)$. $\varphi(g_0)$ is not the identity matrix, so ψ is not the identity map on $\mathbb{R}^d$.

Since A commutes with every element of $GL_d(\mathbb{R})$, in particular it commutes with $\varphi(g_0)$, so by 3.18, ψ sends each eigenspace of A to itself. ψ cannot act as the identity on every U_λ, because if it did, then $\forall \vec{v} \in \mathbb{R}^d$, $\vec{v} = \sum_{i=1}^d \alpha_i \vec{e_i}$ where $\alpha_i \in \mathbb{R}$, and

$$\psi(\vec{v}) = \psi(\sum_{i=1}^d \alpha_i \vec{e_i}) = \sum_{i=1}^d \alpha_i \psi(\vec{e_i}) = \sum_{i=1}^d \alpha_i \vec{e_i} = \vec{v}$$

so ψ would act as the identity on $\mathbb{R}^d$, which is contrary to the choice of ψ.

We've shown by contradiction that there must be an eigenspace U_λ such that $\psi : U_\lambda \to U_\lambda$ is not the identity map. Because $\psi|_{U_\lambda}$ is not the identity map, $\varphi(g_0)|_{U_\lambda}$ is not the identity matrix, so $\varphi : g \mapsto \varphi(g)|_{U_\lambda}$ is a nontrivial representation of G. Note that $\varphi : g \mapsto \varphi(g)|_{U_\lambda}$ means $\varphi : G \to GL(U_\lambda) \cong GL_{dim(U_\lambda)}\mathbb{R}$ so the dimension of φ is the dimension of U_λ.

By definition, m is the minimum dimension of nontrivial representations of G, so the dimension of φ (which is the dimension of U_λ) is at least m. Since A is symmetric, the dimension of U_λ is the multiplitiey of λ, so the multiplicity of λ is at least m as desired. □

Definition 3.23. $\sigma : V \to V$ is a **permutation** on set V if it is a bijection from V to V.

Definition 3.24. Consider a graph $G = (V, E)$. A graph automorphism is a mapping $\sigma : V \to V$ that preserves adjacency, i.e. $\forall i, j \in V$, $i \sim j \Leftrightarrow \sigma(i) \sim \sigma(j)$

Remark 3.25. A graph automorphism of a bipartite graph $\Gamma(V_1, V_2, E)$ consists of permutations $\sigma_1 : V_1 \to V_1$ and $\sigma_2 : V_2 \to V_2$ s.t. $\forall v_1 \in V_1$ and $v_2 \in V_2$, $v_1 \sim v_2 \Leftrightarrow \sigma_1(v_1) \sim \sigma_2(v_2)$.

Definition 3.26. $P(\sigma)$ is a **permutation matrix** of permutation σ if

$$P(\sigma)_{ij} = \begin{cases} 1 & \text{if } \sigma(i) = j \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 3.27. *Let $\Gamma(V_1, V_2, E)$ be a biregular bipartite graph, let A be its adjacency matrix, let σ_1 be a permutation of V_1, and let σ_2 be a permutation of V_2. Then σ_1 and σ_2 constitute a bipartite graph automorphism iff $P(\sigma_1)A = AP(\sigma_2)$*

Proof. The claim is that

$$\forall i \epsilon V_1,\ j \epsilon V_2,\ i \sim j \Leftrightarrow \sigma_1(i) \sim \sigma_2(j) \iff P(\sigma_1)A = AP(\sigma_2)$$

We will translate the right-hand side into some other statement.

By definition, $P(\sigma_1)A = AP(\sigma_2) \Leftrightarrow \forall i, j,\ [P(\sigma_1)A]_{ij} = [AP(\sigma_2)]_{ij}$.

For all i, j, $[AP(\sigma_2)]_{ij} = \sum_{l=1}^{L} A_{il}P(\sigma_2)_{lj}$. Notice that cells of A and cells of P only take values 1 or 0, so terms of the sum are either 1 or 0. The summation is equivalent to summing only the terms that are 1. For a term to be 1, A_{il} and $P(\sigma_2)_{lj}$ must both be 1. By definition, $A_{il} = 1$ iff $i \sim l$, and $P(\sigma_2)_{lj} = 1$ iff $\sigma_2(l) = j$. Hence, $A_{il}P(\sigma_2)_{lj} = 1$ iff $i \sim l$ and $\sigma_2(l) = j$, so

$$\sum_{l=1}^{L} A_{il}P(\sigma_2)_{lj} = \sum_{l\, s.t.\, i \sim l = \sigma_2^{-1}(j)} A_{il}P(\sigma_2)_{lj}.$$

Multiple l's can be adjacent to i, but since σ_2 is one-to-one, only one l can equal $\sigma_2^{-1}(j)$, so

$$[AP(\sigma_2)]_{ij} = \sum_{l\, s.t.\, i \sim l = \sigma_2^{-1}(j)} A_{il}P(\sigma_2)_{lj} = \begin{cases} 1 & if\, i \sim \sigma_2^{-1}(j) \\ 0 & \text{otherwise} \end{cases}$$

For all i,j,$[P(\sigma_1)A]_{ij} = \sum_{k=1}^{K} P(\sigma_1)_{ik}A_{kj}$. The terms of this sum are either 1 or 0, so the sum is equivalent to summing only the terms that are 1. For a term to be 1, $P(\sigma_1)_{ik} = 1$ iff $\sigma_1(i) = k$, and $A_{kj} = 1$ iff $k \sim j$. Hence, $P(\sigma_1)_{ik}A_{kj} = 1$ iff $\sigma_1(i) = k$ and $k \sim j$, so

$$\sum_{k=1}^{K} P(\sigma_1)_{ik}A_{kj} = \sum_{k\, s.t.\, \sigma_1(i) = k \sim j} P(\sigma_1)_{ik}A_{kj}$$

.
Multiple k could be adjacent to j, but since σ_1 is one-to-one, only one $k = \sigma_1(i)$. Hence, the summation can have only one term that is 1, so

$$[P(\sigma_1)A]_{ij} = \sum_{k\, s.t. \sigma_1(i) = k \sim j} P(\sigma_1)_{ik}A_{kj} = \begin{cases} 1 & if\, \sigma_1(i) \sim j \\ 0 & \text{otherwise} \end{cases}$$

For all i,j $[P(\sigma_1)A]_{ij} = [AP(\sigma_2)]_{ij}$ iff the cells are both 1 or both 0 iff $\left(\sigma_1(i) \sim j \text{ and } i \sim \sigma_2^{-1}(j)\right)$ or $\neg \left(\sigma_1(i) \sim j \text{ and } i \sim \sigma_2^{-1}(j)\right)$ Hence, $\sigma_1(i) \sim j$ is equivalent to $i \sim \sigma_2^{-1}(j)$.

To summarize, $P(\sigma_1)A = AP(\sigma_2)$ means $\forall i, j$, $\sigma_1(i) \sim j$ iff $i \sim \sigma_2^{-1}(j)$, so the lemma says:

$$\forall i \in V_1,\, j \in V_2,\, i \sim j \Leftrightarrow \sigma_1(i) \sim \sigma_2(j) \iff \forall i \in V_1,\, j \in V_2,\, \sigma_1(i) \sim j \Leftrightarrow i \sim \sigma_2^{-1}(j)$$

($\Rightarrow$)Suppose

$$i \in V_1, j \in V_2, i \sim j \Leftrightarrow \sigma_1(i) \sim \sigma_2(j). \tag{3.28}$$

We want to show $\sigma_1(i) \sim j \Leftrightarrow i \sim \sigma_2^{-1}(j)$.

$$\sigma_1(i) \sim j \Leftrightarrow \sigma_1(i) \sim \sigma_2(\sigma_2^{-1}(j)) \Leftrightarrow i \sim \sigma_2^{-1}(j) \tag{3.29}$$

where the last equivalence comes from the $\Leftarrow$ direction of 3.28

($\Leftarrow$)Suppose

$$i \in V_1, j \in V_2,\ \sigma_1(i) \sim j \Leftrightarrow i \sim \sigma_2^{-1}(j) \tag{3.30}$$

We want to show $i \sim j \Leftrightarrow \sigma_1(i) \sim \sigma_2(j)$.

$$i \sim j \Leftrightarrow i \sim \sigma_2^{-1}(\sigma_2(j)) \Leftrightarrow \sigma_1(i) \sim \sigma_2(j) \tag{3.31}$$

where the last equivalence comes from the $\Leftarrow$ of 3.30 □

Claim 3.32. Let σ be a permutation and let P be its $n \times n$ permutation matrix. $P^T = P^{-1}$.

Proof. The claim is that $PP^T = P^TP = I_{n\times n}$.

Recall that $P(\sigma)_{ij}$ is 1 if $\sigma(i) = j$ and is 0 otherwise. Since σ is a function, every row vector of P has only one entry that is 1. Since σ is bijective, every column vector of P has only one entry that is 1. No two row vectors can have same the same component be 1, because if there were two such row vectors, there would be a column vector with more than one 1-entry, contradicting that every column vector has only one 1-entry. Similarly, no two column vectors can have the same component be 1. Hence, every pair of distinct row vectors of P is orthogonal and every pair of distinct column vectors of P is orthogonal.

Let the rows of P be $\vec{r_1}, ..., \vec{r_n}$. $(PP^T)_{ii} = \vec{r_i} \cdot \vec{r_i} = \sum_{j=1}^{n} r_{ij} = 1$, since every row vector has only one entry that is 1. For $i \neq j$, $(PP^T)_{ij} = \vec{r_i} \cdot \vec{r_j} = 0$ since row vectors are orthogonal. Hence, $PP^T = I$.

Let the columns of P be $\vec{c_1}, ..., \vec{c_n}$. $(P^TP)_{ii} = \vec{c_i} \cdot \vec{c_i} = \sum_{j=1}^{n} c_{ij} = 1$, since every column vector has only one entry that is 1. For $i \neq j$, $(PP^T)_{ij} = \vec{c_i} \cdot \vec{c_j} = 0$, since column vectors are orthogonal. Hence, $P^TP = I$. □

Lemma 3.33. *Let σ_1, σ_2 constitute a graph automorphism. Then $P(\sigma_2)$ commutes with A^TA.*

Proof.

$$\begin{aligned}
P(\sigma_2)^{-1}A^TAP(\sigma_2) &= P(\sigma_2)^TA^T(I_{k\times k})AP(\sigma_2) \\
&= P(\sigma_2)^TA^T(P(\sigma_1)P(\sigma_1)^{-1})AP(\sigma_2) \\
&= P(\sigma_2)^TA^T(P(\sigma_1)P(\sigma_1)^T)AP(\sigma_2) \\
&= (P(\sigma_2)^TA^TP(\sigma_1))(P(\sigma_1)^TAP(\sigma_2)) \\
&= (P(\sigma_2)^{-1}A^TP(\sigma_1))(P(\sigma_1)^{-1}AP(\sigma_2)) = A^TA
\end{aligned}$$

We have $P(\sigma_2)^{-1}A^TAP(\sigma_2) = A^TA$, so $A^TAP(\sigma_2) = P(\sigma_2)A^TA$. □

Now consider the particular bipartite graph $\Gamma(G_2, G_2, E)$ involved in the proof of Gowers' Theorem. A is its bipartite adjacency matrix, which is a $|G|\text{x}|G|$ real matrix, so A^TA is a $|G|\text{x}|G|$ real matrix. Let λ_2 denote the second largest eigenvalue of A^TA. Choose σ_1, σ_2 that constitute a graph automorphism. Let $\varphi : g \mapsto P(\sigma_2)$ be a nontrivial representation of G, i.e. let φ map some g to a $P(\sigma_2)$ that is not the identity matrix. Let ψ be the linear transformation corresponding to this $P(\sigma_2)$. Using an argument similar to that in the proof of Proposition 3.27, we will show that $m_2 \geq m$.

Remark 3.34. Recall definition 3.23. $U_{\lambda_2}(A^TA) \equiv \left\{\vec{x} \in \mathbb{R}^{|G|}\, s.t.\, A^TA\vec{x} = \lambda_2\vec{x}\right\}$

Proposition 3.35. $m_2 \geq m$

Proof. $P(\sigma_2)$ is not the identity matrix, so ψ does not act as the identity on $\mathbb{R}^{|G|}$. Suppose we could show that ψ does not act as the identity on $U_{\lambda_2}(A^TA)$.

Then $P(\sigma_2)|_{U_{\lambda_2}(ATA)}$ would not be the identity matrix. Hence, $\varphi|_{U_{\lambda_2}(ATA)} : G \to P(\sigma_2)|_{U_{\lambda_2}(ATA)}$ would be a nontrivial representation of G, so its dimension would be at least the minimum dimension of a nontrivial representation of G i.e. m.

By 3.33, $P(\sigma_2)$ commutes with A^TA, so by 3.18, $U_{\lambda_2}(A^TA)$ is invariant under $P(\sigma_2)$, so $P(\sigma_2)|_{U_{\lambda_2}(ATA)} = GL(U_{\lambda_2}(ATA))$, so $\varphi|_{U_{\lambda_2}(ATA)} : G \to GL(U_{\lambda_2}(ATA))$, so $dim(U_{\lambda_2}(ATA))$ = the dimension of $\varphi|_{U_{\lambda_2}(ATA)}$, which we already showed is at least m. Since A^TA is symmetric, the $m_2 = dim(U_{\lambda_2}(ATA))$, which is at least m, so $m_2 \geq m$.

It remains to show that ψ does not act as the identity on $U_{\lambda_2}(A^TA)$. The only way a ψ that is not the identity transformation can act as the identity on $U_{\lambda_2}(A^TA)$ is if each vector in $U_{\lambda_2}(A^TA)$ has identical components, i.e. is a multiple of $\vec{1}$. (To see this, note that ψ does not act as the identity on $U_{\lambda_2}(A^TA)$ iff $P(\sigma_2)|_{U_{\lambda_2}(A^TA)}$ is not the identity matrix.) By 2.16, $A^TA\vec{1} = \lambda_1\vec{1} \neq \lambda_2\vec{1}$, so for $c \in \mathbb{R}$, $A^TA(c\vec{1}) \neq \lambda_2(c\vec{1})$ so no multiple of $\vec{1}$ is in $U_{\lambda_2}(A^TA)$, so ψ cannot act as the identity on $U_{\lambda_2}(A^TA)$. □

3.35 finishes the proof of Gowers' Theorem.

Acknowledgements

Thanks to my mentors Irine Peng and Marius Beceanu for their help.

References

[1] L. Babai. Discrete Math Lecture Notes. http://people.cs.uchicago.edu/ laci/REU07/.
[2] J. A. Rice. Mathematical Statistics and Data Analysis. Duxbury Press. 2006.