7.1 Berlekamp’s switching game revisited

Recall from last time:

Exercise 7.1.1. If there are \( n \) coin flips, \( X_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ coin is } H, \\ -1, & \text{if } i^{\text{th}} \text{ coin is } T \end{cases} \). Let \( Y = \sum X_i = \) #heads – #tails. Also, set \( E(|Y|) = f_n \). Then, show that

\[
 f_n \sim \sqrt{\frac{2}{\pi}} \sqrt{n}. \tag{1}
\]

Corollary 7.1.2. Berlekamp switching game:

\[
 \text{value} \gtrsim \sqrt{\frac{2}{\pi}} n^{\frac{3}{2}} \tag{2}
\]

7.2 Hadamard matrices revisited

Recall that an Hadamard matrix \( A = (a_{ij}) \) of size \( n \) is an \( n \times n \) square matrix with \( a_{ij} = \pm 1 \), whose rows are orthogonal: \( AA^T = nI \). That is, \( (\frac{1}{\sqrt{n}}A) \cdot (\frac{1}{\sqrt{n}}A)^T = I \), i.e., \( (\frac{1}{\sqrt{n}}A)^T = (\frac{1}{\sqrt{n}}A)^{-1} \). We deduce that also \( A^T A = nI \), so the columns must be orthogonal as well.

Lemma 7.2.1. (Lindsey’s Lemma) If \( A = (a_{ij}) \) is an \( n \times n \) Hadamard matrix, and \( T \) is a \( k \times \ell \)-submatrix, then \( |\sum_{(i,j) \in T} a_{ij}| \leq \sqrt{k\ell n} \).

Before the proof, let us consider a neat application.

Question 7.2.2. How large can a homogeneous \( t \times t \)-submatrix of \( A \) be? (if \( A \) is Hadamard).

Claim 7.2.3. \( t \leq \sqrt{n} \).
Proof. Suppose that $T$ is $k \times \ell$ and homogeneous. Then $k\ell = \sum_{(i,j) \in T}a_{ij} \leq \sqrt{k\ell n}$, by Lindsey’s Lemma. So $k^2\ell^2 \leq k\ell n$, i.e., $k\ell \leq n$. If $k = \ell = t$, then $t^2 \leq n$.  

\begin{exercise}
Prove: if $k\ell \geq 100n$ then the +1-s and the −1-s appear in nearly equal numbers in $T$. More precisely, show that that each occupies at least 45% and at most 55% of the cells in $T$.
\end{exercise}

This exercise shows that Lindsey’s Lemma is a “quasirandomness” result: the two values are nearly equally distributed in every sufficiently large submatrix, as they would be if the signs were assigned at random.

\begin{proof}[Proof of Lindsey’s Lemma. (Cauchy-Schwarz magic)] Note first that the sum of all entries of a matrix $A$ is the product $1^T \cdot A \cdot 1$, where $1$ is the all-ones column, and $1^T$ is the all-ones row.

Let now $X \subseteq [n]$ be the set of rows and $Y \subseteq [n]$ be the set of columns that define $T$ (so $|X| = k$ and $|Y| = \ell$). Let $x$ be the incidence vector of $X$, i.e., $x_i = 1$ if $i \in X$ and $x_i = 0$ otherwise. Similarly, let $y$ be the incidence vector of $Y$. Note that $|\|x\|\| = \sqrt{k}$ and $|\|y\|\| = \sqrt{\ell}$. 

Now the sum of all entries of $T$ is $x^T Ay$. We need to estimate $|x^T Ay|$.

The key step is an application of the Cauchy-Schwarz inequality: $|x^T Ay| \leq |\|x\|||Ay||$.

Let $B = \frac{1}{\sqrt{n}} A$. Then $B$ is an orthogonal matrix, so $\|By\| = |\|y\||$ and therefore $|\|Ay|| = \sqrt{n}|\|y||$. We conclude that $|x^T Ay| \leq |\|x\|| \sqrt{n}|\|y|| = \sqrt{k\ell n}$.  
\end{proof}

\section{Gowers’ Theorem revisited}

Our proof of Lindsay’s Lemma was a warm-up to one of the central ingredients of Gowers’ proof, the “quasirandomness theorem.”

Let us recall Gowers’ theorem.

\begin{theorem}[Gowers’ Theorem]
If $G$ is a finite group, $X, Y, Z \subseteq G$ and $|X| \cdot |Y| \cdot |Z| \geq \frac{n^4}{m}$, where $n = |G|$ and $m$ is the minimum dimension of nontrivial representations of $G$, then $(\exists x, y, z)(x \in X, y \in Y, z \in Z; xy = z)$.
\end{theorem}

In fact, the number of solutions is near $\frac{|X| \cdot |Y| \cdot |Z|}{n^4}$, when the assumed inequality is strengthened sufficiently (to $|X| \cdot |Y| \cdot |Z| \geq C \cdot \frac{n^4}{m}$ for large $C$).

\begin{strategy}[Strategy of proof of the theorem]
Let us define a bipartite graph $\Gamma(G, Y)$ using $Y$: the vertices are two copies of the set $G$ (say, $G_1 \sqcup G_2$, where $G_1$ and $G_2$ are copies of $G$), and two elements $g \in G_1, h \in G_2$ are adjacent iff, as elements of the group $G$, $gy = h$ for some $y \in Y$. We call this the “bipartite Cayley graph $\Gamma(G, Y)$.”

We define the density of a bipartite graph with the two sides having $k$ and $\ell$ vertices as $|E|/(k\ell)$ where $E$ is the set of edges; so the density is the proportion of the number of edges compared to the maximum possible number of edges on the same pair of sets of vertices.

Now, the degree of each vertex in $\Gamma(G, Y)$ is the size of $Y$, and thus the total number of edges of $\Gamma(G, Y)$ is $|G| \cdot |Y|$ and the density is $p = \frac{|Y|}{|G|}$.
If this graph is “sort of random,” we expect that the set of edges \( E(X, Z) \) between \( X \subseteq G_1 \) and \( Z \subseteq G_2 \) has size \(|E(X, Z)| \approx p|X| \cdot |Z| = \frac{|X||Y||Z|}{n}\) (where \( n = |G|\)). We will prove that this is indeed the case if \( X, Y, Z \) are large enough.

### 7.4 The Quasirandomness Theorem

The theory of quasirandomness was worked out in the mid to late 80s by Tanner, Alon–Milman, Alon–Chung, Thomason, Chung–Graham–Wilson.

The setting will be a bipartite graph with \( k \) vertices on the “left” and \( \ell \) vertices on the “right.” We define the bipartite adjacency matrix of such a graph as a \( k \times \ell \)(0, 1)-matrix.

We put

\[
a_{ij} = \begin{cases} 
1 & \text{if } i \sim j \text{ (i on the left, j on the right)} \\
0 & \text{otherwise.}
\end{cases}
\]

Now, \( A^T A \) is a symmetric matrix (\((A^T A)^T = A^T A\)). Thus, its eigenvalues are real, and we may write them as \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell \). Moreover, \( A^T A \) is positive semidefinite, i.e., \( (\forall x \in \mathbb{R}^\ell)(x^T A^T A x \geq 0) \). Indeed, because \( x^T A^T A x = \|A x\|^2 \geq 0 \). It follows that \( \lambda_i \geq 0 \) for all \( i \).

Assume further that our graph is biregular: all vertices on the left have the same degree \( = \frac{|E|}{k} \), and all vertices on the right have the same degree \( = \frac{|E|}{\ell} \).

#### Theorem 7.4.1. (Quasirandomness Theorem)

Suppose \( \Gamma \) is a biregular bipartite graph, and let \( X \) be a subset of the left side and \( Z \) a subset of the right side. Put \( p = \frac{|E|}{k\ell} \) (the density of \( \Gamma \)). Let \( \lambda_i \) the \( i \)-th eigenvalue of \( A^T A \), in decreasing order. Then

\[
|E(X, Z)| - p|X| \cdot |Z| \leq \sqrt{\lambda_2} |X| \cdot |Z|.
\] (3)

This is very similar to Lindsey’s lemma, which is not a surprise since both statements have a “quasirandomness” flavor. Their proof also depends on an essentially identical application of Cauchy-Schwarz. The inequality above is saying that if \( \lambda_2 \) is small, then \( E(X, Z) \) will be close to what would be expected if the edges were thrown in at random.

#### Proof of the Quasirandomness theorem.

We want to estimate \( E(X, Z) \) for a biregular bipartite graph. We have \(|E(X, Z)| = x^T A z\), and \(|X| \cdot |Z| = x^T J_{k,\ell} z\), where \( J := J_{k,\ell} \) is the \( k \times \ell \)-matrix of all 1’s; and \( x \) is the incidence vector of \( X \) and \( z \) is the incidence vector of \( Z \).

#### Notation 7.4.2.

From now on, if \( J \) appears with no subscript, it denotes \( J = J_{k,\ell} \). Otherwise, \( J_{m \times n} \) is the \( m \times n \)-matrix with all ones.

So the quantity on the left of (3) may be rewritten as

\[
|E(X, Z)| - p|X| \cdot |Z| = |x^T (A - pJ) z|,
\] (4)

which looks eerily similar to what we were looking at in the case of Lindsey’s lemma. Again, we apply our “Cauchy-Schwarz” gun and obtain

\[
|x^T (A - pJ) z| \leq \|x\| \cdot \|(A - pJ) z\| = \sqrt{|X| \cdot \|(A - pJ) z\|}.
\] (5)
Since \( \|z\| = \sqrt{z^Tz} \), it remains to prove that
\[
\|(A - pJ)z\| \leq \sqrt{\lambda_2} \|z\|. \tag{6}
\]
In other words, we have to show that \( \|(A - pJ)z\|^2 \leq \lambda_2 \cdot \|z\|^2 \). We have
\[
\|(A - pJ)z\|^2 = z^T(A^T - pJ^T)(A - pJ)z. \tag{7}
\]
Also,
\[
\]
Now, we assumed that the sum of every column, \( s_c \) was the same. So, \( s_c = \frac{|E|}{\ell} = \frac{pk\ell}{\ell} = pk \).
We have that
\[
J^T A = pk \cdot J_{\ell \times \ell}. \tag{9}
\]
Also, \( A^T J = (J^T A)^T = (pk J_{\ell \times \ell})^T = pk J_{\ell \times \ell} \) (since \( J_{\ell \times \ell} \) is symmetric). Since \( J^T J = k J_{\ell \times \ell} \), we may rewrite the RHS of (8), obtaining
\[
(A^T - pJ^T)(A - pJ) = A^T A - p^2 k J_{\ell \times \ell}. \tag{10}
\]
Now, let \( s_r = \frac{|E|}{k} = p\ell \) be the sum of each row of \( A \). We have
\[
A1_\ell = s_r \cdot 1_k, \quad A^T 1_k = s_c \cdot 1_r. \tag{11}
\]
Thus, we see that
\[
A^T A1_\ell = s_r \cdot s_c \cdot 1_\ell. \tag{12}
\]
In other words, \( 1_\ell \) is an eigenvector of \( A^T A \) to the eigenvalue \( \lambda_1 = s_r \cdot s_c = p^2 k\ell \). Furthermore, we have \( J_{\ell \times \ell}1_\ell = \ell 1_\ell \), and \( (p^2 k J_{\ell \times \ell}) \cdot 1_\ell = p^2 k\ell \cdot 1_\ell = \lambda_1 \cdot 1_\ell \).

Now, set \( M := A^T A - p^2 k J_{\ell \times \ell} \), the RHS of (10). We conclude that \( M \cdot 1_\ell = 0 \) and \( M = M^T \). By the Spectral Theorem, there exists an orthogonal eigenbasis \( e_i \) to \( M \), so that \( Me_i = \mu_i e_i \). We may assume that \( e_1 = 1_\ell \); therefore \( 1_\ell \) is perpendicular to \( e_i \) for \( i = 2, \ldots, \ell \). It follows that \( J_{\ell \times \ell} e_i = 0 \) for \( i = 2, \ldots, \ell \). Therefore for \( i \geq 2 \) we have \( (A^T A - p^2 k J_{\ell \times \ell}) e_i = A^T A e_i \) for \( i \geq 2 \), so \( \mu_i = \lambda_i \) for \( i \geq 2 \). As a consequence, the maximal eigenvalue of \( M \) is \( \lambda_2 \) (NOT \( \lambda_1 \)). By the Rayleigh principle, the largest eigenvalue is \( \max_{z} \frac{z^T M z}{z^T z} \). So we have \( z^T M z \leq \lambda_2 z^T z \).

We needed to show that \( \|(A - pJ)z\|^2 \leq \lambda_2 \|z\|^2 \). But, \( \|(A - pJ)z\|^2 = z^T M z \leq \lambda_2 \|z\|^2 \).

We see that all we needed to finish is that \( \lambda_2 \) is the largest eigenvalue corresponding to vectors orthogonal to the all-ones vector.

We remark, although our proof did not need it, that \( \lambda_1 \) is indeed larger than \( \lambda_2 \), by the following exercise (completes the proof as stated):

**Exercise 7.4.3.** Suppose \( A = (a_{ij}) \) is a nonnegative \( n \times n \)-matrix. Suppose also that \( v \) is an all-positive eigenvector: \( Av = \rho v \). Let \( \lambda \) be any (complex) eigenvalue. Then \( |\lambda| \leq \rho \).