In this lecture and the next one we will study a special class of regular graphs called Cayley graphs. The eigenvalues and eigenvectors of the adjacency matrix of a Cayley graph have a particularly nice form.

Today we will study Cayley graphs over the group $\mathbb{Z}_2^n$ and their connection to small-biased sets. We will then see how small-biased sets are used in a beautiful algorithm of Prasad Raghavendra for solving linear equations modulo 2.

# 1 Cayley graphs, Abelian groups, and small-biased sets

Recall that a *group* is a set with an operation that is associative ($(ab)c = a(bc)$ for all $a, b, c$) with an identity (there is an element 1 such that $a1 = 1a = a$ for every $a$) and inverses for all elements (for every $a$ there is an $a^{-1}$ such that $aa^{-1} = a^{-1}a = 1$). A set $S$ of group elements is called a *generating set* if every element of $G$ can be written as a finite product of elements in $S$. Here we will only worry about finite groups.

Let $S$ be a generating set for a group $G$ which is closed under inverse (for every $a$ in $S$, $a^{-1}$ is also in $S$). The *Cayley graph* $\mathrm{Cay}(G, S)$ is the $|S|$-regular graph where there is a vertex for every element in $G$ and an edge from $g$ to $sg$ for every $g \in G$ and $s \in S$. (Parallel edges and loops are allowed.) Since $S$ is closed under inverse, this graph is undirected.

We are interested in constructing infinite families of Cayley graphs which are expanding. It is common to start with a specific family of groups $\{G_n\}$ and try to construct a set of generators $S_n$, $|S_n| \leq d$ for $G_n$ so that $\lambda(\mathrm{Cay}(G_n, S_n)) \leq 1 - \varepsilon$.

To illustrate the connection between the algebra of the groups $G_n$ and the expansion of the corresponding Cayley graphs let us start with some groups we already have some experience with, namely $G_n = \mathbb{Z}_2^n$. Unfortunately it will not be possible to obtain expander families out of these groups. Nevertheless, they will serve as a good introduction to Cayley graphs.

Let $S = \{s_1, \ldots, s_d\}$ be a subset of $\mathbb{Z}_2^n$. (In $\mathbb{Z}_2^n$ every element is its own inverse, so $S$ is automatically closed under inverse.) Notice that $S$ is a generating set for $\mathbb{Z}_2^n$ if and only if the rank of $s_1, \ldots, s_d$ viewed as vectors in $\mathbb{Z}_2^n$ is $n$, which is only possible if $d \geq n$. So it is not even possible to generate $\mathbb{Z}_n$ with a number of elements independent of $n$, much less make it into an expanding family. But let us anyway try to answer the following question:

> How small can $\lambda = \lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S))$ get among all sets $S$ of size $d$?

We saw that $\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) = 1$ unless $d \geq n$, so let's see what happens when $d$ becomes larger. Last time we showed that

$$\lambda^t \geq \|\mathbf{p}^t - \mathbf{u}\|$$

for every $t > 0$, where $\mathbf{p}^t$ is the distribution of a random walk after $t$ steps starting from some

vertex $s$. Because of commutativity, after $t = \alpha d$ steps the random walk can reach at most

$$\binom{d}{0} + \binom{d}{1} + \cdots + \binom{d}{t} \leq 2^{dH(\alpha)}$$

vertices. Let us choose $t \leq d/2$ so that $dH(\alpha) = n - 1$. Then at least $2^{n-1}$ of the vertices have not been reached with $t$ steps and

$$\|\mathbf{p}^t - \mathbf{u}\| \geq \sqrt{2^{n-1} \cdot (0 - 2^{-n})^2} = 2^{-(n+1)/2}.$$

Therefore

$$\lambda \geq 2^{-(n+1)/2t} = 2^{-(n-1)/2t - 1/t} = 2^{-H(\alpha)/2\alpha - 1/t}.$$

Applying the upper bound $H(\alpha) \leq \alpha \log_2(1/\alpha) + O(\alpha)$, we obtain

$$\lambda \geq 2^{-\log_2(1/\alpha)/2 - O(1) - O(1/t)} = \Omega(\sqrt{\alpha}).$$

from where $H(\alpha) = O(\lambda^2 / \log(1/\lambda))$, and so $d = \Omega(n/\lambda^2 \log(1/\lambda))$. This bound is tight up to the $\Omega(\log(1/\lambda))$ factors by the following lemma:

**Lemma 1.** $\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) = \max_{a \neq 0} |\mathrm{E}_{s \sim S}[\chi_a(s)]|$.

This equation says that $\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) \leq \delta$ if and only if the set $S$ is $\delta$-biased.[1] In Lecture 2 we showed we can achieve $|S| = O(n/\delta^2)$, and we just saw that it is necessary to have $|S| = \Omega(n/\delta^2 \log(1/\delta))$. It is not known if the logarithmic factor is necessary.

*Proof.* Let $A$ be the normalized $2^n \times 2^n$ adjacency matrix of $\mathrm{Cay}(\mathbb{Z}_2^n, S)$. Then we can write

$$A = \frac{1}{d} \sum_{s \in S} A_s$$

where $A_s(g, h) = 1$ if $h = s + g$ (using additive notation for the operation in $\mathbb{Z}_2^n$, and 0 otherwise).

A very nice property of abelian groups is that all the matrices $A_s$ have the same eigenvectors, and so these must also be the eigenvectors of $A$. The $2^n$ eigenvectors of $A_s$ are the character functions $\chi_a$ viewed as vectors whose entries are indexed by $\mathbb{Z}_2^n$:

$$(\chi_a A_s)(h) = \sum_g \chi_a(g) A_s(g, h) = \chi_a(s + h) = \chi_a(s)\chi_a(h)$$

because the only $g$ for which $A_s(g, h)$ is nonzero is $g = s + h$. So $\chi_a$ is an eigenvector of $A_s$ with eigenvalue $\chi_a(s)$, and by linearity $\chi_a$ is an eigenvector of $A$ with eigenvalue

$$\lambda_a = \frac{1}{d} \sum_{s \in S} \chi_a(s) = \mathrm{E}_{s \sim S}[\chi_a(s)]$$

This gives us a formula for all $2^n$ eigenvalues of $A$. When $a = 0$, $\lambda_0 = 1$, and so

$$\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) = \max_{a \neq 0} |\mathrm{E}_{s \sim S}[\chi_a(s)]|. \qquad \square$$

Getting back to our objective of constructing an expanding family of Cayley graphs over $\mathbb{Z}_2^n$, we see that this is impossible as $\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) = \Omega(\sqrt{d/n})$. The situation is similar over other Abelian groups, and to make this approach work we have to turn to non-Abelian groups.

[1] What we mean more precisely is that the uniform distribution over the elements of $S$ is $\delta$-biased. In general, $S$ can be a multiset, in which case each element is chosen with probability proportional to its multiplicity.

## 2 An algorithm for linear equations modulo 2

Consider the following algorithm for solving linear equations modulo 2. The algorithm maintains a set $S$ of candidate solutions. At each step, the algorithm looks at a new equation in the system and updates the set $S$. Initially, let $S = \{0,1\}^n$; that is, $S$ contains all possible solutions. After looking at the first equation, we keep those $s$ in $S$ that satisfy the equation and throw away those that do not satisfy it. We continue with the rest of the equations one by one. After we have looked at all the equations, what is left in $S$ are only the solutions to the system.

This algorithm is not efficient as the set $\{0,1\}^n$ is exponentially large. We want to run the same algorithm, but on a much smaller set $S$ of candidate solutions. Let us try to specify the properties we want $S$ to have.

1. At each point, all $s \in S$ satisfy all previously seen equations.
2. Upon seeing a new equation $\langle a, x\rangle = b$, $S$ should contain at least one solution to it, provided that it is consistent with the previous equations.

We start with the first equation. Since we do not know what this equation will be ahead of time, we must ensure that $S$ initially contains solutions to all possible equations. This brings to mind a small-biased set: Every linear equation is satisfied by about half the elements of a small-biased set. The following generalization of $\delta$-biased sets will be useful.

**Definition 2.** Let $A$ be a linear subspace of $\{0,1\}^n$. A set $S$ over $\{0,1\}^n$ is $\delta$-biased over $A$ if for every $a \in A$, $|\mathrm{E}_{s\sim S}[\chi_a(s)]| \le \delta$.

Just as in the special case $A = \mathbb{F}_2$, $S$ is $\delta$-biased over $A$ if and only if $\lambda(\mathrm{Cay}(A,S)) \le \delta$.

Let us initially set $S$ to be a $\delta$-biased set over $\{0,1\}^n$. After looking at the first equation $(a_1, b_1)$, we throw out those $x \in S$ that fail to satisfy it – about half of them. We expect to be left with a small-biased set over the linear space $A_1$ of equations orthogonal to $a_1$. How does the discarding affect the bias of the other equations $a \in A_1$? To answer this question we use this formula:

**Claim 3.** *Let $X$ and $Y$ be $\{-1,1\}$-valued random variables. Then for every value $x \in \{-1,1\}$,*

$$\mathrm{E}[Y \mid X = x] = \frac{\mathrm{E}[Y] + x\,\mathrm{E}[XY]}{1 + x\,\mathrm{E}[X]}.$$

If $X = (-1)^{\langle a_1, x\rangle}$ and $Y = \langle a, x\rangle$, then because $S$ is $\delta$-biased we get that $|\mathrm{E}[X]|, |\mathrm{E}[Y]|, |\mathrm{E}[XY]| \le \delta$ and so

$$\mathrm{E}_{x\sim S}[(-1)^{\langle a,x\rangle} \mid \langle a_1, x\rangle = b_1] \le \frac{2\delta}{1-\delta}. \tag{1}$$

This suggests that after we discard the incorrect solutions to the first equation, the bias of $S$ with respect to the other equations may grow by a factor of about two. Thus after seeing about $\log(1/\delta)$ equations, $S$ may fail to contain any feasible solutions to the remaining equations. The following lemma gives a way to reduce the bias of $S$:

**Lemma 4.** *If $S$ is $\delta$-biased over $A$, then the (multi)set $S^t$ consisting of $s_1 + \cdots + s_t$, where $s_1, \ldots, s_t \in S$, is $\delta^t$-biased over $A$.*

*Proof.* For every $a \in A$,

$$|\mathrm{E}[\chi_a(s_1 + \cdots + s_t)]| = |\mathrm{E}[\chi_a(s_1) \cdots \chi_a(s_t)]| = |\mathrm{E}[\chi_a(s_1)]| \cdots |\mathrm{E}[\chi_a(s_t)]| \leq \delta^t. \qquad \square$$

In the language of Cayley graphs, this lemma says that $\lambda(\mathrm{Cay}(A, S^t)) = \lambda(\mathrm{Cay}(A, S))^t$. This statement has a graph-theoretic interpretation (and alternate proof). A random walk on $\mathrm{Cay}(A, S)$ starting at a vertex $a$ chooses $s$ at random from $S$ and moves to the vertex $a + s$. A random walk on $\mathrm{Cay}(A, S^t)$ takes $t$ random steps out of $a$. Thus $\mathrm{Cay}(A, S^t)$ is the graph one obtains after taking $t$ steps in the random walk. Its adjacency matrix is the $t$-th power of the adjacency matrix of $\mathrm{Cay}(A, S)$. Therefore the eigenvalues of $\mathrm{Cay}(A, S^t)$ are the $t$-th power of the eigenvalues of $\mathrm{Cay}(A, S)$.

So to reduce the bias of $S$, we can replace its entries with their $t$-wise sums $s_1 + \cdots + s_t$. But what is the effect of this transformation on condition 1? Suppose that we are looking at an equation $\langle a, x \rangle = b$ that is a linear combination of previous equations. By assumption, all $s_1, \ldots, s_t \in S$ will satisfy this equation. But then

$$\langle a, s_1 + \cdots + s_t \rangle = \langle a, s_1 \rangle + \cdots + \langle a, s_t \rangle = tb \tag{2}$$

and we see that in general, $S^t$ will satisfy this equation only when $t$ is odd! The smallest value of $t$ that both reduces the bias and does not violate previously satisfied equations is $t = 3$.

We now have almost all the elements of the algorithm. Starting with a $\delta$-biased set $S$ for a suitably chosen constant $\delta$, we look at the equations one by one. After seeing a new equation, we discard the entries of $S$ that are not solutions to this equation. Then we replace $S$ with the multiset $\{s_1 + s_2 + s_3 \colon s_1, s_2, s_3 \in S\}$. After we have exhausted all the equations, we return any $s$ from $S$ as a solution.

The running time of this algorithm is determined by the size of $S$. The discarding step reduces the size of $S$ by about a factor of two. The bias reduction step increases the size of the step by a cubic rate. Thus the dominant effect comes from bias reduction and the size of $S$ grows very quickly. The last ingredient we will need is a method for reducing the size of $S$ without affecting its bias too much. The proof is given at the end of the section.

**Lemma 5.** *Let $S$ be a $\delta$-biased set where $\delta \leq 1/2$. Let $S'$ be a random sample consisting of $O(n/\varepsilon^2)$ entries of $S$ (chosen with repetition). With probability at least $1 - 2^n$, $S$ is $(\delta + \varepsilon)$-biased.*

We can now state and analyze Raghavendra's algorithm:

On input $(a_1, b_1), \ldots, (a_m, b_m)$, where $a_1, \ldots, a_m \in \{0,1\}^n, b_1, \ldots, b_m \in \{0,1\}$:
1. Let $S$ be a $\frac{1}{5}$-biased multiset of size $O(n)$ over $\{0,1\}^n$.
2. For $i = 1$ to $m$ repeat the following:
3. Discard all $s \in S$ such that $\langle a_i, s \rangle \neq b_i$.
4. Replace $S$ by the multiset $\{s_1 + s_2 + s_3 \colon s_1, s_2, s_3 \in S\}$.
5. Keep $O(n)$ entries of $S$ chosen at random with repetition and discard the rest.
6. If $S$ is non-empty, output any $s \in S$.
7. Otherwise, output `inconsistent`.

The running time of the algorithm is $O(n^4 m)$. The factor of $O(n^4)$ comes from step 4. By combining steps 4 and 5 into a single step that randomly samples $O(n)$ random vectors of the form $s_1 + s_2 + s_3$, this improves to $O(n^2)$, giving an overall running time of $O(n^2 m)$. We now show correctness.

**Theorem 6.** *With probability at least $1 - m2^{-n}$, Raghavendra's algorithm outputs $s$ such that $\langle a_i, s\rangle = b_i$ for all $i$ if such $s$ exists, and* `inconsistent` *otherwise.*

Let $A_i$ denote the subspace of $\{0,1\}^n$ orthogonal to $a_1, \ldots, a_i$ and let $S_i$ indicate the state of the set $S$ after the $i$-th iteration. Theorem 6 will follow from these two lemmas:

**Lemma 7.** *For every $1 \le i \le m$, all $s \in S_i$ satisfy the first $i$ equations $\langle a_1, s\rangle = b_1, \ldots, \langle a_i, s\rangle = b_i$.*

**Lemma 8.** *For every $1 \le i \le m$, if the first $i$ equations are consistent then with probability $1 - i2^{-n}$, $S_i$ is non-empty and $\frac{1}{5}$-biased over $A_i$.*

In particular, after the last iteration, Lemma 7 guarantees that all $i \in S$ satisfy all the equations, and Lemma 8 says that there is at least one solution in $S$ provided the equations are consistent.

The proofs of the two lemmas are by induction on $i$.

*Proof of Lemma 7.* The base case $i = 0$ holds trivially. Now assume the statement holds for $i - 1$, so the first $i - 1$ equations are satisfied for all $s \in S = S_{i-1}$. After step 4, the $i$-th equation is also satisfied. Step 4 preserves satisfiability by (2), and step 5 only discards elements of $s$. Therefore by the end of the $i$-th iteration, $S = S_i$ satisfies the first $i$ equations. □

*Proof of Lemma 8.* Initially, $S = S_0$ is non-empty and $\frac{1}{5}$-biased, so the base case holds. Now assume the statement holds for $i-1$ and the first $i$ equations are consistent. By inductive assumption, with probability $1 - (i-1)2^{-n}$, the set $S = S_{i-1}$ is non-empty and $\frac{1}{5}$-biased over $A_{i-1}$. By (1), after step 3 $S$ is $2 \cdot \frac{1}{5}/(1 - \frac{1}{5}) = \frac{1}{2}$-biased over $A_i$. By Lemma 4, after step 4 $S$ is $\frac{1}{2}^3 = \frac{1}{8}$-biased over $A_i$. By Lemma 5 with $\varepsilon = \frac{3}{40}$, with probability $1 - 2^{-n}$, after step 5, $S_i$ is $\frac{1}{8} + \frac{3}{40} = \frac{1}{5}$-biased. Clearly $S_i$ is also non-empty. By the union bound, this is true with probability at least $1 - i2^{-n}$. □

### Missing proofs

*Proof of Claim 3.* Using the formula for conditional expectations, we have

$$\mathrm{E}[Y] = \mathrm{E}[Y \mid X = 1]\Pr[X = 1] + \mathrm{E}[Y \mid X = -1]\Pr[X = -1]$$
$$\mathrm{E}[XY] = \mathrm{E}[Y \mid X = 1]\Pr[X = 1] - \mathrm{E}[Y \mid X = -1]\Pr[X = -1].$$

Therefore

$$\mathrm{E}[Y \mid X = 1] = \frac{\mathrm{E}[Y] + \mathrm{E}[XY]}{2\Pr[X = 1]} = \frac{\mathrm{E}[Y] + \mathrm{E}[XY]}{1 + \mathrm{E}[X]}$$

and

$$\mathrm{E}[Y \mid X = -1] = \frac{\mathrm{E}[Y] - \mathrm{E}[XY]}{2\Pr[X = -1]} = \frac{\mathrm{E}[Y] - \mathrm{E}[XY]}{1 - \mathrm{E}[X]}.$$

□

*Proof of Lemma 5.* Let $f$ be any function of the form $\chi_a$ of $-\chi_a$ where $a \neq 0$. Since $S$ is $\delta$-biased, $\mathrm{E}_{s\sim S}[f(s)] \le \delta$. By the Chernoff bound, for independently chosen $s_1, \ldots, s_m \sim S$:

$$\Pr[f(s_1) + \cdots + f(s_m) > \delta m + \varepsilon m] \le e^{-2\varepsilon^2 m}.$$

Taking a union bound over all $2(2^n - 1) \le 2^{n+1}$ choices of $f$, we get

$$\Pr[f(s_1) + \cdots + f(s_m) > \delta m + \varepsilon m \text{ for some } f \in \{\chi_a, -\chi_a \colon a \neq 0\}] \le 2^{n+1} \cdot e^{-2\varepsilon^2 m}.$$

If we set $m = \lceil (n + \frac{1}{2})/\varepsilon^2 \ln 2 \rceil$, this probability is at most $2^{-n}$, proving the lemma. □