Super-approximation
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 What is super-approximation? 1st try.</td>
<td>5</td>
</tr>
<tr>
<td>1.1 Rough description of strong approximation</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Expanders and super-approximation</td>
<td>7</td>
</tr>
<tr>
<td>1.3 Exercises</td>
<td>8</td>
</tr>
<tr>
<td>2 Random-walks on a graph and expanders</td>
<td>13</td>
</tr>
<tr>
<td>2.1 Basics of random-walks on a finite graph</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Discrete Laplacian</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Finding good cuts</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Discrete isoperimetric inequalities</td>
<td>20</td>
</tr>
<tr>
<td>2.5 Exercises</td>
<td>22</td>
</tr>
</tbody>
</table>
Chapter 1

What is super-approximation? 1st try.

1.1 Rough description of strong approximation

To understand the origin of the phrase *super-approximation*, we start with briefly formulating *strong approximation*. Let’s start with the case of $	ext{SL}_2$, the set of two-by-two matrices with entries in a given unital commutative ring and determinant 1. Strong approximation addresses questions of the following form.

**Question 1.** Does the residue module $n$ map $\pi_n : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ induces a surjective map from $\text{SL}_2(\mathbb{Z})$ to $\text{SL}_2(\mathbb{Z}/n\mathbb{Z})$?

In Exercise 3, you can find one approach for giving an affirmative answer to this question. Notice that for every unital commutative ring $R$, $\text{SL}_2(R)$ can be identified with $V(R) := \{(a, b, c, d) \in R^4 | ad - bc = 1\}$. Question 1 is equivalent to asking if every solution of $ad - bc = 1$ in $\mathbb{Z}/n\mathbb{Z}$ has a lift to a solution of this equation in $\mathbb{Z}$.

One can think about Question 1 in terms of transitivity of certain subgroups of the group of automorphism of $V$ as well. As you can see in Exercise 2, $\text{SL}_2(\mathbb{Z})$ is generated by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. This means

$$(a, b, c, d) \mapsto (a \pm b, b, c \pm d, d) \quad \text{and} \quad (a, b, c, d) \mapsto \pm(b, -a, d, -c)$$

induce a transitive action on $V(\mathbb{Z})$. The strong approximation is equivalent to saying that these maps induce a transitive action on $V(\mathbb{Z}/n\mathbb{Z})$ for every positive integer $n$.

Using the reduced row echelon process, one can show a similar result for $\text{SL}_m(\mathbb{Z})$ for $m \geq 3$. This method is essentially based on using unipotent elements ($u$ is called unipotent if all of its eigenvalues are 1). Following the same ideas, one can prove a similar result for symplectic groups. Let’s recall that for every unital commutative ring $R$,

$$\text{Sp}_{2n}(R) = \left\{ \gamma \in \text{SL}_{2n}(R) \mid \gamma \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \gamma^t = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \right\}.$$ 

This means $\pi_m : \text{Sp}_{2n}(\mathbb{Z}) \to \text{Sp}_{2n}(\mathbb{Z}/m\mathbb{Z})$ is surjective.
Next, we discuss that a similar statement does not hold for \( \text{PGL}_2 \). Let’s define \( \text{PGL}_2(R) := \text{GL}_2(R)/R^\times I \) where \( R^\times \) is the group of units of \( R \). Then
\[
\overline{\det} : \text{PGL}_2(R) \to R^\times/(R^\times)^2, \quad \overline{\det}(\gamma R^\times I) := \det(\gamma)(R^\times)^2
\]
is a well-defined surjective group homomorphism. Hence for every two distinct primes \( p \) and \( q \), \( \pi_{pq} : \text{PGL}_2(\mathbb{Z}) \to \text{PGL}_2(\mathbb{Z}/pq\mathbb{Z}) \) is not surjective as
\[
|\mathbb{Z}^\times| = 2 \quad \text{and} \quad |(\mathbb{Z}/pq\mathbb{Z})^\times/((\mathbb{Z}/pq\mathbb{Z})^\times)^2| = 4.
\]
In technical terms, the big difference between \( \text{SL}_2 \) and \( \text{PGL}_2 \) is that \( \text{SL}_2 \) is simply-connected and \( \text{PGL}_2 \) is not. Notice that for every algebraically closed field \( F \),
\[
1 \to \mu_2(F) \to \text{SL}_2(F) \to \text{PGL}_2(F) \to 1
\]
is a short exact sequence where \( \mu_2 := \{ x \in R | x^2 = 1 \} \). Moreover \( \mu_2(F) \) is a finite central subgroup. A group homomorphism with these properties is called a central isogeny (at least in characteristic zero). A(n) (algebraic) group which does not have a non-trivial (algebraic) isogeny is called a simply-connected (algebraic) group. For instance, \( \text{SL}_n \) and \( \text{Sp}_{2n} \) are simply connected algebraic groups, but \( \text{PGL}_n \) is not a simply connected algebraic group. Here we take a rudimentary approach and say that an algebraic group is a group which consists of solutions of certain polynomial equations and the group operations can be given by polynomial maps. Now we can formulate a version of strong approximation (due to Eichler, Kneser, and Platonov):

**Theorem 2** (Strong approximation: the \( S \)-arithmetic case). Suppose \( G \) is a simply-connected algebraic group defined by polynomials with coefficients in \( \mathbb{Z} \). Suppose \( G(\mathbb{C}) \) is a product of almost simple groups (we say \( G \) is semisimple). Assume that \( G(\mathbb{Z}[1/q_0]) \) is an infinite group. Then for every integer \( n \) with large enough prime factors the residue modulo \( n \) congruence map
\[
\pi_n : G(\mathbb{Z}[1/q_0]) \to G(\mathbb{Z}/n\mathbb{Z})
\]
is surjective.

Next we want to see what happens if we restrict \( \pi_n \) to a subgroup \( \Gamma \) of \( G(\mathbb{Z}[1/q_0]) \). Can we still get the entire \( G(\mathbb{Z}/n\mathbb{Z}) \) (at least for integers \( n \) with large prime factors)?

We make one important observation: if there is an integer polynomial map \( f : \mathbb{Q}(\mathbb{C}) \to \mathbb{C} \) such that \( f(G(\mathbb{Z}[1/q_0])) \neq 0 \) but \( f(\Gamma) = 0 \), then it is not possible for \( \pi_p(\Gamma) = G(\mathbb{Z}/p\mathbb{Z}) \) to hold for an arbitrarily large prime \( p \). This is the case, because \( f(G(\mathbb{Z}[1/q_0])) \neq 0 \) implies that for a large enough prime \( p \), there is \( \lambda \in G(\mathbb{Z}[1/q_0]) \) such that \( \pi_p(f(\lambda)) \neq 0 \), and so \( f(\pi_p(\lambda)) \neq 0 \). On the other hand, \( f(\pi_p(\Gamma)) = 0 \). Therefore \( \pi_p(\lambda) \notin \pi_p(\Gamma) \). We refer to this type of limitations as an algebraic obstruction.

If we refer to common solutions of a family of polynomials as closed sets, then to avoid the above algebraic obstruction we have to assume that the smallest closed subset of \( G \) which contains \( \Gamma \) is \( G \). We refer to the topology given by these closed sets as the Zariski topology of \( G \). In this language, the latest condition can be phrased as \( \Gamma \) is Zariski-dense in \( G \). Now we can formulate a stronger version of the strong approximation (due to Weisfeiler).
1.2. EXPANDERS AND SUPER-APPROXIMATION

**Theorem 3** (Strong approximation: the Zariski-dense case). Suppose $G$ is a Zariski-connected simply-connected semisimple group given by integer polynomials. Suppose $\Gamma \subseteq G(\mathbb{Z}[1/q_0])$ is a Zariski-dense subgroup. Then for every integer $n$ with large enough prime factors the residue modulo $n$ congruence map $\pi_n : \Gamma \rightarrow G(\mathbb{Z}/n\mathbb{Z})$

is surjective.

### 1.2 Expanders and super-approximation

Suppose $G$ is a group and $\Omega$ is a subset of $G$. We say $\Omega$ is a symmetric subset if the inverse of every element of $\Omega$ is in $\Omega$. The Cayley graph of $G$ with respect to $\Omega$ is an undirected graph whose set of vertices is $G$ and $g_1, g_2 \in G$ are connected exactly when $g_1^{-1}g_2 \in \Omega$. The Cayley graph of $G$ with respect to $\Omega$ is denoted by $\text{Cay}(G; \Omega)$. Notice that $\text{Cay}(G; \Omega)$ is a $|\Omega|$-regular graph; this means that the degree of every vertex is $|\Omega|$. The set of neighbors of $g$ is $g\Omega = \{gw | w \in \Omega\}$. Continuing, we obtain that the connected component of $g$ is the set

$$\{gw_1 \cdots w_n | n \in \mathbb{Z}^+, w_1, \ldots, w_n \in \Omega\}.$$

Since $\Omega$ is symmetric, $\{w_1 \cdots w_n | n \in \mathbb{Z}^+, w_1, \ldots, w_n \in \Omega\}$ is the subgroup generated by $\Omega$. Hence $\text{Cay}(G, \Omega)$ is connected if and only if $\Omega$ is a generating set of $G$. Therefore the strong approximation is equivalent to saying that if $\Omega$ is a symmetric generating set of a Zariski-dense subgroup of $G(\mathbb{Z}[1/q_0])$, then under the right conditions on $G$ and $n$, $\text{Cay}(G(\mathbb{Z}/n\mathbb{Z}), \pi_n(\Omega))$ is a connected graph. Super-approximation is about whether these graphs are highly connected.

Next we formulate what it means for a family of graphs to be highly connected. One way of thinking about the well-connectivity is in terms of people who live in a society. A society is well-connected if it is not consist of two or more communities that are not well-integrated. This means what links these communities together is much less than their sizes.

We can quantify this using the **Cheeger constant** of a graph. The Cheeger constant of a finite graph $\mathcal{G}$ is

$$h(\mathcal{G}) := \min \left\{ \frac{|E(A, A^c)|}{\min\{|A|, |A^c|\}} \middle| A \subseteq V_{\mathcal{G}} \right\},$$

where $E(A, A^c)$ is the set of all the edges that connect a vertex in $A$ to a vertex in $A^c$. Notice that in a $k$-regular graph starting with a vertex $v_0$, the number of vertices that are of distance at most $n$ from $v_0$ is at least

$$\min\{|V_{\mathcal{G}}|/2, (1 + h(\mathcal{G}))/k^n\}.$$ 

This means these balls are expanding exponentially fast. Motivated by this, we say a family $\{\mathcal{G}_i\}_i$ of $k$-regular graphs is a family of expanders if and only if $\inf_i h(\mathcal{G}_i) > 0$; this means there is a uniform positive constant for the Cheeger constants of all of these graphs.
CHAPTER 1. WHAT IS SUPER-APPROXIMATION? 1ST TRY.

graphs. This implies that the number of vertices in balls of these graphs grow uniformly exponentially fast (till they contain at least half of the vertices).

A result of Selberg implies that \( \{ \text{Cay}(\text{SL}_2(\mathbb{Z}/n\mathbb{Z}), \Omega_i) \} \) is a family of expanders if \( i = 1, 2 \) and \( \gcd(n, i) = 1 \), where

\[
\Omega_i = \left\{ \begin{pmatrix} 1 & \pm i \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \pm i & 1 \end{pmatrix} \right\}.
\]

Selberg’s proof was based on the Kloosterman sum, and his result can be applied to every finitely generated congruence subgroup of \( \text{SL}_2(\mathbb{Z}) \). A subgroup of \( \text{SL}_2(\mathbb{Z}) \) is called a congruence subgroup if it contains \( \ker \pi_n \) for some \( n \), where \( \pi_n : \text{SL}_2(\mathbb{Z}) \to \text{SL}_2(\mathbb{Z}/n\mathbb{Z}) \) is the residue modulo \( n \) congruence map. The group generated by \( \Omega_1 \) is \( \text{SL}_2(\mathbb{Z}) \), and as you can see in Exercise 4, the subgroup generated by \( \Omega_2 \) contains the kernel of \( \pi_4 \).

The group generated by \( \Omega_3 \), however, is of infinite index in \( \text{SL}_2(\mathbb{Z}) \) (see Exercise 8), and so it cannot be a congruence subgroup. Notice that the Zariski-closure of the group generated by \( \Omega_3 \) contains the group generated by \( \Omega_1 \), and so it is Zariski-dense in \( \text{SL}_2(\mathbb{Z}) \). Peter Sarnak refer to this type of groups as thin groups; that means a thin group is a Zariski-dense subgroup of infinite index in \( G(\mathbb{Z}) \) (or more generally \( G(\mathbb{Z}[1/q_0]) \)) for some algebraic group \( G \). Since the group generated by \( \Omega_3 \) is Zariski-dense in \( \text{SL}_2(\mathbb{Z}) \), by the strong approximation, for every large enough prime \( p \) (in fact it is enough to assume that \( p \geq 5 \)), \( \text{Cay}(\text{SL}_2(\mathbb{Z}/p\mathbb{Z}), \Omega_3) \) is connected. Lubotzky asked whether these graphs form a family of expanders. This is referred to Lubotzky’s 1-2-3 problem, and Bourgain and Gamburd in their seminal work gave an affirmative answer to this question.

**Theorem 4** (Bourgain–Gamburd). Suppose \( \Omega \) is a finite symmetric subset of \( \text{SL}_2(\mathbb{Q}) \). Let \( \Gamma \) be the group generated by \( \Omega \). Suppose \( \Gamma \) is Zariski dense in \( \text{SL}_2(\mathbb{Q}) \). Then there is \( p_0 \) such that the family of graphs \( \{ \text{Cay}(\text{SL}_2(\mathbb{Z}/p\mathbb{Z}), \pi_p(\Omega)) \mid p \geq p_0, \text{p prime} \} \) is a family of expanders.

We refer to results of this type as super-approximation. The main goals of these notes are to cover the relevant general strategies, go over the type of tools involved, and survey the best known super-approximation results. This comes with the cost of not going into the details of most of the proofs.

1.3 Exercises

1. (Continued fraction) For a sequence of numbers \( \{ b_i \}_{i=0}^{\infty} \), we use \( [b_0; b_1, \ldots, b_m] \) to denote

\[
b_0 + \cfrac{1}{b_1 + \cfrac{1}{\cdots + \cfrac{1}{b_{m-1} + \cfrac{1}{b_m}}}}
\]

and \( [b_0; b_1, \ldots] \) to denote \( \lim_{m \to \infty} [b_0; b_1, \ldots, b_m] \) (if this limit exists).
1.3. EXERCISES

a) For a sequence of non-zero real numbers \( \{b_i\}_{i=0}^{\infty} \), suppose
\[
\begin{pmatrix} b_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} r_n(x) \\ s_n(x) \end{pmatrix}.
\]
Prove that \( \frac{r_n(x)}{s_n(x)} = [b_0; b_1, \ldots, b_n, x] \). (Hint: use induction on \( n \).)

b) For a sequence of non-zero integers \( \{b_i\}_{i=0}^{\infty} \), let \( p_{-1} = 1 \), \( q_{-1} = 0 \),
\[
p_{n+1} := p_nb_{n+1} + p_{n-1} \quad q_{n+1} := q_nb_{n+1} + q_{n-1}
\]
for every non-negative integer \( n \). Prove that
\[
\begin{pmatrix} b_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix},
\]
for every non-negative integer \( n \). Deduce that \( \frac{p_n}{q_n} = [b_0; b_1, \ldots, b_n] \),
\[
p_nq_{n-1} - p_{n-1}q_n = (-1)^{n+1}, \text{ and } \gcd(p_n, q_n) = 1.
\]

c) For a sequence of positive integers \( \{b_i\}_{i=0}^{\infty} \), suppose \( \frac{p_n}{q_n} \) is the simple form of the rational number \([b_0; b_1, \ldots, b_n]\). Use the previous part to show that
\[
\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n+1}}{q_{n-1}q_n},
\]
and deduce that \( \lim_{n \to \infty} \frac{p_n}{q_n} \) exists, and so \([b_0; b_1, \ldots] \) is well-defined.

d) For a non-zero real number \( x \), we let \( x_0 := x \) and define the sequences \( \{a_i\}_{i=0}^{\infty} \) and \( \{x_i\}_{i=0}^{\infty} \) inductively as follows. We set \( a_i := [x_i] \) for every integer \( i \) and \( x_{i+1} := \frac{1}{(x_i)} \) where \( \{y\} := y - \lfloor y \rfloor \) is the fractional part of \( y \).
We stop if \( x_i \) is an integer. Suppose \( \frac{p_n}{q_n} \) is the simple form of \([a_0; a_1, \ldots, a_n]\).
Show that \( x = [a_0; a_1, \ldots, a_n, x_{n+1}] \) for every non-negative integer \( n \).

e) In the setting of the previous item, prove that
\[
x = \frac{p_nx_{n+1} + p_{n-1}}{q_nx_{n+1} + q_{n-1}},
\]
and deduce that
\[
x - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(q_nx_{n+1} + q_{n-1})}.
\]

f) In the above setting, prove that \( x = [a_0; a_1, \ldots] \), and
\[
\frac{1}{q_n(q_n + q_{n+1})} \leq |x - \frac{p_n}{q_n}| \leq \frac{1}{q_nq_{n+1}}.
\]

g) For an irrational number \( \alpha = [a_0; a_1, \ldots] \), let
\[
M(\alpha) := \lim_{n \to \infty} \sup[a_n; a_{n+1}, \ldots] + [0; a_{n-1}, \ldots, a_1].
\]
Prove that there are infinitely many rational numbers of simple form \( \frac{p}{q} \) such that
\[
|\alpha - \frac{p}{q}| \leq \frac{1}{M(\alpha)q^2}.
\]
h) (Hurwitz’s theorem) Prove that $M(\alpha) \geq \sqrt{5}$ for every rational number $\alpha$, and equality holds for the Golden ratio $[1; 1, 1, \ldots]$.

2. (Generating $\text{SL}_2(\mathbb{Z})$) Suppose $\gamma = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in \text{GL}_2(\mathbb{Z})$.

   a) Suppose $\frac{a}{b} = [c_0; c_1, \ldots, c_n]$. Then
   
   $$\gamma = \begin{pmatrix} c_0 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & e \\ 0 & \pm 1 \end{pmatrix}$$
   
   for some integer $e$.

   b) Prove that $\text{SL}_2(\mathbb{Z}) = \langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \rangle$.

3. (Strong approximation: the $\text{SL}_2(\mathbb{Z})$ case) Suppose $n$ is a positive integer and $\bar{\gamma} = \begin{pmatrix} \bar{a} \\ \bar{b} \end{pmatrix} \in \text{SL}_2(\mathbb{Z}/n\mathbb{Z})$.

   a) Prove that there are integers $a$ and $b$ such that $\pi_n(a) = \bar{a}$, $\pi_n(b) = \bar{b}$, and $\gcd(a, b) = 1$, where $\pi_n$ is the residue modulo $n$ congruence map.

   b) Prove that there are $\lambda \in \text{SL}_2(\mathbb{Z})$ and $\bar{e} \in \mathbb{Z}/n\mathbb{Z}$ such that
   
   $$\pi_n(\lambda)^{-1}\bar{\gamma} = \begin{pmatrix} 1 & \bar{e} \\ 0 & 1 \end{pmatrix}.$$

   c) Prove that $\pi_n : \text{SL}_2(\mathbb{Z}) \to \text{SL}_2(\mathbb{Z}/n\mathbb{Z})$ is surjective.

4. Let $\alpha := \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\beta := \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. Suppose $a$ is odd and $b$ is a non-zero even number. Let $v := \begin{pmatrix} a \\ b \end{pmatrix}$.

   a) (The reduction process) Prove that there is $l \in \mathbb{Z}$ such that
   
   $$\min\{\|\alpha^lv\|_\infty, \|\beta^lv\|_\infty\} < \|v\|_\infty.$$

   b) Prove that there is $\gamma \in \langle \alpha, \beta \rangle$ such that $\gamma v = \pm \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

   c) Prove that $\langle \alpha, \beta, -I \rangle = \ker \pi_2$ where $\pi_2 : \text{SL}_2(\mathbb{Z}) \to \text{SL}_2(\mathbb{Z}/2\mathbb{Z})$ is the residue modulo 2 congruence map.

   d) Prove that $\langle \alpha, \beta \rangle$ contains the kernel of $\pi_4$.

5. (Ping-pong lemma) Suppose $G$ is a group and it acts on a set $X$. Suppose $G_1$ and $G_2$ are two subsets of $G$, $|G_1| \geq 2$, and $|G_2| \geq 3$. Suppose $X_1$ and $X_2$ are two subsets of $X$ such that $X_1 \nsubseteq X_2$ and $X_2 \nsubseteq X_1$. Suppose

   $$(G_1 \setminus \{1\}) \cdot X_2 \subseteq X_1 \quad \text{and} \quad (G_2 \setminus \{1\}) \cdot X_1 \subseteq X_2.$$

   Prove that $(G_1 \cup G_2) \simeq G_1 * G_2$. 

6. Suppose \( a \geq 2 \). Let \( \alpha := \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \beta := \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}, G_1 := \langle \alpha \rangle, \) and \( G_2 := \langle \beta \rangle \).

Let \( X_1 := \{ (x,y) \in \mathbb{R}^2 | |x| \geq \frac{a}{2} |y| \} \) and \( X_2 := \{ (x,y) \in \mathbb{R}^2 | |x| \leq \frac{a}{2} |y| \} \).

a) Consider the natural linear action of \( SL_2(\mathbb{R}) \) on \( \mathbb{R}^2 \). Prove that \( (G_1 \setminus \{ I \}) \cdot X_2 \subseteq X_1 \) and \( (G_2 \setminus \{ I \}) \cdot X_1 \subseteq X_2 \).

b) Prove that \( \alpha \) and \( \beta \) freely generate a free subgroup of \( SL_2(\mathbb{R}) \).

7. Let \( \alpha := \pm \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} \) and \( \beta := \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \) be two elements of the group \( PSL_2(\mathbb{Z}) := SL_2(\mathbb{Z})/\{ \pm I \} \).

a) Use Exercise 2 to show that \( PSL_2(\mathbb{Z}) = \langle \alpha, \beta \rangle \) and deduce that there is a surjective group homomorphism from \( \mathbb{Z}/3\mathbb{Z} \ast \mathbb{Z}/2\mathbb{Z} \) to \( PSL_2(\mathbb{Z}) \).

b) Consider the Möbius group action of \( PSL_2(\mathbb{R}) \) on \( \mathbb{C} \cup \{ \infty \} \); that means
\[
\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az + b}{cz + d},
\]
(justify that it is a group action). Notice that
\[
\alpha \cdot z = -1 - \frac{1}{z}, \quad \alpha^{-1} \cdot z = -\frac{1}{z + 1}, \quad \text{and} \quad \beta \cdot z = -\frac{1}{z}.
\]

Let \( X_1 \) be the set of all positive irrational real numbers and \( X_2 \) be the set of all the negative irrational real numbers. Show that
\[
(\langle \alpha \rangle \setminus \{ I \}) \cdot X_1 \subseteq X_2 \quad \text{and} \quad (\langle \beta \rangle \setminus \{ I \}) \cdot X_2 \subseteq X_1.
\]

c) Prove that there is an isomorphism \( PSL_2(\mathbb{Z}) \simeq \mathbb{Z}/3\mathbb{Z} \ast \mathbb{Z}/2\mathbb{Z} \) which factors through \( \langle \alpha \rangle \ast \langle \beta \rangle \).

8. Prove that \( \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \) and \( \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \) generate a subgroup of infinite index in \( SL_2(\mathbb{Z}) \).
(Hint: Use Exercise 7.)
Chapter 2

Random-walks on a graph and expanders

2.1 Basics of random-walks on a finite graph

A random walk on a graph \( G \) is a sequence of random variables \( \{X_i\}_{i=0}^{\infty} \) with values on the set of vertices \( V_G \) of \( G \) such that, for very non-negative integer \( i \), \( X_{i+1} \) is chosen independently at random from the neighbors of \( X_i \). For every vertex \( v \),

\[
P(X_{i+1} = v) = \sum_{w \in V_G} P(X_i = w)P(w \to v).
\]

Here for every \( w \in V_G \), \( P(w \to v) = \frac{1}{d_w} \) if \( \{w, v\} \in E_G \) where \( \{w, v\} \in E_G \) = 1 if \( w \) is connected to \( v \) in \( G \) and it is zero otherwise and \( d_w \) is the degree of the vertex \( w \); that means the number of edges that have \( w \) as one of their vertices. Let \( \mu_i \) be the distribution of \( X_i \); that means

\[
\mu_i : V \to [0,1], \quad \mu_i(v) = P(X_i = v).
\]

Suppose the set of vertices \( V := V_G \) is \( \{v_1, \ldots, v_n\} \). Then \( \mathfrak{B} := \{\delta_{v_1}, \ldots, \delta_{v_n}\} \) is an orthonormal basis of \( L^2(V) \). For every function \( f \in L^2(V) \), \( \langle f \rangle \) denotes the row matrix \( (f(v_1) \cdots f(v_n)) \) and \( |f| \) denotes the transpose of \( \langle f \rangle \). Notice that \( f = \sum_{i=1}^{n} f(v_i) \delta_{v_i} \), and so \( |f| \) is simply the matrix representation of \( f \) with respect to the basis \( \mathfrak{B} \).

Let \( T \) be the transition matrix of the random-walk; that means the \( (i, j) \)-entry of \( T \) is equal to

\[
P(v_i \to v_j) = \frac{1}{d_{v_i}} \mathbb{1}_{\{v_i, v_j\} \in E_G}.
\]

Then by (2.1), we have

\[
\langle \mu_{i+1} \rangle = \langle \mu_i \rangle T,
\]

and so the probability law after \( l \) steps random-walk is given by \( \langle \mu_l \rangle = \langle \mu_0 \rangle T^l \). We can understand and compute powers of a matrix the best if it is diagonal or at least diagonalizable. We know that a symmetric matrix is diagonalizable. We notice that
We notice that $\bar{A} = \bar{D} \bar{G}^{-1} A \bar{G}$, where $\bar{G}$ is the diagonal matrix $\text{diag}(d_1, \ldots, d_n)$ and $A$ is the adjacency matrix of the graph; that means the $(i, j)$ entry is $1$ if $v_i$ is connected to $v_j$ and $0$ otherwise. Hence for every integer $l$, we have

$$T^l = D^{-l} A \bar{G} D^{-l} A \bar{G} \cdots D^{-l} A \bar{G}.$$ 

Therefore

$$T^l = D^{-l} \bar{M} D^{-l} A \bar{G} D^{-l} A \bar{G} \cdots D^{-l} \bar{M} D^{-l} A \bar{G} \cdots D^{-l} A \bar{G} \cdots D^{-l} \bar{M} D^{-l} A \bar{G}.$$ 

We notice that $M \bar{G}$ is a real symmetric. Hence it has a right orthonormal basis $\{ |\phi_1 \rangle, \ldots, |\phi_n \rangle \}$ with real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Since $M \bar{G}$ is symmetric, $\{ |\phi_1 \rangle, \ldots, |\phi_n \rangle \}$ is left eigenbasis. By (2.2), we deduce $T$ is diagonalizable with a right eigenbasis $\{ D^{-1/2} |\phi_1 \rangle, \ldots, D^{-1/2} |\phi_n \rangle \}$, a left eigenbasis $\{ \langle \phi_1 |D_1^{1/2} |\phi_1 \rangle, \ldots, \langle \phi_n |D_1^{1/2} |\phi_n \rangle \}$, and eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. If $|\mu| = \sum_{i=1}^n c_i |\phi_i | D_1^{1/2}$, then by (2.2) we obtain that

$$|\mu| = \sum_{i=1}^n \lambda_i^2 |c_i | \langle \phi_i | D_1^{1/2}.$$

So it is crucial to gain a better understanding of $\lambda_i$’s. This is achieved by looking at the multiplication by $T$ from left:

$$\mathcal{T} : L^2(V) \to L^2(V) \mid \mathcal{T}(f) = T|f|.$$

Notice that for every $v \in V$, $\mathcal{T}(f)(v) = \sum_w \mathbb{P}(v \to w) f(w)$ is the average of the values of $f$ at the neighbors of $v$. Based on the fact that $\mathcal{T}$ is an averaging operator and the maximum modulus principle, we can gain some basic information on $\lambda_i$’s.

For every $i$, let $\bar{\phi}_i : V \to \mathbb{R}, \bar{\phi}_i(v) := d_v^{-1/2} \phi_i (v)$. Then $\mathcal{T}(\bar{\phi}_i) = \lambda_i \bar{\phi}_i$ for every $i$. After replacing $\tilde{\phi}_i$ with $-\bar{\phi}_i$, if needed, we can and will assume that for some $w_0^{(i)} \in V$, $\bar{\phi}_i(w_0^{(i)}) = \| \bar{\phi}_i \|_\infty = \max_{v \in V} \{ |\bar{\phi}_i (v) | \}$. Hence

$$|\lambda_i| |\bar{\phi}_i(w_0^{(i)})| = |\mathcal{T}(\bar{\phi}_i)(w_0^{(i)})| \leq \sum_v \mathbb{P}(w_0^{(i)} \to v) |\bar{\phi}_i (v) | \leq \| \bar{\phi}_i \|_\infty.$$ 

(2.4)

By (2.4), we obtain that $|\lambda_i| \leq 1$ for every $i$.

**Lemma 5.** Every eigenvalue of $T_\bar{G}$ is in the interval $[-1, 1]$.

Next, we investigate the extreme possible values. Notice that $\mathcal{T} \mathbb{I}_V = \mathbb{I}_V$ where $\mathbb{I}_V$ is the constant function $1$; we can observe this based on the fact that $\mathcal{T} f(v)$ is the
average of the values of \( f \) at the neighbors of \( v \). Hence 1 is definitely an eigenvalue of \( \mathcal{T} \), and so by Lemma 5, \( \lambda_1 = 1 \).

A function in the kernel of \( \mathcal{T} - I \) is called harmonic. Suppose \( f \) is a non-zero real harmonic function. After replacing \( f \) with \(-f\), if needed, we can and will assume that \( f(w_0) = \|f\|_\infty \) for some \( w_0 \in V \). Then

\[
f(w_0) = |f(w_0)| \leq \sum_v \mathbb{P}(w_0 \to v)|f(v)| \leq ||f||_\infty,
\]

which implies that for every \( v \in V \), either \( \mathbb{P}(w_0 \to v) = 0 \) or \( f(v) = f(w_0) \). This means \( f(v) = f(w_0) \) for \( v \) that is connected to \( w_0 \). Repeating this argument, we obtain that \( f(v) = f(w_0) \) for every \( v \) in the connected component of \( w_0 \) in \( \mathcal{T} \). Conversely, characteristic functions of connected components of \( \mathcal{T} \) are harmonic functions. Altogether, we have proved the following statement.

**Lemma 6.** The dimension of the operator \( \mathcal{T} - I \) is equal to the number of connected components of \( \mathcal{T} \). In particular, \( \mathcal{T} \) is connected if and only if \( \lambda_2 < 1 \).

Suppose \( \mathcal{T} \) has eigenvalue \(-1\) and \( \mathcal{T} f = -f \) for a nonzero function \( f \). Replacing \( f \) with \(-f\), if needed, we can and will assume that \( f(w_0) = ||f||_\infty \) for some \( w_0 \in V \). Hence

\[
0 = \sum_v \mathbb{P}(w_0 \to v)(f(w_0) + f(v)) \quad \text{and} \quad f(w_0) + f(v) \geq 0, \quad \text{for every} \ v.
\]

Therefore for every neighbor \( v \) of \( w_0 \), we have \( f(v) = -f(w_0) \). Repeating this argument, we see that the value of \( f \) at every neighbor of a neighbor of \( w_0 \) is again \( f(w_0) \). We deduce that the connected component of \( w_0 \) is a bipartite graph; this means the vertices of this connected component can be partitioned into two sets \( A \) and \( B \), and every edge has an element in \( A \) and an element in \( B \).

**Lemma 7.** In the above setting \( \lambda_n = -1 \) if and only if \( \mathcal{T} \) has a bipartite connected component.

**Proof.** We have already proved that if \( \lambda_n = -1 \), then \( \mathcal{T} \) has a bipartite connected component. For the converse look at Exercise 1. \( \square \)

By (2.3), and Lemmas 6 and 7, we obtain the following result on the rate of convergence of random-walks on a finite connected non-bipartite regular graph.

**Proposition 8.** Suppose \( \{X_i\}_{i=0}^\infty \) is a random-walk on a finite connected non-bipartite \( k \)-regular graph \( \mathcal{G} \) \((k\text{-regular means that the degree of all the vertices are } k)\). Suppose \( \mu_i \) is the distribution of \( X_i \). Suppose \( \lambda_1 \geq \ldots \geq \lambda_n \) are as before the eigenvalues of the transition matrix. Let \( \lambda_{\mathcal{G}} := \max\{|\lambda_2|,|\lambda_n|\} \). Then the following statements hold.

1. \((L^2\text{-convergence})\) For every \( f \in L^2(V) \) and every positive integer \( l \),

\[
\left\| \mathcal{T}^l f - \frac{\langle f, \mathbb{1}_V \rangle}{|V|} \mathbb{1}_V \right\|_2 \leq \lambda_{\mathcal{G}}^l \|f\|_2
\]

where \( \mathcal{T} \) is as before.
2. \((L^1\)-convergence\) For every \(f \in L^1(V)\) and every positive integer \(l\),
\[
\left| \mathbb{E}[f(X_l)] - \frac{\sum_{v \in V} f(v)}{|V|} \right| \leq \lambda_g \|f\|_2;
\]
in particular, for every \(A \subseteq V\),
\[
\left| \mathbb{P}(X_l \in A) - \frac{|A|}{|V|} \right| \leq \lambda_g \sqrt{|A|}.
\]

3. (Mixing) For every \(f, g \in L^2(V)\),
\[
\left| \langle f, \mathcal{T}^l g \rangle - \left( \sum_{v \in V} f(v) \right) \frac{\sum_{v \in V} g(v)}{|V|} \right| \leq \lambda_g \|f\|_2 \|g\|_2.
\]

**Proof.** Suppose \(\{\phi_1, \ldots, \phi_n\}\) is as before an orthonormal basis of \(M_g\). Notice that since \(G\) is \(k\)-regular, \(M_g = T_g\). Also notice that \(\phi_1 = \frac{1}{\sqrt{|V|}} \mathbb{1}_V\), and for every \(f \in L^2(V)\), the orthogonal projection of \(f\) to the space of constant functions is
\[
\frac{\langle f, \mathbb{1}_V \rangle}{|V|} \mathbb{1}_V.
\] (2.5)

For \(f \in L^2(V)\), suppose \(f = \sum_{i=1}^n c_i \phi_i\). Then \(\|f\|_2^2 = \sum_{i=1}^n |c_i|^2\) and by (2.5), we have
\[
\mathcal{T}^l f - \frac{\langle f, \mathbb{1}_V \rangle}{|V|} \mathbb{1}_V = \sum_{i=2}^n \lambda_i c_i \phi_i.
\]
This implies that
\[
\left\| \mathcal{T}^l f - \frac{\langle f, \mathbb{1}_V \rangle}{|V|} \mathbb{1}_V \right\|_2^2 = \sum_{i=2}^n |\lambda_i|^2 |c_i|^2 \leq \lambda_g^2 \sum_{i=2}^n |c_i|^2 \leq \lambda_g \|f\|_2^2.
\]
This completes the proof of the first part.

Assuming that the first \(L^1\)-convergence inequality is proved, we let \(f\) be the characteristic function \(\mathbb{1}_A\) of \(A\). The desired inequality follows from the fact that \(\mathbb{E}[\mathbb{1}_A(X_l)] = \mathbb{P}(X_l \in A)\). For the rest of the inequalities look at the exercise 2. \(\square\)

We refer to \(\lambda_g\) as the **spectral gap** of this random walk. Notice that
\[
\lambda_g = \| \mathcal{T}_g |_{L^2(V)^\circ} \|_{\text{op}}
\]
where \(L^2(V)^\circ := \{ f \in L^2(V) | \sum_{v \in V} f(v) \}\) is the space of functions that are orthogonal to the space of constant functions.

It is intuitive that a random-walk on a well-connected regular graph should quickly converge to equidistribution. This means having a lower bound for the Cheeger constant \(h(\mathcal{G})\) of a finite \(k\)-regular graph \(\mathcal{G}\) should give us an upper bound for \(\lambda_g\). In the rest of this chapter, we prove a variant of this result.
2.2 Discrete Laplacian

Suppose $G$ is a finite $k$-regular graph. Pick an orientation for the edges. For every edge $e \in E_G$, let $e^-$ be the initial vertex and $e^+$ be the terminal vertex of the oriented version. Thinking about a function $f : V_G \to \mathbb{R}$ as the amount of charge on nodes, $df(e) := f(e^+) - f(e^-)$ measures the amount of the resistance times the current on that edge. We can also think of the vertices as 0-cells, the edges as 1-cells, and $d : L^2(V) \to L^2(E)$, $df(e) := f(e^+) - f(e^-)$ as the boundary map. We notice that $d^* : L^2(E) \to L^2(V)$, $d^*g(v) = \sum_{v \sim w} g(w) - \sum_{v \sim w} g(v)$ (See Exercise 3). Thinking about a function $g$ on the edges as the amount of a flow going through that edge, we can think about $d^*g(v)$ as the amount of the flow that sinks in $v$. For every $f \in L^2(V)$, we have

$$d^*df(v) = \sum_{v \sim w} df(e) - \sum_{v \sim w} df(e)$$

$$= \sum_{v \sim w} (f(e^+) - f(e^-)) - \sum_{v \sim w} (f(e^+) - f(e^-))$$

$$= d_e f(v) - \sum_{w \sim v} f(w)$$

$$= k((I - \mathcal{R}_G)(f))(v),$$

where $w \sim v$ means $\{w, v\}$ is an edge in $\mathcal{G}$. Hence

$$\mathcal{L}_G := I - \mathcal{R}_G = \frac{1}{k} d^*d,$$  \hspace{1cm} (2.6)

and it is called the discrete Laplacian of the $k$-regular graph $\mathcal{G}$. Assuming that $\{\phi_1, \ldots, \phi_n\}$ is an orthonormal basis of $\mathcal{F}$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, by (2.6) we obtain that

$$\mathcal{L}_G(\phi_i) = (1 - \lambda_i)\phi_i$$

for every $i$. Hence assuming $\mathcal{G}$ is connected, eigenvalues of $\mathcal{L}_G$ are

$$0 = 1 - \lambda_1 < 1 - \lambda_2 \leq \cdots \leq 1 - \lambda_n \leq 2.$$

For $f \in L^2(V)$, suppose $f = \sum_{i=1}^n c_i \phi_i$. Then $\|f\|_2^2 = \sum_{i=1}^n |c_i|^2$ and

$$\|df\|_2^2 = \langle df, df \rangle = \langle f, d^*df \rangle = \langle f, \mathcal{L}f \rangle$$

$$= \sum_{i,j} c_i \overline{c}_j \langle \phi_i, \mathcal{L}\phi_j \rangle = \sum_{i,j} (1 - \lambda_j) c_i \overline{c}_j \langle \phi_i, \phi_j \rangle$$

$$= \sum_{i=1}^n (1 - \lambda_i) |c_i|^2 = \sum_{i=2}^n (1 - \lambda_i)|c_i|^2.$$  \hspace{1cm} (2.7)

By (2.7), we obtain the following description of $1 - \lambda_2$. 

Lemma 9. In the previous setting,

\[ 1 - \lambda_2 = \min \left\{ \| df \|_2^2 f \in L^2(V) \setminus \{0\} \right\}. \]

Proof. By (2.7), we have

\[ \| df \|_2^2 \geq (1 - \lambda_2^2) \sum_{i=2}^{n} |c_i|^2 \]

where \( f = \sum_{i=1}^{n} c_i \phi_i \). Notice that \( c_i = \langle f, \phi_i \rangle \) for every \( i \); in particular \( c_1 = 0 \) as \( \phi_1 \) is constant and \( f \in L^2(V) \). Therefore

\[ \| f \|_2^2 = \sum_{i=2}^{n} |c_i|^2. \]

Altogether, we have

\[ \| df \|_2^2 \geq (1 - \lambda_2^2) \| f \|_2^2 \]

for every \( f \in L^2(V) \). Therefore

\[ 1 - \lambda_2 \leq \min \left\{ \| df \|_2^2 f \in L^2(V) \setminus \{0\} \right\}. \quad (2.8) \]

We also notice that

\[ \| df_2 \|_2^2 = \langle f_2, L f_2 \rangle = 1 - \lambda_2^2, \]

and this shows that the equality in (2.8) holds. This completes the proof.

Sometimes it is useful to notice that

\[ \| df \|_2^2 = \sum_{e \in E} |f(e^+) - f(e^-)|^2 = \sum_{w \sim v} |f(v) - f(w)|^2, \]

and so

\[ 1 - \lambda_2 \leq \frac{\sum_{w \sim v} |f(v) - f(w)|^2}{\sum_v |f(v)|^2} \]

if \( \sum_v f(v) = 0 \) and \( f \neq 0 \).

Next we show that the Cheeger constant can be described based on an \( L^1 \)-version of Lemma 9. This is done based on finding various good cuts.

2.3 Finding good cuts

In a society, communities shape based on certain features. Inspired by this, in social medias, we try to find certain features that can distinguish various communities. A basic such example is finding a feature that can split people into two communities; this means finding a good cut in the underlying graph. In mathematical language, a feature is simply a function \( f \) on the set of the vertices of the given (social media) graph, and after picking a critical value \( c_0 \), we split the vertices based on whether the value of \( f \) at the given vertex is more or less than \( c_0 \).

Suppose \( G \) is a finite graph with the set of vertices \( V \) and set of edges \( E \). For \( f : V \to \mathbb{R} \) and \( c \in \mathbb{R} \), let

\[ V_{f,c}^- := \{ v \in V | f(v) < c \}, \]

and

\[ h_G(f) := \inf_c \frac{|E(V_{f,c}^-, V \setminus V_{f,c}^-)|}{\min\{|V_{f,c}^-|, |V \setminus V_{f,c}^-|\}}. \]

This means \( h_G(f) \) quantifies how good of a cut we can get using \( f \). We can view \( f \) as a projection of the graph \( G \) to a line. Starting with a measure \( \mu \) on \( \mathbb{R} \), using the
2.3. FINDING GOOD CUTS

projection given by $f$, we can put a weight on each edge. For an edge $e = \{v, w\}$, let $I_e$ be the interval with the end points $f(v)$ and $f(w)$. Then the weight of $e$ corresponding to $\mu$ and $f$ is $\mu(I_e)$.

In this section, we use a probabilistic method to find upper bounds for $h_g(f)$.

**Lemma 10.** Suppose $\mu$ is a measure on $\mathbb{R}$, and $f : V \to \mathbb{R}$. For every edge $e = \{v, w\}$, let $I_e$ be the interval with end points $f(v)$ and $f(w)$. Then

$$\int |E(V_{f_e}^-, V \setminus V_{f_e}^-)|d\mu(c) = \sum_e \mu(I_e).$$

**Proof.** Notice that $e \in E(V_{f_e}^-, V \setminus V_{f_e}^-)$ if and only if $c \in I_e$. Hence

$$\int |E(V_{f_e}^-, V \setminus V_{f_e}^-)|d\mu(c) = \sum_e \int [e \in E(V_{f_e}^-, V \setminus V_{f_e}^-)]d\mu(c) = \sum_e \mu(I_e).$$

This completes the proof. \qed

**Lemma 11.** Suppose $\mu$ is a measure on $\mathbb{R}$ such that $\mu(\{e\}) = 0$ for every $e \in \mathbb{R}$, and $f : V \to \mathbb{R}$. Let $c_0$ be the median of $f(v)$’s as $v$ ranges in $V$. For every $v \in V$, let $I_v$ be the interval with the end points $c_0$ and $f(v)$. Then

$$\int \min\{|V_{f_e}^-|, |V \setminus V_{f_e}^-|\}d\mu(c) = \sum_v \mu(I_v).$$

**Proof.** Notice that $\min\{|V_{f_e}^-|, |V \setminus V_{f_e}^-|\} = |V_{f_e}^-|$ if and only if $c \leq c_0$. Hence

$$\int \min\{|V_{f_e}^-|, |V \setminus V_{f_e}^-|\}d\mu(c) = \int_{c \leq c_0} |V_{f_e}^-|d\mu(c) + \int_{c > c_0} |V \setminus V_{f_e}^-|d\mu(c)$$

$$= \sum_v [f(v) < c_0] \mu(f(v), c_0) + [f(v) \geq c_0] \mu(c_0, f(v)]$$

$$= \sum_v \mu(I_v).$$

This completes the proof. \qed

**Theorem 12.** Suppose $\mu$ is a measure on $\mathbb{R}$ such that $\mu(\{e\}) = 0$ for every $e \in \mathbb{R}$. Suppose $f : V \to \mathbb{R}$ is a function such that $\mu(\{\min_v f(v), \max_v f(v)\}) \neq 0$. Let $c_0$ be the median of $f(v)$’s as $v$ ranges in $V$. For $v \in V$, let $I_v$ be the interval with the end points $c_0$ and $f(v)$, and for $e = \{v, w\} \in E$, let $I_e$ be the interval with the end points $f(v)$ and $f(w)$. Then

$$h_g(f) \leq \sum_{e \in E} \mu(I_e) \sum_{v \in V} \mu(I_v).$$

**Proof.** By Lemmas 10 and 11, we have

$$\int (\sum_v \mu(I_v))|E(V_{f_e}^-, V \setminus V_{f_e}^-)| - (\sum_e \mu(I_e)) \min\{|V_{f_e}^-|, |V \setminus V_{f_e}^-|\}d\mu(c) = 0.$$
Therefore for some $c$ we have that
\[
\left( \sum_v \mu(I_v))|E(V_{f,c}, V \setminus V_{f,c})| - \sum_{e \in E} \mu(I_e) \min\{|V_{f,c}|, |V \setminus V_{f,c}|\} \right) \leq 0,
\]
and so
\[
h_G(f) \leq \frac{|E(V_{f,c}, V \setminus V_{f,c})|}{\min\{|V_{f,c}|, |V \setminus V_{f,c}|\}} \leq \sum_{e \in E} \mu(I_e).
\]

This finishes the proof. □

Special cases of $\mu$ give us interesting results. For instance the case when $\mu$ is the Lebesgue measure implies the following Theorem.

**Theorem 13.** Suppose $\mathcal{G}$ is a finite graph with the set of vertices $V$. Let $h(\mathcal{G})$ be the Cheeger constant of $\mathcal{G}$. Then
\[
h(\mathcal{G}) = \inf \left\{ \frac{\|df\|_1}{\|f\|_1} \mid f \in L^1(V) \setminus \{0\}, \text{Med}(f) = 0 \right\},
\]
where Med$(f)$ is the median of $f(v)$'s as $v$ ranges in $V$.

Notice that for the $L^2$-norm in the denominator we took a shift of $f$ which minimized the $L^2$-norm, and here for the $L^1$-norm we are taking a shift of $f$ which minimizes the $L^1$-norm! It worths pointing out that $df$ does not change as we shift $f$ by a constant.

**Proof of Theorem 13.** Applying Theorem 12 for the case when $\mu$ is the Lebesgue measure $\ell$, we obtain that
\[
h(\mathcal{G}) \leq \frac{\sum_{e \in E} \ell(I_e)}{\sum_{v \in V} \ell(I_v)},
\]
where $E$ is the set of edges of $\mathcal{G}$, for $e = \{v, w\}$, $I_e$ is an interval with the end points $f(v), f(w)$, and for $v \in V$, $I_v$ is an interval with the end points Med$(f) = 0$ and $f(v)$. Hence
\[
\ell(I_e) = |df(e)| \quad \text{and} \quad \ell(I_v) = |f(v)|.
\]
Therefore $h(\mathcal{G}) \leq \frac{\|df\|_1}{\|f\|_1}$ if Med$(f) = 0$ and $f \neq 0$.

Suppose $h(\mathcal{G}) = \frac{|E(A,A^c)|}{|A|}$ for some $A \subseteq V$ with $|A| \leq |V|/2$. Let $f = 1_A$ be the characteristic function of $A$. Since $|A| \leq |V|/2$, Med$(f) = 0$. Notice that $\|df\|_1 = |E(A,A^c)|$ and $\|f\|_1 = |A|$, and so $h(\mathcal{G}) = \frac{\|df\|_1}{\|f\|_1}$. This completes the proof. □

### 2.4 Discrete isoperimetric inequalities

In this section, we use the $L^2$-optimization description of $1 - \lambda_2$ (see Lemma 9) and the bounds that we have found for $h(\mathcal{G})$ using Theorem 12, and prove isoperimetric inequalities.

Using Theorem 12 for the case when $\mu$ is the Lebesgue measure, we found an $L^1$-optimization description of $h(\mathcal{G})$. Thinking about edges as wires laying on a surface,
2.4. DISCRETE ISOPERIMETRIC INEQUALITIES

the Lebesgue measure more or less ends up giving us the total weight on top of a point. Next, we roughly think about this graph balanced about the median of $f$ and measure the torque of each edge. This means we assume that 0 is the median and consider the measure $\mu$ given by the density function

$$d\mu(t) := |t| dt.$$  

Notice that $\mu([a,b]) = \int_a^b |t| dt = \frac{b|b| - a|a|}{2}$. Hence

$$\sum_v \mu(I_v) = \frac{\|f\|^2}{2}, \quad (2.9)$$

where $I_v$ is the interval with the endpoints $\text{Med}(f) = 0$ and $f(v)$. We also have

$$\sum_e \mu(I_e) = \frac{1}{2} \sum_e (f(e^+)|f(e^+)| - f(e^-)|f(e^-)|) \quad (2.10)$$

Notice that for every $a, b \in \mathbb{R}$, we have

$$b|b| - a|a| \leq |b - a|(|b| + |a|). \quad (2.11)$$

By (2.10) and (2.11), we obtain that

$$\sum_e \mu(I_e) \leq \sum_e |df(e)| \left( \frac{|f(e^+)| + |f(e^-)|}{2} \right),$$

and so by the Cauchy-Schwarz inequality, we have

$$\sum_e \mu(I_e) \leq \|df\|_2 \sqrt{\sum_e \left( \frac{|f(e^+)| + |f(e^-)|}{2} \right)^2}. \quad (2.12)$$

Because $(\frac{a+b}{2})^2 \leq \frac{a^2 + b^2}{2}$, by (2.12), we obtain

$$\sum_e \mu(I_e) \leq \|df\|_2 \sqrt{\sum_e |f(e^+)|^2 + |f(e^-)|^2} = \sqrt{\frac{k}{2}} \|df\|_2 \|f\|_2, \quad (2.13)$$

if $\mathcal{G}$ is a $k$-regular graph.

By (2.9), (2.13), and Theorem 12, we deduce the following result.

**Lemma 14.** Suppose $\mathcal{G}$ is a $k$-regular graph and $f : V \to \mathbb{R}$ is a function whose median is 0. Then

$$h(\mathcal{G}) \leq \sqrt{\frac{k}{2}} \frac{\|df\|_2}{\|f\|_2}.$$
2.5 Exercises

1. Suppose \( G \) is a finite graph that has a bipartite connected component. Prove that \( \mathcal{T}_G \) has eigenvalue \(-1\).

   (Hint: Suppose \( V \) has two disjoint subsets \( A \) and \( B \) such that if an edge \( e \) intersects \( A \cup B \), then \(|e \cap A| = |e \cap B| = 1\). Let \( f := \mathbb{1}_A - \mathbb{1}_B \) where for a subset \( Y \) of \( V \), \( \mathbb{1}_Y \) is the characteristic function of \( Y \). Prove that \( \mathcal{T} f = -f \).)

2. Prove the \( L^1 \)-convergence and the mixing property of a random-walk in a finite regular graph given in Proposition 8.

   (Hint. For the mixing, use the Cauchy-Schwarz inequality and obtain
   \[
   \left| \langle f, \mathcal{T} l g \rangle - \langle g, \mathbb{1}_V \rangle \mathbb{1}_V \right| \leq \| f \|_2 \| \mathcal{T} l g - \langle g, \mathbb{1}_V \rangle \mathbb{1}_V \|_2 \leq \lambda \| f \|_2 \| g \|_2,
   \]
   and finish the proof. For the \( L^1 \)-convergence, use the mixing inequality for \( g \) equals to the initial distribution \( \mu_0 \). Notice that
   \[
   \langle f, \mathcal{T} \mu_0 \rangle = \mathbb{E}[f(X_1)] \quad \text{and} \quad \sum_{v \in V} \mu_0(v) = 1.
   \]

3. Suppose \( \mathcal{G} = (V, E) \) is a directed graph. Let
   \[
   d : L^2(V) \to L^2(E), df(e) := f(e^+) - f(e^-).
   \]
   Prove that
   \[
   d^* g(v) = \sum_{v = e^-} g(e) - \sum_{v = e^+} g(e).
   \]
   (Hint. Notice that
   \[
   \langle df, g \rangle = \sum_{e \in E} df(e) g(e) = \sum_{e \in E} \left( f(e^+) - f(e^-) \right) g(e) = \sum_{v \in V} f(v) \left( \sum_{v = e^-} g(e) - \sum_{v = e^+} g(e) \right).
   \]