\( G(V,E) \) is an undirected simple \( d \)-regular graph.

\[
\alpha = \frac{1}{2d} \min_{A \in V} \frac{|C(A)|}{|A|} \quad \text{the number of edges crossing from \( A \) to \( V \setminus A \)}
\]

\( A \) is the adjacency matrix of \( G(V,E) \).

\( P \) is the matrix \( P_{ij} = \begin{cases} \frac{1}{2} & \text{if } i=j \\ \frac{1}{2d} & \text{if } i \neq j \land (ij) \in E \\ 0 & \text{otherwise} \end{cases} \)

\underline{Note}: \((1,1,\ldots,1)\) is an eigenvector of \( P \) with eigenvalue \( 1 \).

\underline{Proof}: Verify that \((1,\ldots,1)P = (1,\ldots,1)\)

\underline{Note}: \( P \) is real & symmetric, it thus has an orthogonal set of eigenvectors with real eigenvalues.

Let \( \hat{e} = (e_1,\ldots,e_n) \) be an eigenvector of \( P \) other than \((1,\ldots,1)\) with eigenvalue \( \lambda \): \( \hat{e}P = \lambda \hat{e} \)

\[ \lambda \hat{e} \]

\( \hat{e} \) is orthogonal to the all 1's vector, thus \( \sum_i e_i = 0 \), thus \( \hat{e} \) has both positive and negative entries.

We will show \( \lambda^2 \leq 1 - \alpha^2 \)
Express the action of \( P \) on any \( \vec{h} \) as averaging along edges.

Notice \( \sum h_i^2 = \sum (h_i + h_j)^2 \)

Let \( \vec{h}_P = \vec{h}P \) and notice \( h_i^P = \frac{1}{2} h_i + \sum_{j: (ij) \in E} \frac{h_j}{2d} \).

Thus \( \sum h_i^P = \sum \left( \sum_{j: (ij) \in E} \frac{h_i + h_j}{2d} \right)^2 \).

square of means \( \leq \sum_{i} \sum_{j: (ij) \in E} \frac{1}{d} \left( \frac{h_i + h_j}{2} \right)^2 = \sum \frac{1}{d} \left( \frac{h_i + h_j}{2} \right)^2 \).

where \( x_k = \frac{h_i + h_j}{2} \).

Now \( \sum h_i^2 - \sum h_i^P^2 \geq \sum_{(ij) \in E} \left( \frac{h_i^2 + h_j^2}{2d} - \frac{1}{d} \left( \frac{h_i + h_j}{2} \right)^2 \right) \).

\[ = \frac{1}{4d} \sum_{ij \in E} (h_i - h_j)^2 \]
2. **Effective Averaging for Expanders**

Let \( h_i \geq \ldots \geq h_n \) be any vector such that

\[
\left| \{ i \mid h_i > 0 \} \right| \leq \frac{n}{2} \quad \text{and} \quad \left| \{ i \mid h_i < 0 \} \right| \leq \frac{n}{2}
\]

and let \( \vec{h} = \vec{h} P \), as before.

**Claim** \[
\sum_i h_i^2 - \sum_i \left( h_i P \right)^2 \geq \alpha \sum_i h_i^2
\]
Proof From the end of page 2 we have:

\[ \sum h_i^2 - \sum (h_i^p)^2 \geq \frac{1}{4d} \sum (h_i - h_j)^2 \]

\[ \geq \frac{1}{4d} \sum (h_i^+ - h_j^+)^2 + \frac{1}{4d} \sum (h_i^- - h_j^-) \]

We will show:

\[ \frac{1}{4d} \sum_{(ij) \in E} (h_i^+ - h_j^+)^2 \geq \sum h_i^2 \]

and

\[ \frac{1}{4d} \sum_{(ij) \in E} (h_i^- - h_j^-)^2 \geq \alpha \sum h_i^- \]
For $h_i^+$ (the case of $h_i^-$ is identical) we have:

$$4 \sum_i (h_i^+)^2 = 2 \sum_{ij \in E} \frac{h_i^+ h_j^+}{d} \geq \frac{1}{d} \sum_{ij \in E} (h_i^+ + h_j^+)^2$$

Thus

$$\frac{1}{4d} \sum_{(ij) \in E} (h_i^+ + h_j^+)^2 \leq 1$$

$$\sum_i h_i^+ \leq \frac{1}{4d} \sum_{ij \in E} (h_i^+ - h_j^+)^2$$

Thus

$$\frac{1}{4d} \sum_{ij \in E} (h_i^+ - h_j^+)^2 \geq \frac{1}{16d^2} \left( \sum_i (h_i^+)^2 \right) \left( \sum_{(ij) \in E} (h_i^+ - h_j^+)^2 \right)$$

Cauchy-Schwarz
Now \( \sum_{(i,j) \in E} \left| \mathbf{b}_i^+ - \mathbf{b}_j^+ \right|^2 = 2 \sum_{ij \in E \atop i < j} \left( \mathbf{b}_i^+ - \mathbf{b}_j^+ \right)^2 \)

\[ = 2 \sum_{ij \in E \atop i < j} \sum_{k=1}^{j-1} \left( \mathbf{b}_k^+ - \mathbf{b}_{k+1}^+ \right)^2 \]

\[ = 2 \left( \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor - 1} \mathbf{b}_i^+ - \mathbf{b}_{i+1}^+ \right)^2 \subseteq \left( \mathbf{A}_k \right)^2 \]

\[ \geq 4d \alpha \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor - 1} \left( \mathbf{b}_k^+ - \mathbf{b}_{k+1}^+ \right)^2 \]

\[ = 4d \alpha \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor - 1} \left( \mathbf{b}_k^+ \right)^2 \]

Thus from the previous page:

\[ \frac{1}{4d} \sum_{i,j \in E} \left( \mathbf{b}_i^+ - \mathbf{b}_j^+ \right)^2 \geq \frac{1}{16d^2} \sum_i \mathbf{b}_i^2 \]

\[ \frac{1}{4d} \sum_{i,j \in E} \left( \mathbf{b}_i^+ - \mathbf{b}_j^+ \right)^2 \geq \alpha^2 \sum_{i} \left( \mathbf{h}_i^+ \right)^2 \]
Now by observing that $h_i$ can be treated identically and going at the top of page 4 we have

$$\sum_{i} h_i^2 - \sum_{i} (h_i^p)^2 \geq \alpha^2 \sum_{i} h_i^2$$
3. NORMALIZATION

Now suppose that \( \vec{e} \) is an eigenvector

and let \((e, e, \ldots, e)\) be a "normalization" so that \(\vec{e} + \vec{c}\) is of the form of \(\vec{1}\).

\[ \langle \vec{c}, \vec{e} \rangle = 0 \]
We have

\[ \| \hat{e} + \hat{c} \| = \| \hat{e} \| + \| \hat{c} \| \]

\[ (\hat{e} + \hat{c}) P = \hat{e} P + \hat{c} P = \lambda \hat{e} + \hat{c} \quad \text{thus} \]

\[ \| (\hat{e} + \hat{c}) P \| = \lambda \| \hat{e} \| + \| \hat{c} \| \]

Thus

\[ \| \hat{e} + \hat{c} \|^2 - \| (\hat{e} + \hat{c}) P \|^2 = (1 - \lambda^2) \| \hat{e} \|^2 \]

But from previous analysis

\[ \| \hat{e} + \hat{c} \|^2 - \| (\hat{e} + \hat{c}) P \|^2 \geq \lambda^2 \| \hat{e} + \hat{c} \|^2 \]

Thus

\[ \lambda^2 \| \hat{e} \|^2 + \frac{\lambda^2}{2} \| \hat{c} \|^2 \]

Thus

\[ 1 - \lambda^2 \geq \lambda^2 \quad \text{or} \quad \lambda^2 \leq 1 - \lambda^2 \]
Conductance and Convergence of Markov Chains
– A Combinatorial Treatment of Expanders –

Milena Mihail*
Harvard University
and
U. C. Berkeley

(Extended Abstract)

Abstract

We give a direct combinatorial argument to bound the convergence-rate of Markov chains in terms of their conductance (these are statements of the nature “random walks on expanders converge fast”). In addition to showing that the linear algebra in previous arguments for such results on time-reversible Markov chains was unnecessary, our direct analysis applies to general irreversible Markov chains.

1 Introduction

Recently there has been considerable interest in rapidly mixing properties of Markov chains, that is Markov chains which come “close” to their stationary distribution after a “small” number of steps – “small” is to be compared with the number of states. From a theoretical perspective such properties introduce a complexity aspect to discrete probability: in contrast to classical Perron-Frobenius analysis [Se73] the convergence bounds are non-asymptotic [Ald83] [AD87] [Ald88] [Alo86] [Ald87] [SJ87]. (From the algorithmic perspective rapidly mixing Markov chains on specific combinatorial populations have resulted in remarkable sampling and approximate counting schemes for hard problems [Br86] [JS88] [DLMV88] [DFK89]).

Analyzing the convergence rate of Markov chains is a formal way to reason about the convergence rate of random walks on expanders. So far, the reasoning used to establish the simple fact that “random walks on expanders converge fast” was strongly algebraic (influenced by non-trivial bounds on spectra that were essential in different applications of expanders, e.g. explicit constructions). In particular, the proofs preceding ours were obtained along the following general lines:

• The distance from stationarity is expressed by some “discrepancy” vector.
• For undirected graphs the adjacency matrix $A$ of the graph is symmetric and possesses an orthogonal basis of eigenvectors. Consequently, the discrepancy vector can be written in this basis, and (under mild conditions) the second largest eigenvalue $\lambda_2$ of $A$ provides an effective characterization of the convergence rate [AD87] [Ald87].
• More important, expansion implies separation of $\lambda_2$ from 1 (the discrete analogue of Cheeger’s theorem on Riemannian manifolds [Che70]), thus fast convergence is established [Ald87] [SJ87].

In contrast to this typical algebraic treatment of expanders, we reason from a purely combinatorial perspective, in fact from first principles, as follows:

• The convergence rate of a random walk on an expander graph can be viewed as the diffusion of an initial “charge” placed on the vertices of the graph.
• The charge is diffused along the edges of an expander according to a simple averaging rule: “each edge averages the charges of its two endpoints”.
• Expansion is equivalent to “well distributed edges over the entire graph”, which suggests a substantial number of edges with significantly different charges at their endpoints. Therefore the averaging is effective, the charge diffusion is rapid, and the convergence is fast.

Aside from providing a simple and straightforward insight for the rapidly mixing properties of expanders, our non-algebraic reasoning resulted in the natural generalization to directed graphs and arbitrary finite Markov chains of rapidly mixing statements that were known to hold only for undirected graphs and time-reversible Markov chains. Similar generalizations to arbitrary Markov Chains with continuous state space have been obtained independently.

*Supported by NSF-CCR86-58143
2 Regular Directed Graphs

Let \( G(V, E) \) be a \( d \)-regular directed graph, i.e., for every vertex the in-degree is equal to the out-degree and equal to \( d \). Let \( A \) be the adjacency matrix of \( G \): \( a_{ij} = 1 \) if \( i \in E \) and \( a_{ij} = 0 \) if \( ij \notin E \).

Consider a random walk on \( G \) with transition matrix \( P \):

\[
P_{ij} = \begin{cases} 
1/2 & \text{if } i = j \\
1/2d & \text{if } i \neq j \text{ and } ij \in E \\
0 & \text{otherwise}
\end{cases}
\]

Notice that self-loops of weight 1/2 have been added on each vertex. Technically, this takes care of periodicities that may occur in directed graphs and such random walks are usually called "strongly aperiodic." As we will discuss in section 3 the strong aperiodicity assumption can be dropped. However, it is instructive to keep it for the time being.

Let \( \mathcal{E}(t) \) be the distribution of the random walk at time \( t \): \( \mathcal{E}(t) = \mathcal{E}(t-1)P = \mathcal{E}(0)P^t \). Let \( \bar{\mathcal{E}} = \lim_{t \to \infty} \mathcal{E}(t) \), \( \bar{\mathcal{E}} \) is the "stationary distribution" of the random walk (it is an elementary fact that if \( G \) is connected, then \( \bar{\mathcal{E}} \) exists and it is unique). Moreover, \( \bar{\mathcal{E}} \) satisfies \( \bar{\mathcal{E}} = \bar{\mathcal{E}}P \); it is easy to check that for a random walk on \( G \), \( \pi_i = |V|^{-1} \) and the stationary distribution of the random walk is the uniform over \( V \). We are interested in the rate at which \( \mathcal{E}(t) \) approaches \( \bar{\mathcal{E}} \). Equivalently, we are interested in the rate at

To measure the distance of \( \mathcal{E}(t) \) from \( \bar{\mathcal{E}} \), we use the usual norm of the discrepancy vector:

\[
\| \mathcal{E}(t) - \bar{\mathcal{E}} \| = \sum_i e_i^2(t)
\]

In what follows we show that the expansion of \( G \) determines the rate at which \( \| \mathcal{E}(t) - \bar{\mathcal{E}} \| \to 0 \). In particular, consider the following version of cutset expansion:

\[
a = \frac{1}{2d} \min_{A \subseteq V^i: |A| \leq |V|/2} \frac{|C(A)|}{|A|}
\]

where \( C(A) \) is the cutset of \( A \): \( C(A) = \{ ij : i \in A, j \in \bar{A}, ij \in E \} \). In Theorem 2.1 we show that after each step the length of the discrepancy vector decreases significantly:

\[
\| \mathcal{E}(t) - \bar{\mathcal{E}}(t+1) \| \geq \alpha^2 \| \mathcal{E}(t) - \bar{\mathcal{E}}(t) \|
\]

Therefore,

\[
\| \mathcal{E}(t) - \bar{\mathcal{E}}(t+1) \| \leq (1 - \alpha^2)^t \| \mathcal{E}(0) - \bar{\mathcal{E}}(0) \|
\]

And the exponential convergence of the random walk follows:

\[
\| \mathcal{E}(t) - \bar{\mathcal{E}}(t) \| \leq (1 - \alpha^2)^t \| \mathcal{E}(0) - \bar{\mathcal{E}}(0) \|
\]

The first results of this flavor were obtained for undirected graphs and in terms of vertex expansion by Aldous who used Alon's bounds on eigenvalues [Alo86] [Ald87]. With respect to the version of cutset expansion defined above, Theorem 2.1 follows by the work of Sinclair and Jerrum for the special case of undirected graphs, with slightly worse constants, and under the same strong aperiodicity assumption. They showed \( \| \mathcal{E}(t) - \bar{\mathcal{E}}(t) \| \leq (1 - \alpha^2/2)^t \| \mathcal{E}(0) - \bar{\mathcal{E}}(0) \| \) [SJ87]. The constant was saved here by considering strong aperiodicity in all stages of the proof. In fact, such a consideration appears necessary to our combinatorial reasoning. Sinclair and Jerrum used strong aperiodicity simply to guarantee a positive spectrum and dominance of convergence rate by \( \lambda_2 \).

We proceed to state the Theorem.

**Theorem 2.1** For any initial distribution \( \mathcal{E}(0) \),

\[
\| \mathcal{E}(t) - \bar{\mathcal{E}}(t+1) \| \geq \alpha^2 \| \mathcal{E}(t) - \bar{\mathcal{E}}(t) \|
\]

**Proof.** The proof proceeds in four stages along the intuition described in the introduction.

Stage 1: Probability Charges

The point here is to view the discrepancy \( \mathcal{E}(t) - \bar{\mathcal{E}}(t) \) as a charge distributed over the vertices of the graph \( G \). And further, to realize that the action of one step of the random walk on \( \mathcal{E}(t) \) obeys a simple rule.

In general, a charge \( \bar{f} = < f_1, \ldots, f_n > \) is an assignment of real values to the vertex set \( V \). For a charge \( \bar{f} \), the norm of
component, so let \( f_i^+ = \max\{ f_i, 0 \} \) and \( f_i^- = \min\{ f_i, 0 \} \). A probability charge is a charge \( \varepsilon \) where \( \sum_{i \in V} e_i = \sum_{i \in V} e_i^+ = \sum_{i \in V} |e_i^-| = 0 \).

Clearly, the discrepancy vector \( \varepsilon(t) \) is a probability charge: \( \sum_{i \in V} e_i(t) = \sum_{i \in V} x_i(t) - \sum_{i \in V} x_i = 0 - 1 = 0 \). Moreover, the action of \( P \) on \( \varepsilon(t) \) is identical to the action of \( P \) on the probability distribution \( \varepsilon(t) \):

\[
\varepsilon(t+1) = \varepsilon(t+1) - \varepsilon(t)P = \varepsilon(t)P - \varepsilon P = \varepsilon(t)P
\]

Stage 2: AVERAGING ALONG EDGES
We are interested in the decrease \( \| \varepsilon(t) - \varepsilon(t+1) \| \). The idea in what follows is to express the action of \( P \) on \( \varepsilon(t) \) as an action of the edges of \( G \). In fact, this will be done for an arbitrary charge \( \tilde{f} \).

First express \( \| \tilde{f} \| \) so that each edge (rather than vertex) of \( G \) is assigned a fraction of the charge of its endpoints:

\[
\| \tilde{f} \| = \sum_{i} f_i^2 \\
= \sum_{i,j \in E} \frac{f_i^+ f_j^+}{2d}
\]

Next describe the action of \( P \) on \( \tilde{f} \) as an averaging of \( \tilde{f} \) along the edges of \( G \). For this, let \( \tilde{f}^P = \tilde{f} P \), and notice:

\[
f_i^P = \frac{1}{2} f_i + \sum_{j \in N^{-1}(i)} \frac{f_j}{2d} = \sum_{j \in N^{-1}(i)} \frac{f_i + f_j}{2d}
\]

where \( N^{-1}(i) = \{ j : ji \in E \} \). Now using the fact that the square of the mean is bounded by the mean of squares we get:

\[
\| \tilde{f}^P \| = \sum_{j} (f_j^P)^2 \\
= \sum_{j} \left( \sum_{i \in N^{-1}(j)} \frac{f_i + f_j}{2d} \right)^2 \\
\leq \sum_{j} \sum_{i \in N^{-1}(j)} \frac{1}{d} \left( \frac{f_i + f_j}{2} \right)^2 \\
= \sum_{i,j \in E} \frac{1}{d} \left( \frac{f_i + f_j}{2} \right)^2
\]

Finally establish that the averaging suggested by (2) and (3) is effective in reducing the norm of \( \tilde{f} \) if the quantities to be averaged are significantly different. This follows

\[
\| \tilde{f} \| - \| \tilde{f}^P \| \geq \sum_{i,j \in E} \left( \frac{f_i^2 + f_j^2}{2d} - \frac{1}{d} \left( \frac{f_i + f_j}{2} \right)^2 \right) \\
= \frac{1}{4d} \sum_{i,j \in E} (f_i - f_j)^2
\]

Realize that (4) precisely suggests that the net decrease \( \| \tilde{f} \| - \| \tilde{f}^P \| \) is due to edges of \( G \) with significantly different charges at their endpoints.

Stage 3: EFFECTIVE AVERAGING FOR EXPANDERS
In this and the next stage we want to pinpoint the idea that in cutset expanders the edges are “well distributed” all over the graph, therefore any placement of a charge on the vertices is forced to result in a large number of edges with significantly different charges at their endpoints. In this stage we will establish the above for a charge \( \tilde{h} \) such that \( |\{ i : h_i > 0 \}| \leq |V|/2 \) and \( |\{ i : h_i < 0 \}| \leq |V|/2 \) (wlog assume that \( |V|/2 \) is odd). For such a charge \( \tilde{h} \) we show:

\[
\frac{1}{4d} \sum_{i,j \in E} (h_i - h_j)^2 \geq \alpha^2 \| \tilde{h} \|
\]

Intuitively and technically (5) will be treated along the lines suggested by Sinclair and Jerrum’s manipulation of eigenvectors, which in turn is closely related to Alon’s treatment of eigenvectors (at this stage of the proof both the above references make minimal use of the property that the function of vertices manipulated is an eigenvector). In particular, the reasoning is as follows: If the vertices of \( G \) are ordered according to the value of their charge, so that \( h_1 \geq h_2 \geq \ldots \) then edges with significantly different charges at their endpoints become “long edges”. If \( A_k = \{ 1, \ldots, k \} \), then every such long edge \( ij \) will appear in each one of the cutsets \( C(A_k) \), \( k = i, \ldots, j \). In turn, the size of these cutsets depends directly on cutset expansion.

Proof of (5). Clearly,

\[
\frac{1}{4d} \sum_{i,j \in E} (h_i - h_j)^2 \geq \frac{1}{4d} \sum_{i,j \in E} (h_i^+ - h_j^+)^2 + \frac{1}{4d} \sum_{i,j \in E} (h_i^- - h_j^-)^2
\]

Now for the charge \( \tilde{h}^+ \) (1) suggests:

\[
\frac{4}{d} \sum_{i} (h_i^+)^2 = 2 \sum_{i,j \in E} \frac{(h_i^+)^2 + (h_j^+)^2}{d} \geq \frac{1}{d} \sum_{i,j \in E} (h_i^+ + h_j^+)^2
\]

Using (7) and the Cauchy-Schwartz inequality we have:

\[
\frac{1}{4d} \sum_{i,j \in E} (h_i^+ - h_j^+)^2
\]
holds for an arbitrary probability charge \( \varepsilon(t) \). To establish this consider an ordering of \( \varepsilon(t) \) so that \( \varepsilon_k(t) \geq \varepsilon_{k+1}(t) \) and let \( m \) be as before: \( m = (|V| + 1)/2 \). Now let \( \tilde{h} \) be a charge such that \( h_i = \varepsilon_i(t) - e_{m}(t) \) and notice that \( h_i \geq 0 \) for \( i < m \), while \( h_i \leq 0 \) for \( i \geq m \), therefore (5) applies to \( \tilde{h} \). The idea in what follows is that the net decrease of \( \varepsilon(t) \) is at least as large as that of \( \tilde{h} \). In particular, (4) and (5) suggest:

\[
\| \varepsilon(t) \| - \| \varepsilon(t+1) \| \geq \frac{1}{4d} \sum_{i,j \in E} (c_i(t) - c_j(t))^2
\]

\[
= \frac{1}{4d} \sum_{i,j \in E} (h_i^2 - h_j^2)
\]

\[
\sum_{ij \in E} (h_i^2 - h_j^2) = \sum_{ij \in E} (h_i^2 - h_{i+1}^2) + \sum_{ij \in E} (h_j^2 - h_{j+1}^2)
\]

\[
= \sum_{i \in E} (h_i^2 - h_{i+1}^2) \left( |C(A_k)| + |C(V \setminus A_k)| \right) \quad (9)
\]

The fact that \( G \) is \( d \)-regular suggests \( |C(A_k)| = |C(V \setminus A_k)| \), and (9) becomes:

\[
\sum_{ij \in E} |h_i^2 - h_j^2| = 2 \sum_{k=1}^{m-1} (h_k^2 - h_{k+1}^2) |C(A_k)| \quad (10)
\]

Let \( m \) be such that \( m = (|V| + 1)/2 \). Clearly \( h_m^2 = 0 \) (because of the special form of \( \tilde{h} \)) and (10) becomes:

\[
\sum_{ij \in E} |h_i^2 - h_j^2| = 2 \sum_{k=1}^{n-1} (h_k^2 - h_{k+1}^2) |C(A_k)| \quad (11)
\]

Moreover by the definition of \( \alpha \) we know that \( |C(A)| \geq \frac{2d}{n} |A_k| \), hence (11) becomes:

\[
\sum_{ij \in E} |h_i^2 - h_j^2| \geq 4d \sum_{k=1}^{m-1} (h_k^2 - h_{k+1}^2) |A_k| \quad (12)
\]

In view of (12), (8) yields the following simple bound:

\[
\frac{1}{4d} \sum_{ij \in E} (h_i^2 - h_j^2) \geq \alpha^2 \sum_i (h_i^2)^2 \quad (13)
\]

Now let \( r \) be such that \( e_r(t) \geq 0 \) and \( e_{r+1}(t) \leq 0 \) and notice:

\[
\| \tilde{h} \| = \sum_i h_i^2
\]

\[
= \sum_{i=1}^{n} (e_i(t) - e_m(t))^2
\]

\[
= \sum_{i=1}^{r} (e_i(t) - e_m(t))^2 + \sum_{i=r+1}^{n} (|e_i(t)| + e_m(t))^2
\]

\[
= \sum_{i=1}^{r} e_i^2(t) + \sum_{i=r+1}^{n} |e_i(t)| + e_m(t)
\]

\[
= \sum_{i=1}^{r} e_i^2(t) + \sum_{i=r+1}^{n} |e_i(t)|
\]

Since \( \varepsilon(t) \) is a probability charge we have \( \sum_{i=1}^{n} e_i(t) = \sum_{i=r+1}^{n} |e_i(t)| = 0 \) and the above becomes:

\[
\| \tilde{h} \| = \sum_{i=1}^{n} e_i^2(t) + |V| e_m^2(t)
\]

\[
= \| \varepsilon(t) \| + |V| e_m^2(t)
\]

\[
\geq \| \varepsilon(t) \|
\]

Finally (14) and (15) complete the proof of Theorem 2.1:

\[
\| \varepsilon(t) \| - \| \varepsilon(t+1) \| \geq \alpha^2 \| \varepsilon(t) \|
\]

Remark: In terms of vertex expansion \( \beta = \min_{x} |\Gamma(A)|/|A| \), we can replace stage 3 by Alon’s treatment of eigenvectors [Alo86] and show: \( \| \varepsilon(t) \| - \| \varepsilon(t+1) \| \geq \alpha^2 \| \varepsilon(t) \| \)
3 General Markov Chains

Consider an irreducible and aperiodic Markov chain over state space $S$, $|S| = n$ (irreducibility and aperiodicity are assumed simply to guarantee unique convergence). Let $P = (p_{ij})$ denote its transition matrix, $\pi(0)$ the initial probability distribution, and $\pi$ the unique stationary distribution. As before, we are concerned with the rate at which $\pi(t)$ approaches $\pi$; equivalently, the rate at which the *discrepancy* $\varepsilon(t) = \pi(t) - \pi$ vanishes. We will characterize this rate in terms of expansion-like properties of the graph of the ergodic parameters associated with $P$.

Previously, results of this nature were obtained by Sinclair and Jerrum [S187]. Sinclair and Jerrum associated with $P$ the underlying graph of $P$: $G_P(S, W)$, where $w_{ij} = \pi_i p_{ij}$. $G_P$ is the weighted graph of ergodic flows of $P$. They further defined the *conductance* $\Phi_P(A)$ of a subset $A$ of $S$ as:

$$\Phi_P(A) = \frac{\sum_{i \in A} \sum_{j \in S \setminus A} w_{ij}}{\sum_{i \in A} \pi_i}$$

The conductance $\Phi_P$ of $P$ is:

$$\Phi_P = \min_{A \subset S: \sum_{i \in A} \pi_i \leq 1/2} \Phi_P(A)$$

Notice that the conductance is the weighted edge analogue of cutset expansion. For the case of strongly aperiodic (i.e. $\pi_i \geq 1/2$ for all $i$) and time-reversible (i.e. $w_{ij} = w_{ji}$) Markov chains Sinclair and Jerrum's bounds imply:

$$\| \varepsilon(t) \| \leq (1 - \Phi_P^2/2)^t \| \varepsilon(0) \|$$

where throughout this section

$$\| \varepsilon(t) \| = \sum_{i} \varepsilon_i^2(t) / \pi_i$$

For time-reversible Markov chains where $w_{ij} = w_{ji}$ the matrix $W$ is symmetric and possesses an orthogonal basis of eigenvectors. Sinclair and Jerrum's analysis makes strong use of the symmetry of $W$ and their argument is algebraic. For general non-reversible and strongly aperiodic Markov chains we can show a slightly better bound:

**Theorem 3.1** $\| \varepsilon(t) \| \leq (1 - \Phi_P^2)^t \| \varepsilon(0) \|$
they are tight. Such an improvement has been obtained by Diaconis [D89] for chains amenable to “canonical-path” arguments for conductance. (b): To explore if the bounds obtained here yield simple algorithms to approximate the expansion of a graph, like the ones proposed in [Alo86] and [BS87]. (c): To extend Aldous’s sample-averaging result for arbitrary Markov chains (Proposition 4.1 in [Ald87], which, by the way, gives a remarkable upper bound on random resources). (d): To check how much of the linear algebra used in previous statements concerning expanders was actually necessary.

Acknowledgment

I wish to thank Umesh Vazirani for the generous feedback of enlightening ideas that he provided, as usual. I also wish to thank Alistair Sinclair and Madhu Sudan for several important remarks.

References


