Logarithmic Sobolev techniques for random walks on graphs

F. R. K. Chung
University of Pennsylvania
Philadelphia 19104

Abstract

Recently, Diaconis and Sarloff-Coste used logarithmic Sobolev inequalities to improve convergence bounds for random walks on graphs. We will give a strengthened version by showing that the random walk on a graph G on n vertices reach the stationarity (under total variation distance) after about $\frac{1}{4\alpha}\log\log n$ steps where α denotes the log-Sobolev constant. Under the relative pointwise distance (which is a slight stronger notion), the random walk converges in about $\frac{1}{2\alpha}\log\log n$ steps.

1 Random walks on graphs

In a graph G, a walk is just a sequence of vertices $(x_0, x_1, \cdots, x_s)$ with $\{x_{i-1}, x_i\} \in E(G)$ for all $1 \le i \le s$. A random walk is determined by the transition probabilities $P(u,v) = Prob(x_{i+1} = v | x_i = u)$, which is independent of i. Clearly, for each vertex u,

$$\sum_v P(u,v) = 1.$$

For any initial distribution $f : V \to \mathbb{R}$ with $\sum_v f(v) = 1$, the distribution after k steps is just fP^k (i.e., a matrix multiplication with f viewed as a row vector where P is the matrix of transition probabilities). The random walk is said to be *ergodic* if there is a unique stationary distribution $\pi(v)$ satisfying

$$\lim_{s\to\infty} fP^s(v) = \pi(v).$$

Necessary and sufficient conditions for the ergodicity of P are (i) *irreducibility*, *i.e.*, for any $u, v \in V$, there exists some s such that $P^s(u,v) > 0$ (ii) *aperiodicity*, *i.e.*, gcd $\{s : P^s(u,v) > 0\} = 1$. The problem of interest is to determine the number of steps s required for P^s to be *close* to its stationary distribution, given an arbitrary initial distribution.

We say a random walk is *reversible* if

$$\pi(u)P(u,v) = \pi(v)P(v,u).$$

An alternative description for a reversible random walk can be given by considering a weighted connected graph with edge weights satisfying

$$w(u,v) = w(v,u) = \pi(v)P(v,u)/c$$

where c can be any constant to be chosen for simplifying the values. (For example, we can take c to be the average of $\pi(v)P(v,u)$ over all (v,u) with $P(v,u) \neq 0$, so that the values for $w(v,u)$ are either 0 or 1 for a simple graph.) The random walk on a weighted graph has as its transition probabilities

$$P(u,v) = \frac{w(u,v)}{d_u},$$

where $d_u = \sum_z w(u,z)$ is the (weighted) degree of u. The two conditions for ergodicity are equivalent to the conditions that the graph be (i) connected and (ii) not bipartite. To simplify notation and eliminate possible confusion, for a random walk problem, we will just deal with the associated weighted graph. In particular, in the next section we will discuss the Laplacian and the heat kernel of a graph which are self-adjoint and very useful for understanding the behavior of the random walk.

2 The Laplacian and heat kernel of a weighted graph

A weighted undirected graph G (possibly with loops) has associated with it a weight function $w : V \times V \to \mathbb{R}$ satisfying

$$w(u,v) = w(v,u)$$

and

$$w(u,v) \geq 0.$$

We note that if $\{u,v\} \notin E(G)$, then $w(u,v) = 0$. A simple (unweighted) graph is just the special case where all the weights are 0 or 1. The degree d_v of a vertex v is defined to be:

$$d_v = \sum_u w(u,v).$$

We define

$$L(u,v) = \begin{cases} d_v - w(v,v) & \text{if } u = v, \\ -w(u,v) & \text{if } u \text{ and } v \text{ are adjacent}, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, for a function $f : V \to \mathbb{R}$, we have

$$Lf(x) = \sum_{\substack{y \\ x \sim y}} (f(x) - f(y))w(x, y).$$

Let T denote the diagonal matrix with the (v, v)-th entry having value d_v. The *Laplacian* of G is defined to be

$$\mathcal{L} = T^{-1/2} L T^{-1/2}.$$

In other words, we have

$$\mathcal{L}(u, v) = \begin{cases} 1 - \dfrac{w(v, v)}{d_v} & \text{if } u = v, \\ -\dfrac{w(u, v)}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

Since $\mathcal{L}$ is symmetric, its eigenvalues which are all real and non-negative are denoted by

$$0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$$

where $n = |V|$. We can use the variational characterization of the eigenvalues as follows:

$$\begin{aligned} \lambda_G := \lambda_1 &= \inf_{g \perp T^{1/2}\mathbf{1}} \frac{\langle g, \ \mathcal{L} g \rangle}{\langle g, \ g \rangle} \qquad (1) \\ &= \inf_{\substack{f \\ \sum f(x) d_x = 0}} \frac{\displaystyle\sum_{x \in V} f(x) L f(x)}{\displaystyle\sum_{x \in V} f^2(x) d_x} \\ &= \inf_{\substack{f \\ \sum f(x) d_x = 0}} \frac{\displaystyle\sum_{x \sim y} (f(x) - f(y))^2 w(x, y)}{\displaystyle\sum_{x \in V} f^2(x) d_x}. \end{aligned}$$

For a connected graph G, the eigenvalues satisfy

$$0 < \lambda_i \leq 2$$

for $i \geq 1$. Various properties of the eigenvalues can be found in [4].

Suppose we write

$$\mathcal{L} = \sum_{i=0}^{n-1} \lambda_i I_i$$

where I_i is the projection to the ith eigenfunction ϕ_i of the graph. For any $t \geq 0$, the heat kernel H_t of G is defined to be the $n \times n$ matrix

$$\begin{aligned} H_t &= \sum_i e^{-\lambda_i t} I_i \\ &= e^{-t\mathcal{L}} \\ &= I - t\mathcal{L} + \frac{\epsilon^2}{2}\mathcal{L}^2 - \ldots \end{aligned}$$

In particular,

$$H_0 = I.$$

For a function $f : V \rightarrow \mathbb{R}$, we consider

$$\begin{aligned} F(x,t) &= \sum_{y \in S \cup \delta S} H_t(x,y) f(y) \\ &= (H_t f)(x). \end{aligned}$$

Then F satisfies the following properties (see [4]):

(i) $F(x, 0) = f(x)$,

(ii) For a fixed x,

$$\sum_y H_t(x,y)\sqrt{d_y} = \sqrt{d_x}$$

(iii) F satisfies the heat equation:

$$\frac{\partial F}{\partial t} = -\mathcal{L}F$$

(iv)

$$\mathcal{L}F(x,t) = \sum_{\substack{y \\ \{x,y\}}} \left(\frac{F(x,t)}{\sqrt{d_x}} - \frac{F(y,t)}{\sqrt{d_y}}\right) = 0$$

(v) $\displaystyle\sum_{\{x,y\} \in E} \left(\frac{F(x,t)}{\sqrt{d_x}} - \frac{F(y,t)}{\sqrt{d_y}}\right)^2 w(x,y) = \sum_x F(x,t)\mathcal{L}F(x,t)$

3 The rate of convergence for random walks

In a random walk with an associated weighted connected graph G, the transition matrix P satisfies

$$\mathbf{1}TP = \mathbf{1}T$$

and therefore the stationary distribution is exactly $\mathbf{1}T/\mathrm{vol}\ G$, where $\mathrm{vol}\ G = \sum_x d_x$. We want to show that when k is large enough, for any initial distribution $f : V \to \mathbb{R}$, fP^k converges rapidly to its stationary distribution.

Here $\|\cdot\|$ denotes the L_2 norm. We have

$$\begin{aligned} \|fP^s - \mathbf{1}T/\mathrm{vol}G\| &\le \|fT^{-1/2}(I-\mathcal{L})^s T^{1/2} - I_0\| \\ &\le \|T^{-1/2}(\sum_{i\neq 0} I_i^s)T^{1/2}\|\,\|f\| \\ &\le (1-\lambda)^s\|f\| \end{aligned}$$

where

$$\lambda = \begin{cases} \lambda_1 & \text{if } 1-\lambda_1 \ge \lambda_{n-1}-1 \\ 2-\lambda_{n-1} & \text{otherwise.} \end{cases} \tag{2}$$

So, after $s \ge (1/\lambda)\log(1/\epsilon)$ steps, the L_2 distance between fP^s and its stationary distribution is less than $\epsilon\|f\|$.

We note that the convergence of the random walk P^s is related to the heat kernel h_s as follows:

$$\begin{aligned} \|P^s - T^{-1/2}I_0T^{1/2}\| &= \|\sum_{i\neq 0} I_i^s T^{1/2}\| \\ &\le \|H_s - I_0\| \\ &\le e^{-s\lambda} \end{aligned} \tag{3}$$

Although λ occurs in the above upper bound for the distance between the stationary distribution and the s-step distribution, in fact, only λ_1 is crucial in the following sense. Note that λ is either λ_1 or $2-\lambda_{n-1}$. Suppose the latter holds, *i.e.*, $\lambda_{n-1} - 1 \ge 1 - \lambda_1$. We can consider a modified random walk, called the lazy walk, on the graph G' formed by adding a loop of weight d_v to each vertex v. The new graph has Laplacian eigenvalues $\tilde{\lambda}_k = \lambda_k/2 \le 1$, which follows from equation (1). Therefore,

$$1 - \tilde{\lambda}_1 \ge 1 - \tilde{\lambda}_{n-1} \ge 0,$$

and the convergence bound in L_2 distance in (4) for the modified random walk becomes

$$2/\lambda_1 \log(\frac{\max_x \sqrt{d_x}}{\epsilon \min_y \sqrt{d_y}}).$$

In general, suppose a weighted graph with edge weights $w(u,v)$ has eigenvalues λ_i with $\lambda_{n-1} - 1 \geq 1 - \lambda_1$. We can then modify the weights by choosing, for some constant c,

$$w'(u,v) = \begin{cases} w(v,v) + cd_v & \text{if } u = v \\ w(u,v) & \text{otherwise.} \end{cases} \tag{4}$$

The resulting weighted graph has eigenvalues

$$\lambda'_k = \frac{\lambda_k}{1+c} = \frac{2\lambda_k}{\lambda_{n-1} + \lambda_k}$$

where

$$c = \frac{\lambda_1 + \lambda_{n-1}}{2} - 1 \leq \frac{1}{2}.$$

Then we have

$$1 - \lambda'_1 = \lambda'_{n-1} - 1 = \frac{\lambda_{n-1} - \lambda_1}{\lambda_{n-1} + \lambda_1}.$$

In particular we set

$$\lambda = \lambda'_1 = \frac{2\lambda_1}{\lambda_{n-1} + \lambda_1}.$$

Therefore the modified random walk corresponding to the weight function w' has an improved bound for the convergence rate in L_2 distance:

$$\frac{1}{\lambda} \log \frac{\max_x \sqrt{d_x}}{\epsilon \min_y \sqrt{d_y}}$$

where $\lambda = \lambda_1$ if $\lambda_{n-1} + \lambda_1 \leq 2$ and $\lambda = 2\lambda_1/(\lambda_{n-1} + \lambda_1)$ otherwise. Note that $\lambda \geq 2\lambda_1/(2 + \lambda_1) \geq 2\lambda_1/3$.

We remark that for many applications in sampling, the convergence in L_2 distance seems to be too weak since it does not capture the convergence at each vertex. A stronger notion of convergence is measured by the relative pointwise distance, which is defined as follows (also see [8]): After s steps, the *relative pointwise distance* (r.p.d.) of P to the stationary distribution $\pi(x)$ is given by

$$\Delta(s) = \max_{x,y} \frac{|P^s(y,x) - \pi(x)|}{\pi(x)}.$$

Let f_x denote

$$f_x(y) = \begin{cases} 1 & \text{if } y = x, \\ 0 & \text{otherwise,} \end{cases}$$

We have

$$\begin{aligned}
\Delta(t) &= \max_{x,y} \frac{|f_y\,(P^t)\,f_x - \pi(x)|}{\pi(x)} \\
&= \max_{x,y} \frac{|f_y\,T^{-1/2}\,(I-\mathcal{L})^t\,T^{1/2}\,f_x - \pi(x)|}{\pi(x)} \\
&= \max_{x,y} |f_y T^{-1/2} (\sum_{i\neq 0} I_i^t) T^{1/2} f_x| \mathrm{vol} G \\
&\leq \max_{x,y} |f_y T^{-1/2} (H_t - I_0) T^{1/2} f_x| \mathrm{vol} G \\
&\leq e^{-t\lambda} \max_{x,y} \|T^{-1/2} f_x\|\ \|T^{-1/2} f_y\| \mathrm{vol} G \\
&\leq e^{-t\lambda} \frac{\mathrm{vol}\ G}{\min_x d_x}. \qquad (5)
\end{aligned}$$

So if we choose t such that

$$t \geq \frac{1}{\lambda} \log \frac{\mathrm{vol}\ G}{\epsilon \min_x d_x},$$

then, after t steps, we have $\Delta(t) \leq \epsilon$ where λ is as defined in (2) or for the lazy walk as defined in (4), λ can be taken to be

$$\lambda = \begin{cases} \lambda_1 & \text{if } 1-\lambda_1 \geq \lambda_{n-1} - 1 \\ \dfrac{2\lambda_1}{\lambda_{n-1}+\lambda_1} & \text{otherwise} \end{cases} \qquad (6)$$

Here we note that the factor of $\frac{\mathrm{vol}\ G}{\min_x d_x}$ in (5) can be further reduced by the Logarithmic Sobolev techniques which we will discuss in the next section.

4 The log-Sobolev constant

Let G denote a weighted graph on n vertices. For a function $f : V(G) \to \mathbb{R}$. We may view f as a column vector, $1 \times n$ matrix or a row vector. The stationary distribution $\pi(x) = d_x/\mathrm{vol}\ G$ will be viewed as a column or row vector. Let Π denote the diagonal matrix with value $\pi(x)$ as the (x, x)-entry.

The log-Sobolev constant α of a weighted graph G is the least constant satisfying the following *log-Sobolev inequality* for any nontrivial function $f : V \to \mathbb{R}$:

$$\sum_{\{x,y\}\in E} (f(x) - f(y))^2 w_{x,y} \leq \alpha \sum_{x\in V} f^2(x) d_x \log \frac{f^2(x)\mathrm{vol}\ G}{\sum_{z\in V} f^2(z) d_z}$$

In other words, α can be expressed as follows:

$$\alpha_G = \alpha = \inf_{f \neq 0} \frac{\sum_{\{x,y\} \in E} (f(x) - f(y))^2 w_{x,y}}{\sum_{x \in V} f^2(x) d_x \log \frac{f^2(x) \text{vol } G}{\sum_{z \in V} f^2(z) d_z}} \tag{7}$$

where f ranges over all nontrivial functions $f : V \to \mathbb{R}$.

Logarithmic Sobolev inequalities first arose in the analysis of elliptic differential operators in infinite dimensions. Many developments and applications can be found in several survey papers [1, 6, 7, 9]. Diaconis and Salaff-Coste [5] introduced a discrete version of logarithmic sobolev inequality to prove that for a regula graph on n vertices

$$\Delta_{TV}(t) \leq e^{1-c} \quad \text{if} \quad t \geq \frac{1}{2\alpha} \log \log n + \frac{c}{\lambda}.$$

We will give a simple proof of the above result by deriving the following slightly stronger statements.

Theorem 1 *In a weighted graph G with log-Sobolev constant α, we have $\Delta(t) \leq e^{2-c}$ if*

$$t \geq \frac{1}{2\alpha} \log \log \frac{\text{vol } G}{\min_x d_x} + \frac{c}{\lambda}.$$

Theorem 2 *In a weighted graph G with log-Sobolev constant α, we have $\Delta_{TV}(t) \leq e^{1-c}/2$ if*

$$t \geq \frac{1}{4\alpha} \log \log \frac{\text{vol } G}{\min_x d_x} + \frac{c}{\lambda}.$$

The proofs for the above theorem will be given in the next section.

5 Proofs of the main theorems

For a function $f : V(G) \to \mathbb{R}$, we define the $(\pi; p)$-norm of f, denoted by ${}_\pi\|f\|_p$, to be

$${}_\pi\|f\|_p = \left(\sum_{x \in V(G)} f^p(x) \pi(x) \right)^{1/p}.$$

In particular,

$$_\pi\|f\|_2 = \left(\sum_x f^2(x)\pi(x)\right)^{1/2} = \|\Pi^{1/2} f\|_2.$$

The main proof for Theorems 1 and 2 consists of two parts. In the first part (Theorem 3), we will see that the inequality (8) relating the p-norm to the 2-norm, for certain p, implies the improved convergence bound for random walks. The second part (Theorem 5) states that the inequality (8) can be derived from the log-Sobolev inequality.

Theorem 3 *Suppose that in a weighted graph G, its heat kernel H_s satisfies*

$$_\pi\|f\ \Pi^{1/2}\ H_s\ \Pi^{-1/2}\|_p \quad \leq \quad _\pi\|f\|_2 \tag{8}$$

for all $f : V(G) \to \mathbb{R}$, and $p = e^{\beta s}$ for some positive value β. Then the random walk on G satisfies

$$\Delta(t) \leq\ e^{2-c}$$

if

$$t \geq \frac{2}{\beta} \log\log \frac{\text{vol}G}{\min_x d_x} + \frac{c}{\lambda}.$$

Proof: We define q satisfying

$$\frac{1}{p} + \frac{1}{q} = 1.$$

For a vertex x of G, let ψ_x denote the characteristic function satisfying $\psi_x(y) = 1$ if $x = y$, and 0 otherwise. For a function $f : V \to \mathbb{R}$, we consider

$$\begin{aligned}
& |\psi_x \Pi^{-1/2} H_s \Pi^{1/2} f| \\
= & \ |\left(\psi_x \Pi^{-1+1/q}\right)\ (\Pi^{1/p-1/2} H_s \Pi^{1/2} f)| \\
\leq & \ \left(\sum_y (\psi_x\ \Pi^{-1}(y))^q \pi(y)\right)^{1/q} \left(\sum_y (f\Pi^{1/2} H_s\ \Pi^{-1/2}(y))^p \pi(y)\right)^{1/p} \\
= & \ _\pi\|\psi_x \Pi^{-1}\|_q\ \ _\pi\|f\Pi^{1/2} H_s \Pi^{-1/2}\|_p
\end{aligned} \tag{9}$$

by using Hölder's inequality.

We consider

$$
\begin{aligned}
{}_\pi\|\psi_x \Pi^{-1}\|_q &= \left(\sum_y (\psi_x \ \Pi^{-1}(y))^q \pi(y)\right)^{1/q} \\
&= (\pi(x)^{1-q})^{1/q} \\
&= \pi(x)^{-1/p} \\
&\leq (\frac{\mathrm{vol} G}{\min_x d_x})^{1/p}.
\end{aligned}
$$

Using the hypothesis that $p = e^{\beta s}$ and the choice of s satisfying

$$
s = \frac{1}{\beta} \log\log \frac{\mathrm{vol} G}{\min_x d_x},
$$

we have

$$
(\frac{\mathrm{vol} G}{\min_x d_x})^{1/p} \quad = \quad e^{e^{\log\log \frac{\mathrm{vol} G}{\min_x d_x} - \beta s}} = e
$$

From (8) and (9), we have, for any f,

$$
\begin{aligned}
|\psi_x \Pi^{-1/2} H_s \Pi^{1/2} f| &\leq e \ {}_\pi\|f \Pi^{1/2} H_s \Pi^{-1/2}\|_p \\
&\leq e \ {}_\pi\|f\|_2.
\end{aligned}
$$

In particular, for the heat kernel and the projection I_0 into the 0-th eigenfunction, we have

$$
\begin{aligned}
|\psi_x \Pi^{-1/2}(H_{s+r} - I_0)\Pi^{1/2} f| &\leq |\psi_x \Pi^{-1/2} H_s (H_r - I_0) \Pi^{1/2} f| \\
&\leq e \ {}_\pi\|\Pi^{-1/2}(H_r - I_0)\Pi^{1/2} f\|_2 \\
&\leq e\|(H_r - I_0)\Pi^{1/2} f\|_2 \\
&\leq e\|(H_r - I_0)\|_2 \ \|\Pi^{1/2} f\|_2 \\
&\leq e^{1-\lambda r} \ {}_\pi\|f\|_2.
\end{aligned}
$$

This is equivalent to

$$
|\psi_x \Pi^{-1/2}(H_{s+r} - I_0) g| \leq e^{1-\lambda r}\|g\|_2
$$

for all g. This implies

$$
\|\psi_x \Pi^{-1/2}(H_{s+r} - I_0)\|_2 \quad \leq \quad e^{1-\lambda r}. \tag{10}
$$

Therefore, the random walk on G converges to the stationary distribution under relative pairwise distance as follows (see (5)):

$$\begin{aligned}\Delta(2s+2r) &\leq \max_{x,y} |\psi_x \Pi^{-1/2}(H_{2s+2r} - I_0)\Pi^{-1/2}\psi_y| \\ &\leq \max_{x,y} \|\psi_x \Pi^{-1/2}(H_{s+r} - I_0)\|_2 \cdot \|\psi_y \Pi^{-1/2}(H_{s+r} - I_0)\|_2 \\ &\leq e^{2-2\lambda r}\end{aligned}$$

by using (10) and the Cauchy schwarz inequality. Now, we take $r = \frac{c}{2\lambda}$, $t = 2s + 2r$, and the proof is complete. □

We can also obtain a similar statement for the convergence bound under the total variation distance.

Theorem 4 *Suppose that in a weighted graph G, its heat kernel H_s satisfies*

$$_{\pi}\|f \ \Pi^{1/2} \ H_s \ \Pi^{-1/2}\|_p \quad \leq \quad _{\pi}\|f\|_2$$

for all $f : V(G) \to \mathbb{R}$, and $p = e^{\beta s}$ for some positive value β. Then the random walk on G satisfies

$$\Delta_{TV}(t) \leq \frac{1}{2}e^{1-c}$$

if

$$t \geq \frac{1}{\beta}\log\log\frac{\text{vol}G}{\min_x d_x} + \frac{c}{\lambda}.$$

Proof: We follow the notation in Theorem 3.

$$\begin{aligned}\Delta_{TV} &= \frac{1}{2}\max_x \sum_y |\psi_x P^{s+r}(y) - \pi(y)| \\ &\leq \frac{1}{2}\max_x \sum_y |\psi_x \Pi^{-1/2}(H_{s+r} - I_0)\Pi^{1/2}(y)| \\ &\leq \frac{1}{2}\max_x \sum_y e^{1-\lambda r}\pi(y) \\ &\leq \frac{1}{2}e^{1-\lambda r}\end{aligned}$$

by using (10). □

Now we proceed to show that the log-Sobolev constant can be used to determine β in the above theorems. This proof is very similar to the continuous case (see [5]).

Theorem 5 *In a graph G with log-Sobolev constant α, its heat kernel H_t satisfies*

$$_\pi\| f\ \pi^{1/2}\ H_t\ \pi^{-1/2}\|_p \quad \le \quad _\pi\|f\|_2$$

for any $t > 0$, $p = e^{4\alpha t} + 1$, and for any $f : V(G) \to \mathbb{R}$.

Proof: From the definition of α, we have

$$\sum_{x\sim y}(f(x) - f(y))^2 w(x,y) \ge \alpha \sum_x f^2(x) d_x \log \frac{f^2(x)^2}{\sum_z f^2(z)\pi(z)}$$

for any nontrivial function f. In particular, we can replace f by $f^{p/2}$ and we have

$$\sum_{x\sim y}(f^{p/2}(x) - f^{p/2}(y))^2 w(x,y) \ge \alpha \sum_x f^p(x) d_x \log \frac{f(x)^p}{\sum_z f^p(z)\pi(z)}. \tag{11}$$

Now we need the following inequality which is not hard to prove:

$$4(p-1)(a^{p/2} - b^{p/2})^2 \le p^2(a-b)(a^{p-1} - b^{p-1}). \tag{12}$$

for all $a, b \ge 0$ and $p \ge 1$. From (11) and (12), we have

$$\begin{aligned}
& \alpha \sum_x f^p(x)\pi(x) \log \frac{f(x)^p}{\sum_z f^p(z)\pi(z)} \\
\le \quad & \sum_{x\sim y}(f^{p/2}(x) - f^{p/2}(y))^2 w(x,y) \\
\le \quad & \frac{p^2}{4(p-1)} \sum_{x\sim y}(f^{p-1}(x) - f^{p-1}(y))(f(x) - f(y)) w_{x,y}
\end{aligned} \tag{13}$$

We now replace f by $g = f\pi^{1/2} H_t \pi^{-1/2}$ in the above inequality and define p as a function of t:

$$p = p(t) = 1 + e^{4\alpha t}.$$

Note that $p' = p'(t) = 4\alpha(p-1)$. From (13), we have

$$\frac{p'}{p^2}\sum_x g^p(x)\pi(x)\log\frac{|g(x)|^p}{\sum_z g^p(z)\pi(z)} - \sum_{x\sim y}(g^{p-1}(x) - g^{p-1}(y))(g(x) - g(y)) w_{x,y} \le 0 \tag{14}$$

Now we define

$$F(t) = \ _\pi\|g\|_p.$$

Clearly, $F(0) = {}_{\pi}\|f\|_2$. If we can show that the derivative $F'(t) \leq 0$, then we have ${}_{\pi}\|g\|_p = F(t) \leq F(0) = {}_{\pi}\|f\|_2$ as desired. It remains to show $F'(t) \leq 0$. Since

$$F(t) = \left(\sum_x (f\Pi^{1/2} H_t \Pi^{-1/2}(x))^p \pi(x)\right)^{1/p} = (G(t))^{1/p},$$

we have

$$F'(t) = \left(-\frac{p'}{p^2}\log G(t) + \frac{G'(t)}{pG(t)}\right) F(t). \tag{15}$$

We note that

$$\begin{aligned} G'(t) &= p\sum_x g^{p-1}\pi(x)(f\Pi^{1/2}\frac{d}{dt}H_t\Pi^{-1/2}(x)) + p'\sum_x g^p(x)\pi(x)\log g(x) \\ &= I + II \end{aligned}$$

We consider the above sum of I as a product of matrices (where A^* denotes the transpose of A):

$$\begin{aligned} I &= p\ g^{p-1}\Pi^{1/2}\frac{d}{dt}H_t\Pi^{1/2}f^* \\ &= -p\ g^{p-1}\Pi^{1/2}\mathcal{L}H_t\Pi^{1/2}f^* \\ &= -p\ g^{p-1}\Pi^{1/2}\mathcal{L}\Pi^{1/2}g^* \\ &= -\frac{p}{\text{vol } G}\sum_{x\sim y}(g^{p-1}(x) - g^{p-1}(y))(g(x)-g(y))w_{x,y} \end{aligned}$$

by using the heat equation in the weighted version of Lemma 10.3. Substituting into (15), we obtain

$$\begin{aligned} & F'(t) \\ =\ & \frac{p'}{p^2}(\sum_x g^p(x)\pi(x)\log g^p(x) - \log G(t)) - \frac{1}{\text{vol } G}\sum_{x\sim y}(g^{p-1}(x) - g^{p-1}(y))(g(x)-g(y)) \\ =\ & \frac{1}{\text{vol } G}\left(\frac{p'}{p^2}\sum_x\sum_x g^p(x)d_x\log\frac{g^p(x)}{\log G(t)} - \sum_{x\sim y}(g^{p-1}(x)-g^{p-1}(y))(g(x)-g(y))w_{x,y}\right) \\ \leq\ & 0 \end{aligned}$$

by using (14). Theorem 5 is proved. □

Together with Theorem 3 and 4, we have completed the proofs for the main results in Theorem 1 and 2.

References

[1] D. Bakry, L'hypercontractivité et son utilisation en théorie des semigroups, In *Ecole d' été de Saint Flour 1992*, Springer LNM 1581.

[2] W. Beckner, Inequalities in fourier analysis, *Annals of Mathematics* 102 (1975) 159-182.

[3] F. R. K. Chung and S. -T. Yau, Eigenvalues of graphs and Sobolev inequalities, *Combinatorics, Probability and Computing*, 4 (1995) 11-26.

[4] F. R. K. Chung, *Spectral Graph Theory*, CBMS Lecture Notes, 1996 , AMS Publication.

[5] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains, preprint.

[6] L. Gross, Logarithmic Sobolev inequalitities, *Amer. J. Math.* 97, (1976) 1061-1083.

[7] L. Gross, *Logarithmic Sobolev inequalitities and contractivity properties of semi-groups*, in Lecture Notes in Math. (1993) Springer LNM 1563.

[8] A.J. Sinclair, *Algorithms for Random Generation and Counting*, Birkhauser, 1993.

[9] D. Stroock, Logarithmic Sobolev inequalitities for Gibbs states, in Lecture Notes in Math. (1993) Springer LNM 1563.