

Modified Logarithmic Sobolev Inequalities in Discrete Settings

Sergey G. Bobkov^{1,3} and Prasad Tetali²

Received June 15, 2004; revised December 3, 2004

Motivated by the rate at which the entropy of an ergodic Markov chain relative to its stationary distribution decays to zero, we study modified versions of logarithmic Sobolev inequalities in the discrete setting of finite Markov chains and graphs. These inequalities turn out to be weaker than the standard log-Sobolev inequality, but stronger than the Poincaré (spectral gap) inequality. We show that, in contrast with the spectral gap, for bounded degree expander graphs, various log-Sobolev constants go to zero with the size of the graph. We also derive a hypercontractivity formulation equivalent to our main modified log-Sobolev inequality. Along the way we survey various recent results that have been obtained in this topic by other researchers.

KEY WORDS: Spectral gap; entropy decay; logarithmic Sobolev Inequalities.

MSC 2000: 60Jxx, 46Exx.

1. INTRODUCTION

Let (M, P, π) be an ergodic, reversible Markov chain with a finite state space M , transition probability matrix P and stationary distribution π . For $f, g: M \rightarrow \mathbf{R}$, let $\mathcal{E}(f, g)$ denote the Dirichlet form defined by

$$\mathcal{E}(f, g) = -\mathbf{E}_\pi(fLg) = -\sum_{x \in M} f(x)Lg(x)\pi(x), \quad (1.1)$$

¹ Department of Mathematics, University of Minnesota, Minneapolis, MN, USA. E-mail: bobkov@math.umn.edu

² School of Mathematics and College of Computing, Georgia Tech, Atlanta, GA, USA. E-mail: tetali@math.gatech.edu

³ To whom Correspondence should be addressed.

where $-L = I - P$ is the associated Laplacian matrix. Then the spectral gap of P or the smallest non-zero eigenvalue of $-L$ can be defined as the optimal positive constant in Poincaré-type inequality

$$\lambda_1 \text{Var}_\pi(f) \leq \mathcal{E}(f, f), \tag{1.2}$$

over all $f: M \rightarrow \mathbf{R}$. As usual, $\text{Var}_\pi(f) = \mathbf{E}_\pi f^2 - (\mathbf{E}_\pi f)^2$ stands for the variance and $\mathbf{E}_\pi f = \int f d\pi$ for the expectation of f with respect to the measure π . One arrives at such a functional (or variational) definition of the spectral gap in a natural way by considering the rate of decay of variance of the distribution of the chain with respect to the stationary distribution. More formally, working in the technically easier continuous time, let $\mu_t = \mu_0 P_t$ be the distribution of the chain at time t , for $t \geq 0$, where $P_t = e^{tL}$ is the semi-group generated by L . Then it is a classical fact that the density $f_t(x) = \frac{\mu_t(x)}{\pi(x)}$, $x \in M$, satisfies the identity

$$\frac{d}{dt} \text{Var}_\pi(f_t) = -2\mathcal{E}(f_t, f_t), \tag{1.3}$$

thus motivating the above definition of λ_1 . On the other hand, only recently attention seems to have been given to the following equally natural property: for all $t > 0$,

$$\frac{d}{dt} D(\mu_t \| \pi) = -\mathcal{E}(f_t, \log f_t), \tag{1.4}$$

where $D(\mu \| \pi) = \sum_{x \in M} \mu(x) \log \frac{\mu(x)}{\pi(x)}$ denotes the so-called informational divergence or the relative entropy of μ with respect to π . Using the standard notation for the entropy functional $\text{Ent}_\pi(f) = \mathbf{E}_\pi f \log f - \mathbf{E}_\pi f \log \mathbf{E}_\pi f$, one is now motivated in studying the inequality

$$\rho_0 \text{Ent}_\pi(f) \leq \frac{1}{2} \mathcal{E}(f, \log f), \tag{1.5}$$

in the class of all positive f on M , since then $\frac{d}{dt} D(\mu_t \| \pi) \leq -2\rho_0 D(\mu_t \| \pi)$. Together with a well-known relation between the total variation norm $\|\mu_t - \pi\|_{\text{TV}} = \sum_{x \in M} |\mu_t(x) - \pi(x)|$ and the relative entropy, the latter leads to the bound

$$\|\mu_t - \pi\|_{\text{TV}}^2 \leq 2 \log \frac{1}{\pi_*} e^{-2\rho_0 t}, \quad t \geq 0, \tag{1.6}$$

where $\pi_* = \min_{x \in M} \pi(x)$. This recovers and in fact improves upon a similar bound, due to Diaconis and Saloff-Coste,⁽¹⁵⁾ employing the logarithmic Sobolev constant ρ , the best one in the standard logarithmic Sobolev inequality

$$\rho \text{Ent}_\pi(f^2) \leq 2\mathcal{E}(f, f). \tag{1.7}$$

As is shown in,⁽¹⁵⁾ for the time $\tau_2 = \inf\{t > 0: \sup_{\mu_0}[\mathbf{E}_\pi |f_t - 1|^2]^{1/2} \leq 1/e\}$ “reach stationarity,” one has

$$\frac{1}{4\rho} \leq \tau_2 \leq \left(1 + \frac{1}{4} \log \log \frac{1}{\pi_*}\right) \frac{1}{2\rho}.$$

Therefore, while ρ captures rather accurately the convergence to stationarity in terms of $\sup_{\mu_0}[\mathbf{E}|f_t - 1|^2]^{1/2}$, which in general is larger than $D(\mu_t \|\pi)$, it seems better to use ρ_0 when one wants to work with either the relative entropy or the total variation norm.

In this paper we aim to study a number of general properties of this constant in the framework of abstract Markov kernels and finite graphs and put it in the hierarchy of various Sobolev-type inequalities actively circulating in the literature in the recent years. In general, the inequality (1.5), defining ρ_0 , may be viewed as a modified form of L. Gross’ inequality (1.7) – a concept, suggested in 1996 by Ledoux⁽²⁶⁾ in connection with the concentration of measure phenomenon. To be more precise, he studied the form involving the length of the gradient,

$$\rho_1 \text{Ent}_\pi(e^f) \leq \int |\nabla f|^2 e^f d\pi, \tag{1.8}$$

as an approach to some of M. Talagrand’s deviation inequalities for product probability measures. Afterwards, some other modifications of (1.8) have appeared in the literature, especially in the discrete settings, cf. e.g. Refs. 6, 7, 27, 28. In particular, the inequality (1.5) was considered by Ledoux⁽²⁸⁾ for product measures on the discrete cube and, as an application, for Poissonian limits. While preparing the present paper, we learnt that the inequality (1.5) was also introduced by Dai Pra, Paganoni and Posta (see Ref. 13 where ρ_0 is referred to as the “entropy constant”) in the context of certain Gibbs measures on \mathbf{Z}^d . In particular, they showed examples of measures which fail to satisfy the classical inequality (1.7), while satisfying (1.5). The crucial identity (1.4) justifying (1.5) in the framework of Markov kernels was probably first observed in the early 90’s by Bakry⁽²⁾ and Stroock,⁽³⁷⁾ and later was used in bounds involving ρ rather than ρ_0 , cf. e.g. Refs. 15, 19. We also learnt that Gao and Quastel⁽²⁰⁾ have recently considered ρ_0 and using martingale tools obtained for it a lower estimate on the symmetric group and for slices of the discrete cube. Unaware of⁽²⁰⁾ and our present work, these examples and some others (such as the top-to-random card shuffle) have been derived by Goel.⁽²¹⁾ A very detailed analysis of the decay of relative entropy in the specific example of the top-to-random shuffle is reported in.⁽³⁶⁾ It should therefore be emphasized that, since in many interesting examples ρ_0 is much

better than ρ , the role of (1.5) needs to be explored in terms of various applications regardless of (1.7). In an upcoming paper⁽³⁵⁾ some of the ideas from this paper are being further employed in developing the so-called transposition inequalities in discrete settings.

These notes (which are an extended and a detailed version of Ref. 8) are organized as follows. In Section 2, we formalize the argument leading to the bound (1.6) under (1.5) in the general non-reversible case. In Section 3, we observe that

$$\rho \leq \rho_0 \leq \rho_1 \leq \lambda_1,$$

and discuss several standard examples illustrating these relations. In many of them, ρ_0 is in fact the order of λ_1 , while $\rho = o(\rho_0)$, thus providing tight bounds on convergence to stationarity using the total variation norm. In Section 4, we introduce a family of inequalities that interpolate between the modified logarithmic Sobolev inequality (1.5) and the Poincaré inequality (1.1). In application to Markov chains, these more general inequalities allow one to control a distance-like quantity $\mathbb{E}_\pi f_t^p - 1$ (between μ_t and π) uniformly over all $p \in (1, 2]$.

It is natural to wonder how the relative entropy decays for random walks on expander graphs (namely the graphs for which λ_1 is bounded away from zero). In Section 5, it is shown (using ρ_1) that in fact both ρ and ρ_0 are of the order of $1/\log |G|$ for bounded degree expanders G . Here we also discuss a concentration content of (1.5) in terms of deviations of Lipschitz functions on G .

Note that the eigenvalue interpretation of λ_1 tells us that there is a function (namely an eigenfunction) which achieves the optimal value in the variational description of λ_1 . The same is not necessarily true for ρ and ρ_0 (e.g. as in the symmetric two-point space). In such a case, it turns out that if the inf in the definition of ρ_0 is not achieved, then in fact $\rho_0 = \lambda_1$! The analogous result for ρ was known to be true, implicit in the work of Rothaus⁽³³⁾ in the context of manifolds and also appears explicitly in the context of finite Markov chains in.⁽³⁴⁾ These questions are discussed in Section 6. Finally, in the last (7th) section we give a hypercontractivity characterization of ρ_0 .

2. CONVERGENCE TO STATIONARITY

Elaborating on the introduction, we start with a stochastic matrix P on a finite set M , and define a Markov process $\{X_t\}_{t \geq 0}$ in M with initial

distribution, say, μ_0 and transition matrices

$$P_t = e^{-t(I-P)}, \quad t \geq 0,$$

with the generator $-L = I - P$. To study the asymptotic behavior of the probability distributions μ_t of random variables X_t for large time, we will assume that:

- (a) There is a stationary distribution π for P , i.e. $\pi P = \pi$.
- (b) $\pi(x) > 0$, for all $x \in M$.
- (c) For all $x, y \in M$, there exists $n \geq 1$ with $P^n(x, y) > 0$.

Since M is finite, the irreducibility property c) implies that such a π exists and is unique, see e.g. [18], and it is easy to see that (c) and (a) in turn imply (b). Moreover, by b), any probability distribution μ on M is absolutely continuous with respect to π .

Thus, let $f_t(x) = \frac{\mu_t(x)}{\pi(x)}$, $x \in M$, be the density of μ_t with respect to π at time $t \geq 0$. As is well known, the measures μ_t approach π , or equivalently, f_t 's are getting close to 1 for large t . A proper quantitative statement may be done, for example, in terms of L^p -distance

$$\|f_t - 1\|_{L^p(\pi)}^p = \int |f_t - 1|^p d\pi, \quad 1 \leq p < +\infty,$$

which becomes the total variation norm $\|\mu_t - \pi\|_{TV}$ in case $p=1$. Another important measure of closeness is the informational divergence,

$$D(\mu_t \|\pi) = \text{Ent}_\pi(f_t) = \int f_t \log f_t d\pi.$$

Recall that $\mu_t = \mu_0 P_t$. Let P^* denote the time-reversal of P defined by the identity $\pi(x)P^*(x, y) = \pi(y)P(y, x)$, $x, y \in M$. Let

$$P_t^* = e^{tL^*} = \sum_{n=0}^{\infty} \frac{t^n (L^*)^n}{n!},$$

be the semigroup associated to the dual $-L^* = I - P^*$. Then the following is a useful basic technical fact.

Lemma 2.1. For any μ_0 and all $t \geq 0$, $f_t = P_t^* f_0$. Consequently, for any $x \in M$,

$$\frac{df_t(x)}{dt} = L^* f_t(x).$$

For example, differentiating the function $\text{Var}_\pi(f_t) = \int f_t^2 d\pi - 1$, we get

$$\frac{d}{dt} \text{Var}_\pi(f_t) = \int \frac{d}{dt} f_t^2 d\pi = 2 \int f_t L^* f_t d\pi = 2 \int L(f_t) f_t d\pi = -2\mathcal{E}(f_t, f_t).$$

Therefore, starting from the Poincaré-type inequality (1.2) with a constant $\lambda_1 > 0$, one gets $\frac{d}{dt} \text{Var}_\pi(f_t) \leq -2\lambda_1 \text{Var}_\pi(f_t)$. Integrating over t , one arrives at the standard estimate:

Theorem 2.2. For every initial distribution μ_0 ,

$$\text{Var}_\pi(f_t) \leq \text{Var}_\pi(f_0) e^{-2\lambda_1 t}, \quad t \geq 0. \tag{2.1}$$

Now, let us repeat a similar argument towards the study of the informational divergence.

Lemma 2.3. For any μ_0 and all $t > 0$, the density f_t is strictly positive on M . Furthermore, the function $t \rightarrow D(\mu_t \parallel \pi)$ is differentiable on $(0, +\infty)$, and

$$\frac{d}{dt} D(\mu_t \parallel \pi) = -\mathcal{E}(f_t, \log f_t), \quad t > 0.$$

Proof. Given $x, y \in M$ and $t \geq 0$, write Taylor’s expansion

$$P_t^*(x, y) = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} (P^*)^n(x, y).$$

Since $\pi(x)(P^*)^n(x, y) = \pi(y)P^n(y, x)$, for all $n \geq 1$, with the assumptions $b) - c)$ we get $P_t^*(x, y) > 0$ whenever $t > 0$. Since $f_t = P_t^* f_0$ and $\sum_x f_0(x) = 1$, this yields the first statement of the lemma.

Thus, in the range $t > 0$, we are allowed to perform differentiation in accordance with Lemma 2.1 and the identity (1.1) for the Dirichlet form:

$$\begin{aligned} \frac{d}{dt} D(\mu_t \parallel \pi) &= \int \frac{d}{dt} f_t \log f_t d\pi \\ &= \int (\log f_t + 1) L^* f_t d\pi = \int L(\log f_t) f_t d\pi = -\mathcal{E}(f_t, \log f_t). \end{aligned}$$

□

Now, similarly to Theorem 2.2, we can start from the modified logarithmic Sobolev inequality (1.5) on M . By Lemma 2.3, $\frac{d}{dt} D(\mu_t \parallel \pi) \leq -2\rho_0 D(\mu_t \parallel \pi)$, for all $t > 0$. Integrating this inequality over t and since the right hand side is continuous at $t=0$, we arrive at:

Theorem 2.4. For every initial distribution μ_0 ,

$$D(\mu_t \|\pi) \leq D(\mu_0 \|\pi)e^{-2\rho_0 t}, \quad t \geq 0. \tag{2.2}$$

Remark 2.5. Note that we did not make the assumption of reversibility of the Markov kernel, namely, the assumption that $\pi(x)P(x, y) = \pi(y)P(y, x)$, for all $x, y \in M$. In the next section, assuming reversibility we will observe that $\rho \leq \rho_0$, thus showing that the estimate (2.2) improves upon

$$D(\mu_t \|\pi) \leq D(\mu_0 \|\pi)e^{-2\rho t}, \quad t \geq 0. \tag{2.3}$$

The latter was obtained by Bakry⁽²⁾ and Stroock,⁽³⁷⁾ see discussion in.⁽¹⁵⁾

Remark 2.6. Together with (2.3), Theorem 3.6 in⁽¹⁵⁾ also involves nonreversible Markov kernels in which case ρ is replaced with $\rho/2$ (a result of Miclo Ref. 31). However observe (as also pointed out in Ref. 15) that there is something inherently reversible about $\mathcal{E}(f, f)$ – namely, using the notation $\mathcal{E}_P(f, f)$ to indicate the dependence on the kernel, it is easily seen that

$$\mathcal{E}_P(f, f) = \mathcal{E}_{\hat{P}}(f, f), \quad \text{with } \hat{P} := \frac{1}{2}(P + P^*),$$

where P^* is the time-reversal of P . Hence one has in general, for the optimal constants in (1.2) and (1.7),

$$\lambda_1(P) = \lambda_1\left(\frac{1}{2}(P + P^*)\right), \quad \rho(P) = \rho\left(\frac{1}{2}(P + P^*)\right).$$

Thus the estimate (2.2) together with $\rho_0(\frac{1}{2}(P + P^*)) \geq \rho(\frac{1}{2}(P + P^*))$ (see Proposition 3.6 below) recovers the nonreversible case of [15] as well.

Remark 2.7. The above proof of Lemma 2.3 which led to Theorem 2.4 is implicitly contained in⁽¹⁵⁾ (or as paraphrased in the appendix of Ref. 19). The important difference is that the usual log-Sobolev inequality (3.2) is taken as a starting point in⁽¹⁹⁾ and in all the above-mentioned papers.

The estimates given in Theorems 2.2 and 2.4 are not comparable in general: each may have its own advantages. When $\rho_0 = \lambda_1$ or when these constants are of similar magnitude, the estimate (2.2) can be more useful than the estimate (2.1). First note, there is a general inequality $\int f d\pi \text{Ent}_\pi f \leq \text{Var}_\pi(f)$, holding true for any measurable function on an arbitrary probability space. Applying this to $f = f_t$, we get

$$D(\mu_t \|\pi) \leq \text{Var}_\pi(f_t).$$

Hence, in the second theorem, a smaller distance (the informational divergence at time t) is estimated from above by a smaller quantity (the informational divergence at the initial time multiplied by an exponentially decreasing factor).

Another natural and typical objective is obtaining the rates of convergence in total variation norm $\|\mu_t - \pi\|_{TV}$ uniformly over all possible μ_0 . Then, in order to apply (2.1), one can use the bound $\|\mu_t - \pi\|_{TV}^2 \leq \text{Var}_\pi(f_t)$. The right hand side of (2.1) is maximized when μ_0 is one of the Dirac measures δ_x which leads to

$$\|\mu_t - \pi\|_{TV}^2 \leq \frac{1}{\pi_*} e^{-2\lambda_1 t}, \quad \pi_* = \min_{x \in M} \pi(x). \tag{2.4}$$

It is also possible to relate the total variation norm to the informational divergence, using the following well known inequality, see e.g. Lemma 12.6.1 in Refs. 12, 15 or 19: for every probability measure μ on M ,

$$\|\mu_t - \pi\|_{TV}^2 \leq 2D(\mu\|\pi). \tag{2.5}$$

With estimate (2.2) this leads to a certain refinement of (2.4) in the case where ρ_0 is approximately λ_1 :

Corollary 2.8. For every initial distribution μ_0 on M , for all $t \geq 0$,

$$\|\mu_t - \pi\|_{TV}^2 \leq 2 \log \frac{1}{\pi_*} e^{-2\rho_0 t}. \tag{2.6}$$

A general 2-state chain can be used to show that in (2.6), the dependence on t can be sharp; (once again, ρ_0 and λ_1 are of the same order in such an example.)

It might also be worth mentioning that the bounds (2.4) and (2.6) can be sharpened by virtue of Theorem 2.2 under mild symmetry assumptions of the initial density f_0 about its mean value $\int f_0 d\pi = 1$. In particular, we have:

Corollary 2.9. For every initial distribution μ_0 such that $\int (f_0 - 1)^3 d\pi = 0$,

$$\text{Var}_\pi(f_t) \leq 2 \log \frac{1}{\pi_*} e^{-2\lambda_1 t}, \quad t \geq 0. \tag{2.7}$$

Proof. Let $\xi = f_0 - 1$ so that $\int \xi d\pi = 0$ and $\int \xi^3 d\pi = 0$. We may assume that $f_0 > 0$ and $f_0 \neq 0$ identically on M . Fix $p \in (1, 2]$ and consider the function

$$\psi(s) = \int (1 + s\xi)^p d\pi = \|1 + s\xi\|_p^p, \quad |s| \leq 1.$$

We have $\psi(0) = 1$, $\psi'(0) = 0$, and

$$\psi''(s) = p(p-1) \int \xi^2 (1 + s\xi)^{p-2} d\pi = p(p-1) \|\xi\|_2^2 \int (1 + s\xi)^{p-2} dv,$$

where v is a probability measure on M with density $\frac{\xi^2}{\|\xi\|_2^2}$ with respect to the measure π . Since the function $t \rightarrow t^{p-2}$ is convex in $t > 0$, and since $1 + s\xi > 0$ for all $|s| \leq 1$, we obtain by Jensen's inequality that

$$\int (1 + s\xi)^{p-2} dv \geq \left(\int (1 + s\xi) dv \right)^{p-2} = \left(\int (\xi^2 + s\xi^3) \frac{d\pi}{\|\xi\|_2^2} \right)^{p-2} = 1.$$

Hence, $\psi''(s) \geq p(p-1) \|\xi\|_2^2$ which implies $\psi(s) \geq 1 + \frac{p(p-1)}{2} \|\xi\|_2^2 s^2$. For $s = 1$, the latter yields

$$\|f_0 - 1\|_2^2 \leq \frac{2}{p(p-1)} (\|f_0\|_2^2 - 1), \quad 1 < p \leq 2,$$

with equality for $p = 2$. Letting $p \downarrow 1$, we obtain

$$\|f_0 - 1\|_2^2 \leq 2 \text{Ent}_\pi(f_0) = 2D(\mu_0 \| \pi),$$

an inequality sharpening the estimate (2.5) for the total variation norm. It remains to apply Theorem 2.2 and to bound the informational divergence similarly to Corollary 2.8. □

3. HIERARCHY OF INEQUALITIES AND SOME EXAMPLES

In this section we make a systematic study of relationship between various logarithmic Sobolev inequalities and the Poincaré inequality in discrete settings. Let (M, μ) be a probability space, and let \mathcal{A} be a linear space of bounded measurable functions on M . Further assumptions on \mathcal{A} are:

Axiom 1. If $f, g \in \mathcal{A}$, then $fg \in \mathcal{A}$ (that is, \mathcal{A} is an algebra).

Axiom 2. If $f \in \mathcal{A}$, then $e^f \in \mathcal{A}$.

Definition 3.1. Any bilinear form $\mathcal{E}: \mathcal{A} \times \mathcal{A} \rightarrow \mathbf{R}$ will be called a Dirichlet form.

Although the definition of Dirichlet forms has nothing to do with the measure μ , it appears that many standard examples are constructed through a measure. In what follows we will be primarily interested in a discrete setting of finite undirected graphs or finite Markov chains. However, since traditionally these functional inequalities have been studied in a continuous setting, we also briefly mention such a setting.

Example 3.2. (a continuous setting). Let M be an open subset of \mathbf{R}^d , and let \mathcal{A} be the family of all smooth, compactly supported functions on M . Put

$$\mathcal{E}(f, g) = \int_M \langle \nabla f(x), \nabla g(x) \rangle d\mu(x),$$

where $\langle \cdot, \cdot \rangle$ is a canonical scalar product in \mathbf{R}^d , and where $\nabla f(x) = (\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_d})$ denotes the usual gradient of f at the point $x \in M$. This gradient is local in the sense that $\nabla u(f) = u'(f) \nabla f$, for any smooth u such that $u(f) \in \mathcal{A}$.

The example can be generalized by considering for M a Riemannian manifold of dimension d . If M is compact, one typically takes for μ the normalized Lebesgue measure on M .

Example 3.3. (a graph setting). Let $G = (M, \mathcal{M})$ be a finite, connected, undirected graph with vertex set M and edge set $\mathcal{M} = \{(x, y) \in M \times M : x \sim y\}$. Let μ be an arbitrary probability measure on the vertices and let $\mu(x) = \mu(\{x\})$, $x \in M$. Given a function f on M , one can define the gradient $\nabla f(x)$ at each vertex $x \in M$ as the vector $\{f(x) - f(y)\}_{y \sim x}$ of the length $d(x)$, the degree of x . Hence, the corresponding Dirichlet form becomes

$$\begin{aligned} \mathcal{E}(f, g) &= \int \langle \nabla f(x), \nabla g(x) \rangle d\mu(x) \\ &= \sum_x \sum_{y \sim x} (f(x) - f(y))(g(x) - g(y)) \mu(x). \end{aligned}$$

Here \mathcal{A} represents the space of all functions on M .

Example 3.4. (an abstract discrete setting and reversible Markov kernels). Again, let (M, μ) be a finite probability space, and let $P: M \times M \rightarrow [0, +\infty)$ be a non-negative function, called a kernel in the sequel. For all

functions f, g on M , one may define the associated Dirichlet form by

$$\mathcal{E}(f, g) = \frac{1}{2} \int \sum_{y \in M} (f(x) - f(y))(g(x) - g(y))P(x, y) d\mu(x).$$

It corresponds to the gradient operator $\nabla f(x) = \{\frac{1}{\sqrt{2}}(f(x) - f(y))\sqrt{P(x, y)}\}_{y \in M}$, so again the gradient formula as in Example 3.2 works well. If P is a reversible Markov kernel, then the above definition is also equivalent to the general definition given by (1.1), and we have the additional property that $\mathcal{E}(f, g) = \mathcal{E}(g, f)$. It turns out that the formula suggested by Example 3.2 is particularly suited to study reversible kernels, due to the apparent symmetry, while (1.1) is more general.

In the rest of this section, whenever we assume that P is a Markov kernel, we also assume implicitly that P is in fact reversible.

Consider a probability space (M, μ) and a Dirichlet form $\mathcal{E}: \mathcal{A} \times \mathcal{A} \rightarrow \mathbf{R}$, as above. Then one can introduce Poincaré-type (or spectral gap) and logarithmic Sobolev inequalities as

$$\lambda_1 \text{Var}(f) \leq \mathcal{E}(f, f), \quad f \in \mathcal{A}, \tag{3.1}$$

$$\rho \text{Ent}(f^2) \leq 2\mathcal{E}(f, f), \quad f \in \mathcal{A}. \tag{3.2}$$

As mentioned already in the introduction, by a modified logarithmic Sobolev inequality (of the Dirichlet type), we mean an inequality of the form

$$\rho_0 \text{Ent}(e^f) \leq \frac{1}{2} \mathcal{E}(e^f, f), \quad f \in \mathcal{A}. \tag{3.3}$$

If the Dirichlet form comes through a gradient like in all the previous examples, one may also consider modified logarithmic Sobolev inequalities of the gradient type. The most popular versions which appeared in connection with the concentration of measure phenomenon (cf. Ref. 27) are:

$$\rho_1 \text{Ent}(e^f) \leq \frac{1}{2} \int |\nabla f|^2 e^f d\mu, \quad f \in \mathcal{A}, \tag{3.4}$$

$$\rho_1 \text{Ent}(e^f) \leq \frac{1}{2} \int |\nabla e^f|^2 e^{-f} d\mu, \quad f \in \mathcal{A}. \tag{3.5}$$

To be more precise, here one assumes that any point $x \in M$ is assigned with a linear operator $\mathcal{A} \ni f \rightarrow \nabla f(x) \in \mathbf{R}^{d(x)}$ such that the functions of the form $x \rightarrow \langle \nabla f(x), \nabla g(x) \rangle$ are μ -integrable and bounded, whenever $f, g \in \mathcal{A}$. Formally replacing f with $\log f$, the inequality (3.5) takes a more familiar form

$$\rho_2 \text{Ent}(f) \leq \frac{1}{2} \int \frac{|\nabla f|^2}{f} d\mu, \quad f \in \mathcal{A}, \quad f \text{ positive}. \tag{3.6}$$

More precisely, we obtain (3.5) from the last inequality (3.6) by applying it to e^f . At this step the axiom 2 is used. For the converse implication, one needs a different assumption, that $\log f \in A$ as long as f belongs to \mathcal{A} and is positive. Thus, in all the examples we considered before, the inequalities (3.5) and (3.6) are equivalent, but we prefer the first, exponential form in order to keep maximal generality and to save more analogs between (3.5) and the other exponential form (3.4).

If the gradient is local like in Example 3.2, all the log-Sobolev inequalities (3.2), (3.3), (3.4) and (3.5) are equivalent to each other, and moreover $\rho = \rho_0 = \rho_1 = \rho_2$ for optimal values. As for the general case, first we show that, under reasonable assumptions, the spectral gap inequality is weaker than any of these inequalities.

Proposition 3.5. Assume that

- (1) the function $g(x)=1$ belongs to \mathcal{A} and $\mathcal{E}(f, 1) = \mathcal{E}(1, f) = 0$, for all $f \in \mathcal{A}$;
- (2) for all $f, g \in \mathcal{A}$ and for any uniformly bounded sequence f_n converging to f (μ -almost everywhere), we have $\mathcal{E}(f_n, g) \rightarrow \mathcal{E}(f, g)$, as $n \rightarrow \infty$.

Then, for the optimal constants in (3.1)–(3.5), we have $\max\{\rho, \rho_0, \rho_1, \rho_2\} \leq \lambda_1$.

Proof. To show $\rho \leq \lambda_1$, note that, for every c real, $\mathcal{E}(f+c, f+c) = \mathcal{E}(f, f)$. Since $\text{Ent}((f+c)^2) \rightarrow 2\text{Var}(f)$, as $c \rightarrow \infty$, the application of (3.2) to functions of the form $f+c$ yields (3.1) with ρ in the place of λ_1 . To prove $\rho_0, \rho_1, \rho_2 \leq \lambda_1$, apply the inequalities (3.3)–(3.5) to functions $\frac{1}{n}f$ with $n \rightarrow \infty$. \square

Note that the assumptions (1) and (2) are not needed for deriving $\rho_1, \rho_2 \leq \lambda_1$. The assumption (2) is automatically fulfilled as long as there exists a linear operator L associated with the Dirichlet form \mathcal{E} . In particular, this is clearly true for an abstract discrete setting.

Now let us specialize the log-Sobolev inequalities to discrete settings where they may differ considerably in terms of the magnitudes of ρ, ρ_0, ρ_1 and ρ_2 .

Proposition 3.6. In the reversible Markov kernel setting, for the optimal constants in (3.1)–(3.5), we have

$$0 \leq \rho \leq \rho_0 \leq \rho_1 \leq \rho_2 \leq \lambda_1.$$

Proof. Let (M, μ) be a finite probability space with a reversible Markov kernel P . First we show that the logarithmic Sobolev inequality (3.2) implies the modified logarithmic Sobolev inequality (3.3) with $\rho_0 = \rho$. Thus, fix a function f on M . Starting from (3.2), apply it to the function $e^{f/2}$ to get

$$\rho \text{Ent}(e^f) \leq 2\mathcal{E}(e^{f/2}, e^{f/2}).$$

Hence, in order to derive (3.3) with the same constant on the left, it suffices to show that

$$\mathcal{E}(e^{f/2}, e^{f/2}) \leq \frac{1}{4}\mathcal{E}(e^f, f).$$

This estimate is actually observed in [15]. To remind the argument, note that, according to the definition (in Example 3.4) of the discrete Dirichlet form, we need to check that

$$\left(e^{f(x)/2} - e^{f(y)/2} \right)^2 \leq \frac{1}{4} \left(e^{f(x)} - e^{f(y)} \right) (f(x) - f(y)),$$

for all $x, y \in M$. Putting $a = e^{f(x)/2}$, $b = e^{f(y)/2}$, we are reduced to the inequality $(a - b)^2 \leq \frac{1}{2}(a^2 - b^2) \log \frac{a}{b}$ in the range $a, b > 0$, which can easily be verified to be true.

Now, in view of Proposition 3.5, we need only to show that (3.3) \Rightarrow (3.4) \Rightarrow (3.5) with $\rho_2 = \rho_1 = \rho_0$. Clearly, it suffices to compare the right hand sides in these inequalities and to see that, for every f on M ,

$$\mathcal{E}(e^f, f) \leq \int |\nabla f|^2 e^f d\mu \leq \int |\nabla e^f|^2 e^{-f} d\mu. \tag{3.7}$$

Since $|\nabla f(x)|^2 = \frac{1}{2} \sum_{y \in M} (f(x) - f(y))^2 P(x, y)$, we have

$$\begin{aligned} \int |\nabla f|^2 e^f d\mu &= \frac{1}{2} \sum_{x, y \in M} (f(x) - f(y))^2 e^{f(x)} P(x, y) \mu(x) \\ &= \frac{1}{2} \sum_{x, y \in M} (f(x) - f(y))^2 e^{f(y)} P(x, y) \mu(x), \end{aligned}$$

by reversibility. So

$$\int |\nabla f|^2 e^f d\mu = \frac{1}{2} \sum_{x, y \in M} (f(x) - f(y))^2 \frac{e^{f(x)} + e^{f(y)}}{2} P(x, y) \mu(x).$$

Similarly,

$$\int |\nabla e^f|^2 e^{-f} d\mu = \frac{1}{2} \sum_{x,y \in M} (e^{f(x)} - e^{f(y)})^2 \frac{e^{-f(x)} + e^{-f(y)}}{2} P(x,y) \mu(x).$$

On the other hand,

$$\mathcal{E}(e^f, f) = \frac{1}{2} \sum_{x,y \in M} (f(x) - f(y))(e^{f(x)} - e^{f(y)}) P(x,y) \mu(x).$$

To establish (3.7), it suffices to compare the corresponding terms in these three representations. Thus, put $a = f(x)$, $b = f(y)$ for fixed $x, y \in M$: we need to show that

$$(a-b)(e^a - e^b) \leq (a-b)^2 \frac{e^a + e^b}{2} \leq (e^a - e^b)^2 \frac{e^{-a} + e^{-b}}{2}.$$

Since all the three sides are symmetric with respect to (a,b) , we may assume $a \geq b$. Putting $a = b + h$, we are reduced to

$$h(e^h - 1) \leq h^2 \frac{e^h + 1}{2} \leq (e^h - 1)^2 \frac{e^{-h} + 1}{2}, \quad h \geq 0. \quad (3.8)$$

Write the first inequality as $e^h - 1 \leq h \frac{e^h + 1}{2}$. It turns into an equality at the point $h = 0$, while after differentiation it becomes $e^h \leq \frac{1}{2} + \frac{h+1}{2} e^h$. Again, there is equality at $h = 0$, and differentiating it, we arrive at $e^h \leq \frac{h+2}{2} e^h$ which is evidently true. This proves the first inequality in (3.8).

The second inequality is simplified as $h^2 e^h \leq (e^h - 1)^2 \iff h e^{h/2} \leq e^h - 1 \iff \frac{h}{2} \leq \text{sh}(\frac{h}{2})$. It readily holds, as well and thus Proposition 3.6 is proved.

Note that the normalizing property $\sum_y P(x,y) = 1$, $x \in M$, was not used in the proof of Proposition 3.6. Moreover, the proof holds good for the graph setting with μ being uniform on the set of vertices. M. Sammer and the second author also observed recently that a part of the above proposition, namely that $\rho \leq \rho_0 \leq \lambda_1$ easily extends to the nonreversible Markov setting as well. \square

Example 3.7. (symmetric discrete cube). Let $M = \{0, 1\}^n$ be the discrete cube. For $x \in M$, if y is the neighbour of x obtained by flipping coordinate i , then we write $y = s_i(x)$. The canonical Dirichlet form on M is defined by

$$\mathcal{E}(f, g) = \int \sum_{i=1}^n (f(x) - f(s_i(x)))(g(x) - g(s_i(x))) d\mu(x), \quad (3.9)$$

where the measure μ is uniform. In this case,

$$\rho = \rho_0 = \rho_1 = \rho_2 = \lambda_1 = 4. \tag{3.10}$$

Formally we are not in a Markov kernel setting. However, one may simply multiply the Dirichlet form by $\frac{1}{2n}$ to get the corresponding constants. That $\rho=4$ is due to Gross;⁽²²⁾ that $\lambda_1=4$ is immediate in “dimension” one ($n=1$). Both constants are dimension-free since the entropy, as well as, the variance represent subadditive functionals over product probability measures (see discussion in Section 4). Hence $\rho = \lambda_1$ and the remaining equalities in (3.10) follow immediately from Proposition 3.6, noting that this may be treated as a graph with uniform measure.

Example 3.8. (non-symmetric discrete cube). Now, for $p \in (0, 1)$, equip M with the product measure $\mu = \mu_p^n$ with marginal μ_p assigning mass p to 1 and mass $q = 1 - p$ to 0. In this case for the Dirichlet form (3.9)

$$\lambda_1 = \frac{1}{pq}, \quad \rho = \frac{1}{pq} \frac{2(p-q)}{\log p - \log q}, \quad \frac{1}{2pq} \leq \rho_0, \rho_2 \leq \frac{1}{pq}. \tag{3.11}$$

The first equality is trivial (again, up to the tensorization of the variance); the second one was obtained by Diaconis and Saloff-Coste in;⁽¹⁵⁾ see also⁽²³⁾ and.⁽³⁴⁾ Proofs of inequalities for ρ_0 and ρ_2 can be found in⁽²⁷⁾ and,⁽⁷⁾ respectively. As for the remaining constant, we have

$$\rho_1 \leq \frac{2(\log p - \log q)}{p - q}. \tag{3.12}$$

Note that in huge contrast with Proposition 3.6, as $pq \rightarrow 0$,

$$\rho_1 \ll \rho \ll \rho_0 \approx \rho_2 \approx \lambda_1,$$

(although the best value of ρ_1 is not known). This pathological situation concerns only the modified log-Sobolev inequality (3.4) of gradient type. Apparently, it may be explained with the fact that the gradient is not defined via the Dirichlet form (in contrast with (3.1), (3.2) and (3.3)) and essentially depends on the kernel itself. Indeed, already in dimension one, for any $f: \{0, 1\} \rightarrow \mathbf{R}$,

$$\int |\nabla f|^2 e^f d\mu = (f(1) - f(0))^2 \left(p e^{f(1)} + q e^{f(0)} \right). \tag{3.13}$$

If $p \neq q$, the right hand side is not invariant under replacement $f(1) \leftrightarrow f(0)$. On the other hand, in accordance with definition (3.9) in dimension one,

$$\mathcal{E}(e^f, f) = (f(1) - f(0)) \left(e^{f(1)} - e^{f(0)} \right), \quad (3.14)$$

which is invariant (and does not depend on p , at all).

To prove (3.12), we may and do restrict ourselves to the case $n = 1$. Using (3.13), write down the modified log-Sobolev inequality (3.4) on $M = \{0, 1\}$:

$$\begin{aligned} pf(1)e^{f(1)} + qf(0)e^{f(0)} - \left(pe^{f(1)} + qe^{f(0)} \right) \log \left(pe^{f(1)} + qe^{f(0)} \right) \\ \leq \frac{1}{2\rho_1} (f(1) - f(0))^2 \left(pe^{f(1)} + qe^{f(0)} \right). \end{aligned}$$

Applying this inequality to $f(1) = \log q - \log p = -f(0)$, we get $pe^{f(1)} + qe^{f(0)} = 1$, and the inequality becomes

$$(q - p)(\log q - \log p) \leq \frac{2}{\rho_1} (\log q - \log p)^2.$$

Example 3.9. (general Markov kernel on a two point set). Now, on $M = \{0, 1\}$ consider an arbitrary Markov kernel, that is, a stochastic matrix of the form

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \quad 0 \leq a, b \leq 1.$$

To avoid trivialities, we assume $a + b > 0$. Clearly, the kernel P has always an invariant measure μ which can be viewed as a probability vector on M with coordinates

$$p \equiv \mu_1 = \frac{a}{a+b}, \quad q \equiv \mu_0 = \frac{b}{a+b}.$$

Note, that P is reversible with respect to μ . Hence the associated Dirichlet form is given by

$$\mathcal{E}(f, g) = \frac{ab}{a+b} (f(0) - f(1))(g(0) - g(1)) = (a+b) \text{cov}_\mu(f, g),$$

where f, g are arbitrary functions on M . It follows from the above example (more precisely from (3.11)), and also using Proposition 3.6 that,

$$\frac{a+b}{2} \leq \rho_0 \leq \rho_1 \leq \rho_2 \leq \lambda_1 = a+b, \quad \rho = \frac{2(a-b)}{\log a - \log b}. \quad (3.15)$$

The left of the inequality can slightly be improved to

$$\rho_0 \geq \frac{a+b}{2} + \sqrt{ab},$$

with equality when $a = b$. We omit the derivation as an exercise to the interested reader.

Example 3.10. (the complete graph). A bit more generally, let (M, \mathcal{M}) be the complete graph on a non-empty finite set M . Moreover, assume M is equipped with a probability measure μ such that $\mu_* = \min_{x \in M} \mu(x) > 0$, and consider the function $P(x, y) = \mu(y)$. Then, (P, μ) is a reversible Markov kernel, and in accordance with the Markov kernel setting, the Dirichlet form may be expressed in terms of the covariance

$$\mathcal{E}(f, g) = \text{cov}_\mu(f, g) \equiv \int fg d\mu - \int f d\mu \int g d\mu. \tag{3.16}$$

In particular, for $M = \{0, 1\}$ with $\mu = \mu_p$, this Dirichlet form is pq times the Dirichlet form (3.14). Since the inequalities (3.1)–(3.3) are defined through the Dirichlet form, we can apply (3.11) (in part concerning ρ_0 and ρ) and then Proposition 3.6 to conclude that

$$\frac{1}{2} \leq \rho_0 \leq \rho_1 \leq \rho_2 \leq \lambda_1 = 1, \tag{3.17}$$

and that

$$\rho = \frac{2(p - q)}{\log p - \log q}. \tag{3.18}$$

Thus, in contrast with Example 3.8, the optimal constants in all modified log-Sobolev inequalities are of order λ_1 . Actually, the set of inequalities (3.17) remains to hold for an arbitrary complete graph M with the remark that, for a single point set M , all the optimal constants are equal to $+\infty$. Indeed, by Jensen’s inequality and by (3.16), $\text{Ent}(e^f) \leq \text{cov}(f, e^f) = \mathcal{E}(f, e^f)$, so $\rho_0 \geq \frac{1}{2}$. On the other hand, $\lambda_1 = 1$, and it remains to apply Proposition 3.6. As for the constant ρ , every complete graph M satisfies (3.18) with $p = \mu_*$, $q = 1 - \mu_*$. This is shown in Ref. 15 on the basis of the two point case (3.11).

As in the previous example, the bound $\rho_0 \geq \frac{1}{2}$ in (3.17) is not optimal and can slightly be improved to

$$\rho_0 \geq \frac{1}{2} + \sqrt{\mu_*(1 - \mu_*)}.$$

Example 3.11. (slices of the n -cube and Bernoulli-Laplace model). A fundamental example is a slice $\Omega(n, k)$ of the discrete cube: the graph with vertices being k -subsets of an n -set ($1 \leq k \leq n$, $n \geq 2$). Two subsets are adjacent if and only if they can be obtained from each other by a single swap of a pair of elements. (Note that this is also the so-called uniform matroid.) In particular, for $k=1$ or $k=n$, we obtain a complete graph of size n . By symmetry, $\Omega(n, k)$ and $\Omega(n, n-k)$ are isomorphic, so the range $k \leq \frac{n}{2}$ is only of interest.

There is a natural reversible Markov kernel associated with the graph $\Omega(n, k)$, which assigns the transition probability $P(x, y) = \frac{1}{k(n-k)}$, whenever x and y are neighbors. In this case, the spectral gap was studied by P. Diaconis and M. Shahshahani who showed in Ref. 16 that $\lambda_1 = \frac{n}{k(n-k)}$, cf. also Ref. 14. Equivalently, using the normalization of the graph setting as in Example 3.3, we have $\lambda_1 = 2n$ which is surprisingly independent of k . An asymptotic behavior of the logarithmic Sobolev constant ρ as a function of (k, n) was studied in Ref. 29 by Lee and Yau. Using a martingale approach, they prove that, in the graph setting,

$$\frac{c_1 n}{\log \frac{n}{k}} \leq \rho \leq \frac{c_2 n}{\log \frac{n}{k}}, \quad 1 \leq k \leq \frac{n}{2},$$

for some numerical constants $c_1, c_2 > 0$. A lower bound with $\log n$ replacing $\log \frac{n}{k}$ was previously obtained by Diaconis and Saloff-Coste⁽¹⁵⁾ where they also raised the question on the correct asymptotic.

As for modified log-Sobolev constants, all of them turn out to be of order λ_1 ! It will be shown in the next section by a direct inductive argument that

$$\frac{n+2}{2} \leq \rho_0 \leq 2n.$$

An asymptotically equivalent lower bound was also obtained by Gao and Quastel in a recent paper⁽²⁰⁾ with a different approach similar to the one of⁽²⁹⁾

Example 3.12. (random transpositions). Another related example is the group S_n of permutations of n elements. Each of $n!$ transpositions has $\frac{n(n-1)}{2}$ neighbours, thus inducing a canonical graph structure on S_n . In the corresponding Markov kernel setting, the spectral gap is $\lambda_1 = \frac{2}{n-1}$, cf. Ref. 17 (Diaconis mentions in Ref. 14 that the first 10 values were computed by J. Deken).

The logarithmic Sobolev constant ρ is approximately $\log(n)$ -times smaller.⁽²⁹⁾ Similarly to the previous example, ρ_0 turns out to be of order λ_1 , namely,

$$\frac{1}{2(n-1)} \leq \rho_0 \leq \frac{2}{n-1}.$$

We discuss the argument in the next section, cf. also Ref. 20. On the probabilistic language, the above implies in particular that, for of the chain on permutations using (uniform) random transpositions, the mixing time in the total variation norm is at most $O(n \log n)$, which is tight, whereas only an $O(n \log^2 n)$ bound follows from ρ , since $\rho = \Theta(1/(n \log n))$.

4. BETWEEN MODIFIED LOG-SOBOLEV AND POINCARÉ

For reversible kernels, both inequalities (2.1) and (2.2) can be united by a more general scheme under a certain stronger hypothesis. Namely, given (M, P, π) with P being a reversible kernel, for a number $p \in (1, 2]$, one may start with the Sobolev-type inequality

$$\alpha(p) [\|f\|_p^p - \|f\|_1^p] \leq \frac{p}{2} \mathcal{E}(f, f^{p-1}), \tag{4.1}$$

where f is an arbitrary positive function on M , and $\|f\|_p^p = \int f^p d\pi$.

If $p=2$, we are reduced to the Poincaré-type inequality (3.1), so the optimal constant $\alpha(2)$ is just the spectral gap λ_1 . For $1 < p < 2$, applying (4.1) to functions of the form $1 + \varepsilon f$ and letting $\varepsilon \rightarrow 0$, we obtain the relation

$$\alpha(p) \leq \lambda_1.$$

On the other hand, dividing both sides of (4.1) by $p-1$ and letting $p \downarrow 1$, we get in the limit the modified logarithmic Sobolev inequality (3.3), so $\alpha(1+) = \rho_0$.

The proofs of Theorems 2.2 and 2.4 are readily extended to the more general statement:

Theorem 4.1. Under the hypothesis (4.1) with $p \in (1, 2]$, for every initial distribution μ_0 on M ,

$$\|f_t\|_p^p - 1 \leq [\|f_0\|_p^p - 1] e^{-2\alpha(p)t}, \quad t \geq 0. \tag{4.2}$$

In the continuous setting with Dirichlet form $\mathcal{E}(f, g) = \int \langle \nabla f, \nabla g \rangle d\pi$ the inequality (4.1) may be rewritten equivalently by replacing p with $2/q$ and putting $f = g^q$. It then takes the form

$$\alpha(2/q) \left[\|g\|_2^2 - \|g\|_q^2 \right] \leq (2-q) \mathcal{E}(g, g), \quad 1 \leq q < 2. \tag{4.3}$$

This inequality was introduced in 1989 by Beckner⁽⁴⁾ as a kind of sharp interpolation between Poincaré and logarithmic Sobolev inequality: it was established for the canonical Gaussian measure with optimal constants $\alpha(2/q) = 1$ thus generalizing the famous Gross' theorem ((4.3) was also proved there for uniform distributions on Euclidean spheres). Recently, a similar inequality was derived for product measures in \mathbf{R}^n with marginal densities $c_r e^{-|x|^r}$, $1 \leq r \leq 2$, by Latala and Oleszkiewicz, cf. Ref. 25.

Let us note that, while for $q = 1$ the inequality (4.3) represents the spectral gap, the limiting case $q = 2$ reduces to the usual logarithmic Sobolev inequality (3.2), where the optimal constant may be much smaller than the one in (3.3). Therefore, (4.1) has a correct form to fit the features of the modified log-Sobolev inequality in the discrete setting. The essential difference between (4.1) and (4.3) already appears for complete graphs as we can see from the following:

Proposition 4.2. For every complete graph M on at least two vertices, equipped with an arbitrary probability measure π , for every $p \in (1, 2]$,

$$\frac{p}{2} \leq \alpha(p) \leq 1.$$

Proof. The right hand side inequality is immediate since $\lambda_1 = 1$. Recalling that $\mathcal{E}(f, g) = \text{cov}_\pi(f, g)$, the left hand side inequality is just

$$\|f\|_p^p - \|f\|_1^p \leq \text{cov}_\pi(f, f^{p-1}) = \|f\|_p^p - \|f\|_1 \|f\|_{p-1}^{p-1},$$

that is, $\|f\|_{p-1} \leq \|f\|_1$. The latter holds due to $p - 1 \leq 1$. □

It can easily be shown that the lower bound $\alpha \geq \frac{p}{2}$ is sharp in the class of all complete graphs. This is true even if we fix M and let π be arbitrary. However, the bound can be sharpened a little for prescribed measures. In particular, we have:

Proposition 4.3. For a two point complete graph M with uniform probability measure π , for every $p \in (1, 2]$,

$$\alpha(p) = \lambda_1 = 1.$$

Proof. Letting $M = \{0, 1\}$, $f(1) = a$, $f(0) = b$, the inequality (4.1) turns into

$$\frac{a^p + b^p}{2} - \left(\frac{a+b}{2}\right)^p \leq \frac{p}{8}(a-b)(a^{p-1} - b^{p-1}), \quad a, b > 0.$$

By homogeneity, we may assume that $a + b = 1$, and by symmetry, that $a \geq b$. Thus, setting $a = 1 + h$, $b = 1 - h$, we are reduced to

$$\frac{(1+h)^p + (1-h)^p}{2} - 1 \leq \frac{p}{4} h \left((1+h)^{p-1} - (1-h)^{p-1} \right),$$

under the assumption $0 \leq h \leq 1$. The best way to derive this inequality is apparently to write Taylor's expansion in powers of h . The left hand side is just

$$p(p-1)\frac{h^2}{2!} + p(p-1)(p-2)(p-3)\frac{h^4}{4!} + p(p-1)(p-2)(p-3)(p-4)(p-5)\frac{h^6}{6!} + \dots$$

while the right hand side is given by

$$\frac{ph}{2} \left[(p-1)h + (p-1)(p-2)(p-3)\frac{h^3}{3!} + (p-1)(p-2)(p-3)(p-4)(p-5)\frac{h^5}{5!} + \dots \right].$$

Since $1 < p \leq 2$, all the coefficients in front of h^{2n} are non-negative, so it remains to apply a simple bound $\frac{1}{2n!} \leq \frac{1}{2(2n-1)!}$, $n \geq 1$. Proposition 4.3 follows. \square

One may further enlarge the class of examples by considering (Cartesian) products of graphs $M = M_1 \times \dots \times M_d$ with product measures $\pi = \pi_1 \otimes \dots \otimes \pi_d$, among which the discrete cube seems to be of the most interest. This can be done with the help of the crucial property of the functional

$$\mathcal{L}(f) = \|f\|_p^p - \|f\|_1^p,$$

being of the subadditive type. This means that, for any f on M ,

$$\mathcal{L}(f) \leq \int \sum_{i=1}^d \mathcal{L}_{x_i}(f) d\pi(x), \tag{4.4}$$

where $\mathcal{L}_{x_i}(f)$ denotes an application of \mathcal{L} to the i -th coordinate with respect to the marginal measure π_i (while the remaining coordinates are fixed). In case of the entropy functional $\mathcal{L}(f) = \text{Ent}(f)$, a similar property was established by Lieb.⁽³⁰⁾ More general functionals on abstract probability spaces (M, π) of the form

$$\mathcal{L}(f) = \int \Phi(f) d\pi - \Phi \left(\int f d\pi \right)$$

have been studied only recently. If Φ is a convex function on an interval (a, b) finite or not (a necessary assumption), the property (4.4) always holds true for product spaces for all measurable f with values in (a, b) if

and only if the functional \mathcal{L} is convex in f . We refer the reader to⁽²⁶⁾ for a simple inductive proof of this characterization. Note that the entropic case corresponds to the choice $\Phi(t) = t \log t$, $t > 0$, and the convexity of \mathcal{L} readily follows from the classical representation

$$\text{Ent}_\pi(f) = \sup_{\int e^g d\pi \leq 1} \int f g d\pi.$$

As for the general convex Φ , the convexity of \mathcal{L} can easily be shown to be equivalent to the property that the function

$$\frac{1}{\Phi''(t)}, \quad a < t < b,$$

is concave (provided that Φ is in $C^2(a, b)$ with $\Phi'' > 0$). Thus, the power functions $\Phi(t) = t^p$ generate convex functionals \mathcal{L} if and only if $1 \leq p \leq 2$. We may apply this to the discrete cube.

Proposition 4.4. Let $M = \{0, 1\}^d$ with uniform probability measure π , and let $p \in (1, 2]$. Then, in the Markov kernel setting,

$$\alpha(p) = \lambda_1 = \frac{2}{d}.$$

Perhaps, a similar statement for the discrete cube can be done for the inequality (4.3): Beckner⁽⁴⁾ mentions on the possibility to derive (4.3) for the Gaussian measure by starting from the discrete cube in the spirit of Gross⁽²²⁾ approach to logarithmic Sobolev inequality (thus, Beckner implicitly refers to the convexity of the functional \mathcal{L} generated by $\Phi(t) = t^p$).

Example 4.5. (product graphs) From the above discussion, we also note the following generalization. Let $M = G^n$ be the Cartesian product of n copies of G , with product probability measure μ^n , where μ is arbitrary on the vertices of G , and let $p \in (1, 2]$. Then we have $\alpha(p)[G^n] = \alpha(p)[G]$, for all $n \geq 1$.

The following observation is aimed at the study of the interpolating inequality (4.1) for non-product graphs.

Lemma 4.6. For any $p \in (1, 2]$, the function

$$R(a, b) = (a - b)(a^{p-1} - b^{p-1}), \quad a, b > 0,$$

is convex in the positive quadrant.

Proof. Let $t = p - 1$, $0 < t \leq 1$. Then, clearly the second partial derivatives

$$\begin{aligned} \frac{\partial^2 R}{\partial a^2} &= ta^{t-2}((1+t)a + (1-t)b), \\ \frac{\partial^2 R}{\partial b^2} &= tb^{t-2}((1+t)b + (1-t)a), \end{aligned}$$

are positive. Since $\frac{\partial^2 R}{\partial a \partial b} = -t(a^{t-1} + b^{t-1})$, after algebraic simplifications we come to

$$\frac{1}{t^2} \det(\text{Hess}(R)) = (1-t^2)a^{t-2}b^{t-2}(a-b)^2 - (a^{t-1} - b^{t-1})^2.$$

We need to show that the right hand side is non-negative. Replacing $a = ub$, we arrive at

$$\sqrt{1-t^2}u^{\frac{t}{2}-1}|u-1| \geq |u^{t-1}-1|, \quad u > 0. \tag{4.5}$$

First consider the case $0 < u \leq 1$, when the inequality simplifies as

$$\psi(u) \equiv \sqrt{1-t^2}u^{\frac{t}{2}-1}(1-u) - (u^{t-1}-1) \geq 0.$$

Since $\psi(1) = 0$, it suffices to see that ψ is non-increasing on $(0, 1)$, that is,

$$\psi'(u) = \sqrt{1-t^2} \left(\left(\frac{t}{2} - 1 \right) u^{\frac{t}{2}-2} - \frac{t}{2} u^{\frac{t}{2}-1} \right) + (1-t)u^{t-2} \leq 0.$$

Dividing by $u^{\frac{t}{2}-2}$, we have to show that

$$g(u) \equiv (1-t)u^{\frac{t}{2}} - \sqrt{1-t^2} \left(\left(1 - \frac{t}{2} \right) + \frac{t}{2}u \right) \leq 0.$$

Clearly, $g(0) < 0$ and $g(1) \leq 0$, so we only need to check the above inequality at the point u_0 where $g'(u_0) = 0$ (if it exists). Differentiating, we get $u_0^{\frac{t}{2}-1} = \frac{\sqrt{1-t^2}}{1-t}$, so

$$\begin{aligned} g(u_0) &= (1-t) \frac{\sqrt{1-t^2}}{1-t} u_0 - \sqrt{1-t^2} \left(\left(1 - \frac{t}{2} \right) + \frac{t}{2}u_0 \right) \\ &= \sqrt{1-t^2} \left(1 - \frac{t}{2} \right) (u_0 - 1), \end{aligned}$$

which is indeed negative.

Similarly, the case $u \geq 1$ can be settled, for which (4.5) becomes

$$\psi(u) \equiv \sqrt{1-t^2}u^{\frac{t}{2}-1}(u-1) - (1-u^{t-1}) \geq 0.$$

Now, since $\psi(1)=0$, we need to see that ψ is non-decreasing on $(1, +\infty)$, that is,

$$\psi'(u) = \sqrt{1-t^2} \left(\frac{t}{2} u^{\frac{t}{2}-1} - \left(\frac{t}{2} - 1 \right) u^{\frac{t}{2}-2} \right) - (1-t)u^{t-2} \geq 0.$$

Dividing by $u^{\frac{t}{2}-2}$, we are reduced to

$$g(u) \equiv \sqrt{1-t^2} \left(\frac{t}{2} u + \left(1 - \frac{t}{2} \right) \right) - (1-t)u^{\frac{t}{2}} \geq 0.$$

We have $g(1) \geq 0$ and $g(+\infty) = +\infty$, so we only need to check the above inequality at the point $u_0 > 0$ with $g'(u_0) = 0$ in case such a point exists. But differentiating, we get $u_0^{\frac{t}{2}-1} = \frac{\sqrt{1-t^2}}{1-t}$ which is impossible since the left hand side of the equality is smaller than 1, while the right hand side is bigger than 1. Hence, g is strictly increasing, finishing the proof of Lemma 4.6. \square

As an illustration, consider the graph $M = \Omega(n, k)$ of slices of the discrete cube. Recall that the statement about the modified log-Sobolev inequality,

$$\text{Ent}_\pi(f) \leq \frac{1}{n+2} \mathcal{E}(f, \log f), \quad (4.6)$$

was mentioned in Example 3.10 for the Dirichlet form corresponding to the graph setting, namely,

$$\mathcal{E}(f, g) = \int \sum_{y \sim x} (f(x) - f(y))(g(x) - g(y)) d\pi(x), \quad (4.7)$$

where π is uniform probability measure on M . Making use of Proposition 4.2 and the convexity of $R(a, b)$ above, (Lemma 4.6) leads to the following generalization of (4.6).

Proposition 4.7. Let $1 \leq k \leq n-1$ be integer, and $p \in (1, 2]$. For every positive function f on $\Omega(n, k)$, with respect to the uniform probability measure

$$\|f\|_p^p - \|f\|_1^p \leq \frac{1}{n+2} \mathcal{E}(f, f^{p-1}). \quad (4.8)$$

Equivalently, in the Markov kernel setting,

$$\frac{p(n+2)}{4k(n-k)} \leq \alpha(p) \leq \lambda_1 = \frac{n}{k(n-k)}.$$

The constant on the right, $\frac{1}{n+2}$, is of correct order uniformly over all admissible triples (n, k, p) . The particular case $p=2$ in (4.8) yields the spectral gap inequality

$$\text{Var}_\mu(f) \leq \frac{1}{n+2} \mathcal{E}(f, f). \tag{4.9}$$

As we already mentioned, the optimal value of the constant is equal to $\frac{1}{2n}$ (Ref. 16). For this constant, equality in (4.9) is attained for any linear function f on \mathbf{R}^n (provided that the graph is naturally embedded in \mathbf{R}^n as in the proof below). On the other hand, dividing both sides of (4.8) by $p-1$ and letting $p \rightarrow 1$, we arrive at the modified logarithmic Sobolev inequality (4.6) stated above.

Proof of Proposition 4.7. We may identify $\Omega = \Omega(n, k)$ with a slice of the discrete cube, $\{x \in \{0, 1\}^n : x_1 + x_2 + \dots + x_n = k\}$, so that Ω inherits the structure of a graph from the discrete cube: neighbours are the pairs of points which differ exactly in two coordinates. The canonical inner metric $\rho = \rho_{n,k}$ in Ω is given by

$$\rho(x, y) = \frac{1}{2} \text{card}\{i \leq n : x_i \neq y_i\}, \quad x, y \in \Omega,$$

that is, one half of the Hamming distance.

For $1 \leq k \leq n-1$, let $A_{n,k}$ denote the best constant in

$$\|f\|_p^p - \|f\|_1^p \leq A_{n,k} \mathcal{E}(f, f^{p-1}), \tag{4.10}$$

where f is an arbitrary positive function on Ω . In terms of the function R of Lemma 4.6 this inequality takes the form

$$\int f^p d\mu \leq \left(\int f d\mu \right)^p + A_{n,k} \frac{1}{C_n^k} \sum_{\rho(x,y)=1} R(f(x), f(y)), \tag{4.11}$$

where C_n^k denotes $\binom{n}{k}$ and the summation is performed over all ordered pairs $(x, y) \in \Omega \times \Omega$ such that $\rho(x, y) = 1$. By symmetry, $A_{n,k} = A_{n,n-k}$.

We know that $A_{n,1} \leq \frac{1}{2n}$. As for $k \geq 2$, we will deduce a recursive inequality relating $A_{n,k}$ to $A_{n-1,k-1}$, and then we may proceed by induction. Thus, fix $k \geq 2$ and a positive function f on Ω with $\int f d\mu = 1$ (this can be assumed in view of homogeneity of (4.10)–(4.11)). Introduce the subgraphs

$$\Omega_i = \{x \in \Omega : x_i = 1\}, \quad 1 \leq i \leq n,$$

and equip them with uniform probability measures μ_i . Since all Ω_i can be identified with $\Omega(n-1, k-1)$, we may write the definition (4.11) for these graphs:

$$\int_{\Omega_i} f^p d\mu_i \leq \left(\int_{\Omega_i} f d\mu_i \right)^p + \frac{A_{n-1,k-1}}{C_{n-1}^{k-1}} \sum_{x \in \Omega_i} \sum_{y \in \Omega_i, \rho(x,y)=1} R(f(x), f(y)).$$

Setting $\varphi(i) = \int_{\Omega_i} f d\mu_i$ and summing these inequalities over all $i \leq n$ with weight $\frac{1}{n}$, we obtain

$$\frac{1}{n} \sum_{i=1}^n \int_{\Omega_i} f^p d\mu_i \leq \frac{1}{n} \sum_{i=1}^n \varphi(i)^p + \frac{A_{n-1,k-1}}{nC_{n-1}^{k-1}} \sum_{i=1}^n \sum_{x \in \Omega_i} \sum_{y \in \Omega_i, \rho(x,y)=1} R(f(x), f(y)). \tag{4.12}$$

Since $\frac{1}{n} \sum_{i=1}^n \mu_i = \mu$, the first term in (4.12) is equal to $\int f^p d\mu$. The second term is estimated from above, according to the case $k=1$, by

$$\left(\frac{1}{n} \sum_{i=1}^n \varphi(i) \right)^p + \frac{A_{n,1}}{C_n^1} \sum_{i \neq j} R(\varphi(i), \varphi(j)).$$

But $\frac{1}{n} \sum_{i=1}^n \varphi(i) = \int f d\mu = 1$. Hence, (4.12) implies

$$\int f^p d\mu - 1 \leq \frac{A_{n,1}}{n} \sum_{i \neq j} R(\varphi(i), \varphi(j)) + \frac{A_{n-1,k-1}}{nC_{n-1}^{k-1}} \sum_{i=1}^n \sum_{x \in \Omega_i} \sum_{y \in \Omega_i, \rho(x,y)=1} R(f(x), f(y)).$$

To treat the triple sum, fix $x, y \in \Omega$ with $\rho(x, y) = 1$. The number of all i such that $x \in \Omega_i$ and $y \in \Omega_i$ simultaneously is equal to $k-1$. Hence, the triple sum will contribute

$$(k-1) \sum_{x \in \Omega} \sum_{y \in \Omega, \rho(x,y)=1} R(f(x), f(y)) = (k-1)C_n^k \mathcal{E}(f, \log f).$$

Since $\frac{(k-1)C_n^k}{nC_{n-1}^{k-1}} = \frac{k-1}{k}$, we thus get

$$\int f^p d\mu - 1 \leq \frac{A_{n,1}}{n} \sum_{i \neq j} R(\varphi(i), \varphi(j)) + \frac{(k-1)A_{n-1,k-1}}{k} \mathcal{E}(f, \log f). \tag{4.13}$$

To treat the sum in (4.13), for each pair $(i, j), i \neq j$, define the bijective map $s_{ij}: \{0, 1\}^n \rightarrow \{0, 1\}^n$,

$$(s_{ij}x)_r = x_r, \quad \text{for } r \neq i, j, \quad \text{and} \quad (s_{ij}x)_j = x_i, \quad (s_{ij}x)_i = x_j.$$

It acts as a bijection between Ω_i and Ω_j and pushes forward μ_i onto μ_j , provided that $k \geq 2$. In particular, $\varphi(j) \equiv \int f(y) d\mu_j(y) = \int f(s_{ij}x) d\mu_i(x)$.

Now, by Lemma 4.6, the function R is convex in the quadrant $a, b > 0$. Consequently, by Jensen's inequality,

$$R(\varphi(i), \varphi(j)) = R\left(\int f(x)d\mu_i(x), \int f(s_{ij}x)d\mu_i(x)\right) \leq \int R(f(x), f(s_{ij}x))d\mu_i(x).$$

Therefore,

$$\sum_{i \neq j} R(\varphi(i), \varphi(j)) \leq \frac{1}{C_{n-1}^{k-1}} \sum_{i \neq j} \sum_{x \in \Omega_i} R(f(x), f(s_{ij}x)). \tag{4.14}$$

Note that $y = s_{ij}x$ always implies $\rho(x, y) \leq 1$, and in case $x \in \Omega_i$, the equality $\rho(x, y) = 1$ is only possible when $x_i = 1, x_j = 0$. Hence, the double sum in (4.14) contains only terms $R(f(x), f(y))$ with $\rho(x, y) = 1$ (the cases $\rho(x, y) = 0$ can be excluded). In turn, fixing any pair $(x, y) \in \Omega$ such that $\rho(x, y) = 1$, there is a unique pair (i, j) such that $i \neq j$ and $y = s_{ij}x$. Thus, the right hand side of (4.14) is just

$$\frac{1}{C_{n-1}^{k-1}} \sum_{x \in \Omega} \sum_{y \in \Omega, \rho(x,y)=1} R(f(x), f(y)) = \frac{n}{k} \mathcal{E}(f, f^{p-1}),$$

and we get from (4.13)

$$\int f^p d\mu - 1 \leq \frac{A_{n,1} + (k-1)A_{n-1,k-1}}{k} \mathcal{E}(f, f^{p-1}).$$

Hence, $A_{n,k} \leq \frac{1}{k}(A_{n,1} + (k-1)A_{n-1,k-1})$, or in terms of $B_{n,k} = kA_{n,k}$,

$$B_{n,k} \leq A_{n,1} + B_{n-1,k-1}.$$

Applying this inequality successively $k - 1$ times and recalling that $A_{r,1} \leq \frac{1}{2^r}$, we arrive at

$$\begin{aligned} B_{n,k} &\leq \frac{1}{2n} + \frac{1}{2n-1} + \dots + \frac{1}{2(n-(k-2))} + B_{n-(k-1),1} \\ &\leq \frac{1}{2n} + \frac{1}{2(n-1)} + \dots + \frac{1}{2(n-(k-2))} + \frac{1}{2(n-(k-1))}. \end{aligned}$$

If $k \leq \frac{n}{2}$, each of the above k terms does not exceed $\frac{1}{n+2}$, so $B_{n,k} \leq \frac{k}{n+2}$. This implies the desired estimate $A_{n,k} \leq \frac{1}{n+2}$. In the case $k \geq \frac{n}{2}$, we may use $A_{n,k} = A_{n,n-k}$, and Proposition 4.7 follows. \square

A similar statement with a completely similar proof can be made about the symmetric group $M = S_n$ in which case we have (4.6) with the same constant $\frac{1}{n+2}$ for the Dirichlet form (4.7):

Proposition 4.8. Let $p \in (1, 2]$. For every positive function f on S_n , $n \geq 2$, with respect to the uniform probability measure

$$\|f\|_p^p - \|f\|_1^p \leq \frac{1}{n+2} \mathcal{E}(f, f^{p-1}).$$

Equivalently, in the Markov kernel setting,

$$\frac{p(n+2)}{2n(n-1)} \leq \alpha(p) \leq \lambda_1 = \frac{2}{n-1}.$$

Now let us return to Theorem 4.1 and the inequality (4.2) about the mixing time. Since the norm $\|f_0\|_p$ is maximized for Dirac measure $\mu_0 = \delta_x$, for some $x \in M$, we obtain similarly to (2.4) a more general bound

$$\|f_t\|_p^p - 1 \leq \frac{1 - \pi_*^{p-1}}{\pi_*^{p-1}} e^{-2\alpha(p)t}, \quad t \geq 0,$$

where $\pi_* = \min_x \pi(x)$. Letting $p \downarrow 1$ helps us recover the previous estimate on the informational divergence, cf. (2.2) and (2.6),

$$\text{Ent}_\pi(f_t) \leq \log \frac{1}{\pi_*} e^{-2\rho_0 t}, \quad t \geq 0.$$

5. CONCENTRATION AND BOUNDS ON DIAMETER

Throughout this section we assume that $G = (M, \mathcal{M})$ is a finite, connected, undirected graph of cardinality $|G| = \text{card}(M) \geq 2$. For simplicity we assume that G is d -regular, although typically the weaker assumption, that the maximum degree is at most d , is sufficient. We denote by $D = D(G)$ the diameter of G . As usual, by a Lipschitz function on G we mean a function $f: M \rightarrow \mathbf{R}$ such that $|f(x) - f(y)| \leq 1$, whenever x and y are neighboring vertices. Let μ be uniform probability over M so that $\mu(x) \equiv \mu(\{x\}) = \frac{1}{|G|}$.

For some $\rho_0 > 0$, the graph G satisfies the logarithmic Sobolev inequality

$$\rho_0 \text{Ent}_\mu(e^f) \leq \frac{1}{2} \mathcal{E}(f, e^f), \tag{5.1}$$

for all $f: M \rightarrow \mathbf{R}$, with a canonical Dirichlet form in the graph setting

$$\mathcal{E}(f, g) = \sum_{x \in M} \sum_{y \in M: y \sim x} (f(x) - f(y))(g(x) - g(y))\mu(x).$$

Since the constant ρ_0 might be of considerable interest in the study of Markov chains, one may also wonder how it influences the global behavior of Lipschitz functions and how it is related to basic characteristics such as the diameter of a graph and the degree of its vertices. As it turns out, deviations of Lipschitz functions from their μ -means may be controlled by the tails of the so-called double exponential distribution function $F(h) = e^{-e^{-h}}$ on the negative half-axis (with tails that decrease to zero much faster than the Gaussian ones, for example). This is already seen from the bound on the Laplace transform of Lipschitz functions.

Theorem 5.1. Let f be Lipschitz with $\mathbf{E}_\mu f = 0$. Let $A = d/\rho_0$. Then,

$$\begin{aligned} \mathbf{E}_\mu e^{tf} &\leq e^{At^2}, & 0 \leq t \leq 1 \\ &\leq e^{At(\log t + 1)}, & t \geq 1. \end{aligned}$$

Proof. The argument is standard. By (5.1), for all f on M , we have

$$\begin{aligned} \rho_0 \text{Ent}_\mu(e^f) &\leq \frac{1}{2|G|} \sum_{x \in M} \sum_{y \in M: y \sim x} (f(x) - f(y))(e^{f(x)} - e^{f(y)}) \\ &= \frac{1}{|G|} \sum_{x \in M} \sum_{\substack{y \in M: y \sim x \\ f(x) > f(y)}} (f(x) - f(y))(e^{f(x)} - e^{f(y)}) \\ &= \frac{1}{|G|} \sum_{x \in M} \sum_{\substack{y \in M: y \sim x \\ f(x) > f(y)}} (f(x) - f(y))(1 - e^{-(f(x) - f(y))})e^{f(x)} \\ &= \frac{1}{|G|} \sum_{x \in M} \sum_{y \in M: y \sim x} \phi((f(x) - f(y))^+)e^{f(x)}, \end{aligned}$$

where we set $\phi(t) = t(1 - e^{-t})$. In particular, if f is Lipschitz, then for all $t \geq 0$,

$$\rho_0 \text{Ent}_\mu(e^{tf}) \leq d\phi(t)\mathbf{E}_\mu e^{tf}. \tag{5.2}$$

Assume $\mathbf{E}_\mu f = 0$ and let $\mathbf{E}_\mu e^{tf} = e^{tu(t)}$. Then $\text{Ent}_\mu(e^{tf}) = t^2 u'(t)e^{tu(t)}$, so (5.2) becomes

$$u'(t) \leq \frac{d}{\rho_0} \frac{\phi(t)}{t^2}, \quad t \geq 0.$$

Integrating this inequality and using $u(0) = 0$ (due to $\mathbf{E}_\mu f = 0$), we conclude that

$$u(t) \leq \frac{d}{\rho_0} \int_0^t \frac{\phi(s)}{s^2} ds = \frac{d}{\rho_0} \int_0^t \frac{1 - e^{-s}}{s} ds.$$

But $\phi(t) = t(1 - e^{-t}) \leq \min(t, t^2)$ implies, setting $A = \frac{d}{\rho_0}$,

$$\begin{aligned} u(t) &\leq At, & 0 \leq t \leq 1 \\ &\leq A + A \log t, & t \geq 1, \end{aligned}$$

proving the theorem. \square

Corollary 5.2. Given Lipschitz f on M with $\mathbf{E}_\mu f = 0$, for any $h \geq 2A = 2d/\rho_0$,

$$\mu\{f \geq h\} \leq \exp\left\{-\frac{A}{e^2} e^{\frac{h}{A}}\right\}.$$

Proof. Using the standard Chebyshev argument, the theorem implies that for all $t \geq 1$, $h > 0$,

$$\mu\{f \geq h\} \leq e^{-th} \mathbf{E}_\mu e^{tf} \leq e^{At(\log t + 1) - th}.$$

The corollary now follows by minimizing the exponent over all $t \geq 1$. \square

Similarly, the use of the first inequality of Theorem 5.1 allows one to control “small” deviations of f above levels h from the remaining range $0 \leq h \leq 2A$. Namely, in this case we have a standard gaussian bound:

$$\mu\{f \geq h\} \leq \exp\left\{-\frac{h^2}{4A}\right\}, \quad A = \frac{d}{\rho_0}.$$

Together with Corollary 5.2 this yields:

Corollary 5.3. The diameter $D = D(G)$ of a d -regular graph G satisfies,

$$D \leq \begin{cases} 4\sqrt{A \log |G|}, & \text{if } |G| \leq e^A, \\ 4A + 2A \log \frac{\log |G|}{A}, & \text{if } |G| \geq e^A, \end{cases}$$

where $A = d/\rho_0$.

Proof. Given a Lipschitz function f on G such that $\mathbf{E}_\mu f = 0$, let $h_1 = \max_{x \in M} f(x)$. Since the set $\{f \geq h_1\}$ contains at least one vertex, by Corollary 5.2 and the comment following it,

$$\frac{1}{|G|} \leq \mu\{f \geq h_1\} \leq e^{-v(h_1)},$$

where

$$v(h) = \begin{cases} h^2/(4A), & \text{if } 0 \leq h \leq 2A, \\ A \exp\{h/A - 2\}, & \text{if } h \geq 2A. \end{cases}$$

Hence, introducing the inverse function

$$v^{-1}(z) = \begin{cases} 2\sqrt{Az}, & \text{if } 0 \leq z \leq A, \\ A(2 + \log(z/A)), & \text{if } z \geq A, \end{cases}$$

we obtain that $h_1 \leq v^{-1}(\log |G|)$. For the function $-f$, the latter inequality takes the form $-h_0 \leq v^{-1}(\log |G|)$, where $h_0 = \min_{x \in M} f(x)$. Adding up the two estimates, we arrive at

$$\max_{x,y \in M} (f(x) - f(y)) \leq 2v^{-1}(\log |G|).$$

This inequality is translation invariant, so the condition on the mean may be removed. The corollary now follows by choosing the Lipschitz function $f(x) = d(x, x_0)$, $x_0 \in M$, and maximizing over all x_0 . \square

Now, instead of (5.1), let us see what can be derived if we start from the logarithmic Sobolev inequality.

$$\rho_1 \text{Ent}_\mu(e^f) \leq \frac{1}{2} \int |\nabla f|^2 e^f d\mu. \tag{5.3}$$

Recall that $|\nabla f(x)|^2 = \sum_{y \in M: y \sim x} (f(x) - f(y))^2$, $x \in M$.

Proposition 5.4. $D \leq 2\sqrt{\frac{2d}{\rho_1} \log |G|}$.

Remark 5.5. This improves upon similar results by Alon and Milman⁽¹⁾ and F. R. K. Chung, A. Grigoryan, and S.-T. Yau [11], where the bounds are of the type $D \leq c\sqrt{\frac{d}{\lambda_1} \log |G|}$, for some universal constant $c > 0$. Using the elementary inequality $\rho \geq \frac{\lambda_1}{\log |G|}$ (see e.g. Ref. 34) and the relation $\rho_1 \geq \rho$, as we commented after the proof of Proposition 3.6), it is clear that the Proposition 5.4 is a refinement. Results in⁽¹⁰⁾ and⁽³²⁾ also provide improvements over,⁽¹⁾ but are in general incomparable with ours.

Remark 5.6. The proposition also implies that, for expander graphs of bounded degree, ρ , ρ_0 and ρ_1 are all of the order of $1/\log |G|$, where

the constants would depend on the bounds on the degree and the expansion, or equivalently, the spectral gap. Indeed since we have,

$$\frac{\lambda_1}{\log |G|} \leq \rho \leq \rho_0 \leq \rho_1 \leq \frac{8d \log |G|}{D^2},$$

and since for graphs with degree at most d , the diameter is at least $\log_d |G|$, up to a universal constant.

Remark 5.7. Starting with the standard log-Sobolev inequality, with $\rho > 0$,

$$\rho \text{Ent}_\mu f^2 \leq 2\mathcal{E}(f, f),$$

and using an argument similar to the above, it can be shown that $D \leq 5d/\rho$. It remains to be seen how this compares with Corollary 5.3 and Proposition 5.4 above. The example of the discrete cube, with $M = \{0, 1\}^d$, which is d -regular with diameter d and $\rho = \rho_1 = 4$ shows that our bounds on the diameter are tight up to universal constants.

Proof of Proposition 5.4. We follow an argument of M. Ledoux for deriving from (5.3) a bound on the Laplace transform, cf. e.g. [26] or [27]. Applying the inequality (5.3) to tf , with $t \in \mathbf{R}$ and Lipschitz f such that $\mathbf{E}_\mu f = 0$, we get

$$\rho_1 \text{Ent}_\mu(e^{tf}) \leq \frac{dt^2}{2} \mathbf{E}_\mu e^{tf}. \tag{5.4}$$

Setting $\mathbf{E}_\mu e^{tf} = e^{tu(t)}$, $\text{Ent}_\mu(e^{tf})$ becomes $t^2 u'(t) e^{tu(t)}$. Plugging into (5.4) yields that $u'(t) \leq d/(2\rho_1)$, which in turn implies that $u(t) \leq dt/(2\rho_1)$. Hence

$$\mathbf{E}_\mu e^{tf} \leq e^{dt^2/2\rho_1}. \tag{5.5}$$

Now let us tensorize (5.5) on $G \square G$, the Cartesian product of G with itself. It is well known (cf. e.g. Ref. 24) that making use of the additive property of entropy, one has $\rho_1(G \square G) = \rho_1(G)$. Consider g on $M \times M$ of the form $g(x, y) = f(x) - f(y)$. In particular, if f is Lipschitz on G , then g will be Lipschitz on $G \square G$, and moreover, $\mathbf{E}_{\mu \times \mu} g = 0$. Thus applying (5.5) with g , and noting that $G \square G$ is regular with degree $2d$ yields:

$$\mathbf{E}_{\mu \times \mu} e^{t(f(x)-f(y))} \leq e^{dt^2/\rho_1}.$$

On the other hand, letting $M = \max f(x)$ and $m = \min f(x)$, we have

$$\mathbf{E}_{\mu \times \mu} e^{t(f(x)-f(y))} \geq \frac{e^t(M-m)}{|G|^2}.$$

Thus we may conclude that for all $t \in \mathbf{R}$,

$$M - m \leq \frac{dt}{\rho_1} + \frac{2 \log |G|}{t}.$$

Minimizing over t yields, for all Lipschitz f on M ,

$$\max f - \min f \leq 2 \sqrt{\frac{2d}{\rho_1} \log |G|}. \tag{5.6}$$

To conclude the proof of the proposition, let us take $f(x) = d(x, x_0)$, for $x_0 \in M$, and maximize the left hand side of (5.6) over all choices of x_0 . \square

6. EXTREMAL FUNCTIONS

Let (P, π) be a reversible Markov kernel on the finite set M . We assume for definiteness that $\text{card}(M) \geq 2$ and that the stationary probability measure π charges every point in M . Recall that the associate Dirichlet form is given by

$$\mathcal{E}(f, g) = \frac{1}{2} \sum_{x, y \in M} (f(x) - f(y))(g(x) - g(y))P(x, y)\pi(x).$$

In this section, we are going to study extremal functions in logarithmic Sobolev inequalities. As is well-known, the infimum over all non-constant functions f on M ,

$$\inf \frac{\mathcal{E}(f, f)}{\text{Var}_\pi(f)},$$

is attained at some f and represents the spectral gap λ_1 – the optimal constant in the Poincaré-type inequality (1.2). Since

$$\mathcal{E}(f, g) = - \int (Lf)gd\pi, \quad L = P - I,$$

this optimal constant is therefore the smallest eigenvalue λ_1 of $-L$ among those to which non-constant functions f belong as eigenfunctions:

$$-Lf = \lambda_1 f. \tag{6.1}$$

Of course, it may occur that $\lambda_1 = 0$. This is possible if and only if for some non-constant function f on M , we have $\mathcal{E}(f, f) = 0$, that is, $f(x) = f(y)$, whenever $x \neq y$ and $P(x, y) > 0$. Equivalently, after some re-enumeration of the elements in M , the matrix P should be of the form

$$P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix},$$

where P_1 and P_2 are square submatrices of sizes at least 2. In the sequel it is natural to exclude this case from the consideration. Actually, one usually assumes a little more, namely, that (P, π) is irreducible in the sense that, for every pair x, y in M there is a natural number $n = n(x, y)$ such that $P^n(x, y) > 0$ (hence, the case where P_1 or P_2 have size 1 is excluded, as well). In the language of Markov chains, this means that all the states are communicable (which in case of a graph, for example, means that the graph is connected). Thus, the irreducibility ensures that $\lambda_1 > 0$. so this number can be characterized as the minimal non-zero eigenvalue of $-L$.

Since in the class of non-negative f , we have $c_1 \text{Var}_\pi(f) \leq \text{Ent}_\pi(f^2) \leq c_2 \text{Var}_\pi(f)$ with some $c_2 > c_1 > 0$ depending on π (see Remark 6.7 below), the above characterization of the property $\lambda_1 > 0$ is also equivalent to saying that $\rho > 0$, that is, the infimum

$$\inf \frac{\mathcal{E}(f, f)}{\text{Ent}_\pi(f^2)}, \tag{6.2}$$

is positive, as well. The variational problem on finding an extremal function f at which the infimum (6.2) is attained leads to the non-linear equation

$$-Lf = \rho f \log f, \tag{6.3}$$

under the constraint $\|f\|_2^2 = \int f^2 d\pi = 1$.

To be more precise, note the following: if we apply the Eq. (6.3) to functions of the form $\frac{1+\varepsilon f}{\|1+\varepsilon f\|_2}$ and let $\varepsilon \rightarrow 0$, we will come back to the linear Eq. (6.1) with $\lambda_1 = \rho$. Hence, (6.1) can be regarded as an infinitesimal or limiting form of (6.3), and in this sense the latter may be called “generalized”.

Definition 6.1. A non-negative function f on M normalized by $\|f\|_2 = 1$ is called a solution to the generalized Eq. (6.3) if it satisfies (6.3), or (6.1) with $\lambda_1 = \rho$.

With this convention we have the following theorem:

- Theorem 6.2.** (a) There exists a number $\rho > 0$ such that the generalized Eq. (6.3) has a non-negative, non-constant solution f on M .
 (b) Among such numbers there is a minimal value.
 (c) This minimal value represents the optimal constant in the logarithmic Sobolev inequality

$$\rho \text{Ent}_\pi(f^2) \leq 2\mathcal{E}(f, f).$$

Note in advance that in the irreducible case any non-negative solution f to (6.3) with any $\rho > 0$ must be positive on M . Indeed, if $f(x) = 0$ for some $x \in M$, then, by (6.3), $Pf(x) = 0$. The latter implies $f(y) = 0$ whenever $P(x, y) > 0$. Hence, $f(z) = 0$ whenever $P^2(x, z) > 0$. Repeating the argument, we get $f(z) = 0$ whenever $P^n(x, z) > 0$ for some $n \geq 1$. Hence, f must be zero on M .

Corollary 6.3. If within the class of all non-negative functions the only solution to the Eq. (6.3) with an arbitrary $\rho > 0$ is the trivial solution $f = 1$, then $\rho = \lambda_1$.

One may formulate this sufficient condition in a different way: if the infimum in (6.2) is not attained, then $\rho = \lambda_1$. Actually, this will be a part of the proof of Theorem 6.2 – to deduce from $\rho < \lambda_1$ the property that the infimum *is* attained! We were informed by Saloff-Coste that this result is implicit in⁽³³⁾ in the context of Riemannian manifolds, and that it is stated explicitly with a sketch of the proof in Ref. 34. We include the proof here for completeness. We further extend this result (see below) to the modified log-Sobolev inequality defining ρ_0 . It is also worth mentioning that Chen and Shen⁽⁹⁾ utilize the above corollary in establishing that $\rho = \lambda_1$ for the simple random walk on the discrete circle $\mathbf{Z} \bmod m$ when m is even. Computation of the precise value of ρ for odd $m \geq 5$ remains open, while for $m = 3$ it has been computed in⁽¹⁵⁾ and shown to be strictly smaller than the value of λ_1 .

It may indeed occur that (6.3) has no non-negative, non-constant solution. This is the case, for example, for $M = \{0, 1\}$ with uniform measure π . So in this case Corollary 6.3 yields $\rho = \lambda_1$, which is Gross' theorem on the two point space.

On the other hand, when $\rho < \lambda_1$, there is no need to consider the generalized equation (that is, to attach (6.1) to (6.3)). To illustrate this, consider the (weighted) two point space $M = \{0, 1\}$ with the measure π assigning the mass $p \in (0, 1)$ to the point 1 and $q = 1 - p$ to the point 0. In this case the operator $-L$ can be identified with the matrix

$$-L = I - P = \begin{bmatrix} p & -p \\ -q & q \end{bmatrix}.$$

Hence, if $f(0) = u$, $f(1) = v$ with $u, v \geq 0$, the Eq. (6.3) becomes the system

$$\begin{cases} p(u - v) = \rho u \log u \\ -q(u - v) = \rho v \log v, \end{cases}$$

under the constraint $qu^2 + pv^2 = 1$. Adding up the two equations with weights q and p , we get $qu \log u + pv \log v = 0$ which means that the solution (u, v) is determined by p , only, but not by the parameter $\rho > 0$. It is already an easy exercise to see that in case $p \neq q$ the only possible non-constant solution is given by $u = \sqrt{\frac{p}{q}}$, $v = \sqrt{\frac{q}{p}}$ which leads to

$$\rho = \frac{2(p - q)}{\log p - \log q}.$$

Thus, by virtue of Theorem 6.2, we have recovered the second equality in (3.11).

Proof of Theorem 6.2. We assume from the very beginning that $\lambda_1 > 0$ (so $\rho > 0$, as well). There is nothing to prove in case $\rho = \lambda_1$, so assume $\rho < \lambda_1$.

We wish to minimize the functional

$$W(f) = \frac{\mathcal{E}(f, f)}{\text{Ent}_\pi(f^2)}$$

in the class of all non-constant functions f on M . Since $\mathcal{E}(|f|, |f|) \leq \mathcal{E}(f, f)$ with equality only when f is either non-negative or non-positive, we may restrict ourselves to non-negative f 's. Moreover, since $W(cf) = W(f)$ for all constants c , we may assume for definiteness that $\max_{x \in M} f(x) = 1$.

First we show that the minimum to $W(f)$ is indeed attained in this class. Let $\frac{\rho}{2}$ denote the infimum (thus, with ρ being the logarithmic Sobolev constant). Then, there is a sequence of non-negative functions $\{f_n\}_{n \geq 1}$ on M with $\max_{x \in M} f_n(x) = 1$ and such that

$$\lim_{n \rightarrow \infty} W(f_n) = \frac{\rho}{2}. \quad (6.4)$$

For each $x \in M$, the sequence $\{f_n(x)\}_{n \geq 1}$ has a converging subsequence. Since M is finite, such a subsequence can be chosen common for all x . Hence, without loss of generality, we may assume that $f_n(x) \rightarrow f(x)$, as $n \rightarrow \infty$ for all $x \in M$. Necessarily, $0 \leq f \leq 1$ on M . Moreover,

there should be $\max_{x \in M} f(x) = 1$. Indeed, $f_n(x_n) = 1$ for some $x_n \in M$, so for some $x_0 \in M$, we have $x_n = x_0$ for infinitely many indices. Hence, $f(x_0) = 1$. In addition, choosing a suitable subsequence, we may assume that $f_n(x_0) = 1$, for all $n \geq 1$.

Suppose that $f_n(x) \rightarrow 1$, for all $x \in M$, so that $f_n = 1 + g_n$ for some sequence $g_n \rightarrow 0$ on M . Then

$$\mathcal{E}(f_n, f_n) = \mathcal{E}(g_n, g_n), \quad \text{Ent}_\pi(f_n^2) = 2\text{Var}_\pi(g_n) + O(\|g_n\|_\infty^3),$$

by Taylor's expansion in the entropy asymptotic (for all n large enough, say, when $\|g_n\|_\infty \leq \frac{1}{2}$). Recall that $g_n(x_0) = 0$, for all $n \geq 1$. Hence,

$$\begin{aligned} 2\text{Var}_\pi(g_n) &= \sum_{x,y} |g_n(x) - g_n(y)|^2 \pi(x)\pi(y) \geq \sum_x |g_n(x) - g_n(x_0)|^2 \pi(x)\pi(x_0) \\ &= \sum_x |g_n(x)|^2 \pi(x)\pi(x_0) \geq \pi_*^2 \|g_n\|_\infty^2. \end{aligned}$$

Therefore, $\text{Ent}_\pi(f_n^2) = 2(1 + o(1))\text{Var}_\pi(g_n)$, so

$$\liminf_{n \rightarrow \infty} W(f_n) = \liminf_{n \rightarrow \infty} \frac{\mathcal{E}(g_n, g_n)}{2\text{Var}_\pi(g_n)} \geq \inf_g \frac{\mathcal{E}(g, g)}{2\text{Var}_\pi(g)} = \frac{\lambda_1}{2} > \frac{\rho}{2},$$

and we get contradiction with (6.4). Thus the limiting function f cannot identically be 1. In that case, since $f(x_0) = 1$, we get $\text{Ent}_\pi(f^2) > 0$. Since, as $n \rightarrow \infty$,

$$\mathcal{E}(f_n, f_n) \rightarrow \mathcal{E}(f, f), \quad \text{Ent}_\pi(f_n^2) \rightarrow \text{Ent}_\pi(f^2),$$

we conclude that $W(f_n) \rightarrow W(f)$. Consequently, minimum to W is attained at f .

Now, take a non-negative, non-constant function f on M representing a minimizer for W . Let it be normalized so that $\|f\|_2 = 1$. Thus, for any non-constant function h on M , we have $W(f) \leq W(h)$. We apply this inequality to $h = f + \varepsilon g$ with an arbitrary function g on M and real ε small enough, in order to derive the Eq. (6.3). We have

$$\|f + \varepsilon g\|_2^2 = 1 + 2\varepsilon \int fg d\pi + \varepsilon^2 \int g^2 d\pi,$$

so, as $\varepsilon \rightarrow 0$,

$$\|f + \varepsilon g\|_2^2 \log \|f + \varepsilon g\|_2^2 = 2\varepsilon \int fg d\pi + O(\varepsilon^2).$$

Similarly, since the derivative $(t^2 \log t^2)' = 2t(\log t^2 + 1)$ is continuous on the whole real line,

$$\int (f + \varepsilon g)^2 \log(f + \varepsilon g)^2 d\pi = \int f^2 \log f^2 d\pi + 2\varepsilon \int f g (\log f^2 + 1) d\pi + o(\varepsilon),$$

and we find

$$\text{Ent}_\pi(f + \varepsilon g)^2 = \text{Ent}_\pi(f^2) + 2\varepsilon \int f \log f^2 g d\pi + o(\varepsilon).$$

On the other hand,

$$\mathcal{E}(f + \varepsilon g, f + \varepsilon g) = \mathcal{E}(f, f) + 2\varepsilon \mathcal{E}(f, g) + \varepsilon^2 \mathcal{E}(g, g).$$

Combining both the asymptotics and using $\rho \text{Ent}_\pi(f^2) = 2\mathcal{E}(f, f)$, we get

$$\rho \text{Ent}_\pi(f + \varepsilon g)^2 - 2\mathcal{E}(f + \varepsilon g, f + \varepsilon g) = 4\varepsilon \left[\rho \int f \log f g d\pi - \mathcal{E}(f, g) \right] + o(\varepsilon).$$

By the log-Sobolev inequality, the left hand side is non-positive. But the right hand side is non-positive for all ε small enough if and only if $\mathcal{E}(f, g) = \rho \int f \log f g d\pi$, or equivalently, if

$$\int (Lf + \rho f \log f) g d\pi = 0.$$

Since this holds for all g on M , the integrand with respect to $g d\pi$ must be zero:

$$Lf + \rho f \log f = 0.$$

Thus, the logarithmic Sobolev constant ρ is among positive numbers such that the above equation, that is, the Eq. (6.3), has a non-negative, non-constant solution f with $\|f\|_2 = 1$.

Minimality: let ρ' be another one with this property, i.e., such that for some non-negative, non-constant function f on M with $\|f\|_2 = 1$, we have

$$-2Lf = \rho' f \log f^2.$$

Multiplying this equality by f and integrating with respect to π , we get

$$2\mathcal{E}(f, f) = \rho' \text{Ent}_\pi(f^2).$$

But $2\mathcal{E}(f, f) = \rho \text{Ent}_\pi(f^2)$ for all f on M , so necessarily $\rho' \geq \rho$. This completes the proof of Theorem 6.2. \square

Now let us turn to the modified logarithmic Sobolev inequality (3.3) which can be written equivalently as

$$\rho_0 \text{Ent}(f) \leq \frac{1}{2} \mathcal{E}(f, \log f), \tag{6.5}$$

in the class of all positive functions f on M . This is a different inequality, and the corresponding variational problem leads to a different equation

$$-Lf - fL(\log f) = 2\rho_0 f \log f, \quad \|f\|_1 = 1, \tag{6.6}$$

under the constraint $\|f\|_1 = \int f d\pi = 1$.

Note that if we apply this equation to functions of the form $\frac{1+\varepsilon f}{\|1+\varepsilon f\|_1}$ with small ε , we come back again to the linear Eq. (6.1) with $\lambda_1 = \rho_0$. Hence, (6.1) can also be regarded as an infinitesimal or limiting form of (6.6), and in this sense the latter may once again be called ‘‘generalized’’:

Definition 6.4. A positive function f on M normalized by $\|f\|_1 = 1$ is called a solution to the generalized Eq. (6.6) if it satisfies (6.6), or (6.1) with $\lambda_1 = \rho_0$.

With this convention we have the following analogue of Theorem 6.2.

- Theorem 6.5.** (a) There exists a number $\rho_0 > 0$ such that the generalized Eq. (6.6) has a positive, non-constant solution f on M .
 (b) Among such numbers there is a minimal value.
 (c) This minimal value represents the optimal constant in the logarithmic Sobolev inequality (6.5).

Proof. The argument is similar to the one used in the proof of Theorem 6.2. Again, there is nothing to prove in case $\rho_0 = \lambda_1$, so assume $\rho_0 < \lambda_1$. We wish to minimize the new functional

$$W(f) = \frac{\mathcal{E}(f, \log f)}{\text{Ent}_\pi(f)},$$

in the class of all positive functions f on M . Since $W(cf) = W(f)$ for all constants c , we may assume for definiteness that $\max_{x \in M} f(x) = 1$. \square

First we show that the minimum to $W(f)$ is attained in this class and is positive. Let $2\rho_0$ denote the infimum (thus, with ρ_0 being the modified logarithmic Sobolev constant). Then, there is a sequence of positive functions $\{f_n\}_{n \geq 1}$ on M with $\max_{x \in M} f_n(x) = 1$ and such that

$$\lim_{n \rightarrow \infty} W(f_n) = 2\rho_0. \tag{6.7}$$

As in the proof of Theorem 6.2, we may assume that this sequence is converging point-wise to a limiting function f with values in $[0, 1]$, and moreover, for some $x_0 \in M$, $f(x_0) = f_n(x_0) = 1$, for all $n \geq 1$.

If $f_n(x) \rightarrow 1$, for all $x \in M$, so that $f_n = 1 + g_n$ for some $g_n \rightarrow 0$ on M , then, by Taylor's expansion, once $\|g_n\|_\infty \leq \frac{1}{2}$, we have

$$\begin{aligned} \mathcal{E}(f_n, \log f_n) &= (1 + O(\|g_n\|_\infty))\mathcal{E}(g_n, g_n), \\ \text{Ent}_\pi(f_n) &= \frac{1}{2}\text{Var}_\pi(g_n) + O\left(\|g_n\|_\infty^3\right). \end{aligned}$$

As we know, $2\text{Var}_\pi(g_n) \geq \pi_*^2\|g_n\|_\infty^2$ which implies that $\text{Ent}_\pi(f_n) = \frac{1+o(1)}{2}\text{Var}_\pi(g_n)$. Therefore,

$$\liminf_{n \rightarrow \infty} W(f_n) = \liminf_{n \rightarrow \infty} \frac{\mathcal{E}(g_n, g_n)}{\frac{1}{2}\text{Var}_\pi(g_n)} \geq \inf_g \frac{2\mathcal{E}(g, g)}{\text{Var}_\pi(g)} = 2\lambda_1 > 2\rho_0,$$

and we get a contradiction to (6.7). Thus the limiting function f cannot be identically 1. Hence, since $f(x_0) = 1$, we get $\text{Ent}_\pi(f) > 0$.

Now, let us see that f must be positive. Put

$$M_0 = \{x \in M : f(x) = 0\}, \quad M_1 = \{y \in M : f(y) > 0\}.$$

The second set contains the point x_0 , so it is non-empty. Note that the irreducibility is exactly the property that, for any partition of M into two non-empty subsets M_0 and M_1 , there are $x \in M_0$, $y \in M_1$ such that $P(x, y) > 0$. But in our case for such points, we have

$$\mathcal{E}(f_n, \log f_n) \geq (f_n(x) - f_n(y))(\log f_n(x) - \log f_n(y))\pi(x)P(x, y) \rightarrow +\infty,$$

as $n \rightarrow \infty$, which implies $W(f_n) \rightarrow +\infty$, so $\rho_0 = +\infty$! Hence, the set M_0 must be empty.

Thus, $0 < f(x) \leq 1$, for all $x \in M$, so

$$\mathcal{E}(f_n, \log f_n) \rightarrow \mathcal{E}(f, \log f), \quad \text{Ent}_\pi(f_n) \rightarrow \text{Ent}_\pi(f),$$

Since $\text{Ent}_\pi(f) > 0$, we conclude that $W(f_n) \rightarrow W(f)$. Consequently, minimum to W is attained at f . Now, take a positive, non-constant function f on M representing a minimizer for W :

$$\rho_0 \text{Ent}(f) = \frac{1}{2}\mathcal{E}(f, \log f). \tag{6.8}$$

Let $\|f\|_1 = 1$. Given an arbitrary function g on M , $f + \varepsilon g$ is positive on M for small ε , and moreover,

$$\|f + \varepsilon g\|_1 \log \|f + \varepsilon g\|_1 = \varepsilon \int g d\pi + O(\varepsilon^2).$$

Similarly, by Taylor's expansion,

$$\int (f + \varepsilon g) \log(f + \varepsilon g) d\pi = \int f \log f d\pi + \varepsilon \int (\log f + 1) g d\pi + O(\varepsilon^2),$$

and we find

$$\text{Ent}_\pi(f + \varepsilon g) = \text{Ent}_\pi(f) + \varepsilon \int \log f g d\pi + O(\varepsilon^2).$$

On the other hand,

$$\begin{aligned} \mathcal{E}(f + \varepsilon g, \log(f + \varepsilon g)) &= \mathcal{E}\left(f + \varepsilon g, \log f + \varepsilon \frac{g}{f} + O(\varepsilon^2)\right) \\ &= \mathcal{E}(f, \log f) + \varepsilon \mathcal{E}(\log f, g) + \varepsilon \mathcal{E}\left(f, \frac{g}{f}\right) + O(\varepsilon^2). \end{aligned}$$

Combining both the asymptotics and using (6.8), we get

$$\begin{aligned} &2\rho_0 \text{Ent}_\pi(f + \varepsilon g) - \mathcal{E}(f + \varepsilon g, \log(f + \varepsilon g)) \\ &= \varepsilon \left[2\rho_0 \int \log f g d\pi - \mathcal{E}(\log f, g) - \mathcal{E}\left(f, \frac{g}{f}\right) \right] + O(\varepsilon^2). \end{aligned} \quad (6.9)$$

Now, using reversibility,

$$\begin{aligned} \mathcal{E}\left(f, \frac{g}{f}\right) &= \frac{1}{2} \sum_{x,y} (f(x) - f(y)) \left(\frac{g(x)}{f(x)} - \frac{g(y)}{f(y)} \right) \pi(x) P(x, y) \\ &= \frac{1}{2} \sum_{x,y} \left(g(x) + g(y) - f(x) \frac{g(y)}{f(y)} - f(y) \frac{g(x)}{f(x)} \right) \pi(x) P(x, y) \\ &= \sum_x \left[g(x) \pi(x) - \frac{g(x)}{f(x)} \sum_y f(y) \pi(x) P(x, y) \right] \\ &= \int \left[1 - \frac{Pf(x)}{f(x)} \right] g(x) d\pi(x). \end{aligned}$$

Hence, recalling that $\mathcal{E}(\log f, g) = -\int L(\log f) g d\pi$, the expression in (6.9) can be written as

$$\mathcal{E} \int \left[2\rho_0 \log f + L(\log f) - 1 + \frac{Pf}{f} \right] g d\pi + O(\varepsilon^2).$$

Since it is non-positive, the coefficient in front of ε must be zero:

$$\int \left[2\rho_0 \log f + L(\log f) - 1 + \frac{Pf}{f} \right] g d\pi = 0.$$

Since g is arbitrary, we obtain

$$2\rho_0 \log f + L(\log f) - 1 + \frac{Pf}{f} = 0.$$

Multiplying by f and recalling $L = P - I$, we finally get

$$2\rho_0 f \log f + fL(\log f) + Lf = 0,$$

which is exactly the desired Eq. (6.6). Thus, the logarithmic Sobolev constant ρ_0 is among positive numbers such that the Eq. (6.6), has a positive, non-constant solution f with $\|f\|_1 = 1$.

Minimality: let ρ'_0 be another one with this property, i.e. such that for some positive, nonconstant function f on M with $\|f\|_1 = 1$, we have

$$\rho'_0 f \log f = -\frac{1}{2} fL(\log f) - \frac{1}{2} Lf.$$

Integrating this equality with respect to π , we get

$$\rho'_0 \text{Ent}_\pi(f) = \frac{1}{2} \mathcal{E}(f, \log f)$$

But $\rho_0 \text{Ent}_\pi(f) \leq \frac{1}{2} \mathcal{E}(f, \log f)$, for all f on M , so necessarily $\rho'_0 \geq \rho_0$. This completes the proof of Theorem 6.5.

Remark 6.6. As we have already mentioned, in the class of all non-negative functions f on a finite probability space (M, π) , the entropy and variance functionals are connected by the relation

$$c_1 \text{Var}_\pi(f) \leq \text{Ent}_\pi(f^2) \leq c_2 \text{Var}_\pi(f).$$

In a Markov kernel setting, the second inequality may be viewed as the logarithmic Sobolev inequality $\text{Ent}_\pi(f^2) \leq c_2 \mathcal{E}(f, f)$ on the complete graph M with transition probabilities $P(x, y) = \pi(x)$. This case is listed in Example 3.10: the best constant, found in [15] by Diaconis and Saloff-Coste, is given according to (3.18) by

$$c_2 = \frac{\log p - \log q}{p - q}, \quad \text{where } p = \pi_* = \min_{x \in M} \pi(x), \quad q = 1 - p.$$

Thus, c_2 is of order $\log \frac{1}{\pi_*}$. By Remark 6.6, this statement may be strengthened to $\mathcal{L}(f) \leq C \log \frac{1}{\pi_*} \text{Var}_\pi(f)$, is a universal constant.

As for the constant c_1 , it can be chosen to be independent of π . Here we include a simple argument given in [L-O]. In general, by Hölder's

inequality, any function of the form $u(t) = (\mathbf{E}_\pi f^{1/t})^{2t}$ is convex, so the ratio

$$\varphi(t) = \frac{u(t) - u(1/2)}{t - 1/2},$$

is non-decreasing in $t > \frac{1}{2}$. But $\varphi(1) = -2\text{Var}_\pi(f)$, while $\varphi(\frac{1}{2}+) = -2\text{Ent}_\pi(f^2)$. Hence, one may take $c_1 = 1$. A careful analysis shows that this constant is optimal.

7. HYPERCONTRACTIVITY

In this last section, we are considering a description of the logarithmic Sobolev constant ρ_0 in terms of a hypercontractive property of Markov semigroups. Our treatment is very general, including both the continuous and the discrete cases. Let (M, μ) be a probability space and let \mathcal{A} be a linear space of bounded measurable functions on M . Let $L : \mathcal{A} \rightarrow \mathcal{A}$ be a linear operator associated with the Dirichlet form $\mathcal{E} : \mathcal{A} \times \mathcal{A} \rightarrow \mathbf{R}$ in the sense that

$$-\int f Lg d\mu = \mathcal{E}(f, g), \quad \text{for all } f, g \in \mathcal{A}. \tag{7.1}$$

It is easy to see that whenever such an operator exists, it is unique. Furthermore, assume a family of linear operators $P_t : \mathcal{A} \rightarrow \mathcal{A}$, $t \geq 0$, is associated with \mathcal{E} , with the properties that:

- (1) P_0 is the identity operator;
- (2) P_t form a semigroup: $P_{t+s} = P_t P_s$, for all $t, s \geq 0$;
- (3) P_t has generator L satisfying the relation (7.1);
- (4) For all $f \in \mathcal{A}$ and $t_0 > 0$, $\sup_{0 \leq t \leq t_0} \|P_t f\|_\infty < +\infty$.

The property of having a generator is understood in $L^1(\mu)$ sense: for all $f \in \mathcal{A}$ and $t \geq 0$,

$$\lim_{\varepsilon \rightarrow 0} \frac{P_{t+\varepsilon} f - P_t f}{\varepsilon} = L(P_t f), \tag{7.2}$$

with convergence in the norm of the space $L^1(\mu)$. Then we have:

Theorem 7.1. Given a number ρ_0 , the following properties are equivalent:

- (a) The Dirichlet form \mathcal{E} satisfies the modified logarithmic Sobolev inequality

$$\rho_0 \text{Ent}(e^f) \leq \frac{1}{2} \mathcal{E}(e^f, f), \quad f \in \mathcal{A}. \tag{7.3}$$

(b) For all $t \geq 0$ and $f \in \mathcal{A}$,

$$\|e^{P_t f}\|_{e^{2\rho_0 t}} \leq \|e^f\|_1. \tag{7.4}$$

All the norms here are taken in the Lebesgue spaces $L^q(\mu)$ (although we say “norm” even if $q < 1$). The equivalence of (7.4) and log-Sobolev inequality (3.2) is well known in the continuous setting (cf. Bakry and Emery);⁽³⁾ also see Ref. 15 for a thorough discussion of this equivalence in the context of finite Markov chains. Here we are dealing with the most general formulation fitting both continuous and discrete cases. The main point and motivation is that, in discrete spaces, the constant ρ_0 can be much better than ρ .

We need the following standard fact (proof omitted) to prove the theorem.

Lemma 7.2. For any continuously differentiable function $\varphi : \mathbf{R} \rightarrow \mathbf{R}$, the function of the form $t \rightarrow \int \varphi(P_t f) d\mu$ is differentiable on the half-axis $[0, +\infty)$ and has derivative

$$\frac{d}{dt} \int \varphi(P_t f) d\mu = \int \varphi'(P_t f) L(P_t f) d\mu, \quad t \geq 0. \tag{7.5}$$

If the derivative φ' is bounded on the whole real line, the assumption (4) is not needed. However, we will need Lemma 7.2 for exponential functions $\varphi(s) = e^{qs}$ and for $\varphi(s) = e^{q^2 s}$. Nevertheless, (4) is automatically implied by (3), when the set M is finite. In this case, $L^\infty(\mu)$ -norm is equivalent to $L^1(\mu)$ -norm, while (3) yields continuity of the function $t \rightarrow \|P_t f\|_1$. Note also that, for finite sets, there always exists a unique operator L and a semi-group P_t associated with \mathcal{E} .

Proof of Theorem 7.1. First note that, for any bounded measurable function g on M , the function of the form

$$u(q) = \|e^g\|_q = \left(\int e^{qg} d\mu \right)^{\frac{1}{q}}, \quad q > 0,$$

is smooth in q and has derivative

$$u'(q) = \frac{1}{q^2} \left(\int e^{qg} d\mu \right)^{\frac{1}{q}-1} \text{Ent}(e^{qg}), \quad q > 0. \tag{7.6}$$

Now let us fix $q > 0$, and for $f \in \mathcal{A}$, consider the function of the form

$$v(t) = \|e^{P_t f}\|_q = \left(\int e^{q P_t f} d\mu \right)^{\frac{1}{q}}, \quad t \geq 0.$$

Applying the chain rule and Lemma 7.2, we can perform differentiation of $v(t)$:

$$v'(t) = \left(\int e^{qP_t f} d\mu \right)^{\frac{1}{q}-1} \int e^{qP_t f} L(P_t f) d\mu. \tag{7.7}$$

These observations show that the function of the two variables $G(t, q) = \|e^{P_t f}\|_q, (t, q) \in [0, +\infty) \times (0, +\infty)$, has partial derivatives which can be found according to (7.6)–(7.7):

$$\frac{\partial G}{\partial t} = \left(\int e^{qP_t f} d\mu \right)^{\frac{1}{q}-1} \int e^{qP_t f} L(P_t f) d\mu, \quad t \geq 0, \tag{7.8}$$

$$\frac{\partial G}{\partial q} = \frac{1}{q^2} \left(\int e^{qP_t f} d\mu \right)^{\frac{1}{q}-1} \text{Ent}(e^{qP_t f}), \quad q > 0. \tag{7.9}$$

To prove that G is really differentiable with respect to the pair of the variables, consider the function $K(t, q) = \|e^{P_t f}\|_q^q = \int e^{qP_t f} d\mu$. Fix $t \geq 0$ and $q > 0$. By property (4) of the semi-group P_t , the functions $P_{t+\varepsilon}$ are uniformly bounded for bounded ε (say, for $|\varepsilon| < 1$). Hence, for any $\varepsilon > -t$ with $|\varepsilon| < 1$ and $\delta > -q$ (both parameters will tend to zero independently of each other), we have

$$\begin{aligned} K(t + \varepsilon, q + \delta) &= \int e^{qP_{t+\varepsilon} f} e^{\delta P_{t+\varepsilon} f} d\mu \\ &= \int e^{qP_{t+\varepsilon} f} (1 + \delta P_{t+\varepsilon} f + O(\delta^2)) d\mu \\ &= \int e^{qP_{t+\varepsilon} f} d\mu + \delta \int e^{qP_{t+\varepsilon} f} P_{t+\varepsilon} f d\mu + O(\delta^2), \end{aligned}$$

where the constants in both the appearances of $O(\delta^2)$ can be chosen depending on (t, q) , only (and not on $x \in M$). To continue the above expansion – now over ε , it remains to apply Lemma 7.2 with $\varphi(s) = e^{q s}$ for the first integral and with $\varphi(s) = e^{q s} s$ for the second one. As a result, we will arrive at an expression of the form

$$K(t + \varepsilon, q + \delta) = K(t, q) + c_1 \varepsilon + \delta(c_2 + c_3 \varepsilon + o(\varepsilon)) + O(\delta^2),$$

as $\varepsilon, \delta \rightarrow 0$. This proves that K is differentiable, and so is $G = K^{1/q}$.

Now, given a differentiable function $q : [0, +\infty) \rightarrow [0, +\infty)$, consider

$$F(t) = G(t, q(t)) = \|e^{P_t f}\|_{q(t)}.$$

This function is also differentiable on $[0, +\infty)$ as a superposition of differentiable functions. According to the chain rule and by (7.8)–(7.9), we find that, for any $t \geq 0$,

$$\begin{aligned} F'(t) &= \frac{\partial G(t, q(t))}{\partial t} + \frac{\partial G(t, q(t))}{\partial q} q'(t) \\ &= \left(\int e^{q(t)P_t f} d\mu \right)^{\frac{1}{q(t)}-1} \left(\int e^{q(t)P_t f} L(P_t f) d\mu + \frac{q'(t)}{q^2(t)} \text{Ent}(e^{q(t)P_t f}) \right), \end{aligned}$$

or equivalently,

$$\begin{aligned} q(t) \left(\int e^{q(t)P_t f} d\mu \right)^{1-\frac{1}{q(t)}} F'(t) \\ = \int e^{q(t)P_t f} L(q(t)P_t f) d\mu + \frac{q'(t)}{q(t)} \text{Ent}(e^{q(t)P_t f}). \end{aligned}$$

Introduce $f_t = q(t)P_t f$. Using (1.1), the first integral on the right is just $-\mathcal{E}(e^{f_t}, f_t)$, and we get

$$q(t) \left(\int e^{f_t} d\mu \right)^{1-\frac{1}{q(t)}} F'(t) = \frac{q'(t)}{q(t)} \text{Ent}(e^{f_t}) - \mathcal{E}(e^{f_t}, f_t).$$

For the particular choice $q(t) = e^{2\rho_0 t}$, the equality becomes

$$q(t) \left(\int e^{f_t} d\mu \right)^{1-\frac{1}{q(t)}} F'(t) = 2\rho_0 \text{Ent}(e^{f_t}) - \mathcal{E}(e^{f_t}, f_t). \quad (7.10)$$

Now, starting from (7.3), the right hand side of (7.10) will be non-positive, so $F'(t) \leq 0$, for all $t \geq 0$. Hence F is non-increasing, so $F(t) \leq F(0)$. The latter is exactly the conclusion (7.4) since $f_0 = f$. Conversely, if (7.4) holds true, that is, if $F(t) \leq F(0)$, for all $t \geq 0$, then, necessarily $F'(0) \leq 0$. This implies that the right hand side of (7.10) is non-negative at $t=0$, which is exactly (7.3), thus establishing Theorem 7.1. \square

ACKNOWLEDGMENTS

We thank L. Saloff-Coste for providing us with several relevant references. We also thank the anonymous referees for a very careful reading of our manuscript. Research supported in part by NSF Grants DMS-0103929, DMS-0405587 (Sergey G. Bobkov) Research supported in part by NSF Grants DMS-0100298, DMS-0401239; research done while visiting Microsoft Research (Prasad Tetali)

REFERENCES

1. Alon, N., and Milman, V. D. (1985). λ_1 , isoperimetric inequalities for graphs and super-concentrators. *J. Comb. Theory Ser. B* **38**, 73–88.
2. Bakry, D. (1994). L'hypercontractivité et son utilisation en théorie des semigroups. *Ecole d'Été de Saint Flour, 1992. Lect. Notes in Math.*, Vol. 1581, Springer-Berlin.
3. Bakry, D., and Emery, M. (1994). Diffusions hypercontractive. *Séminaire de Probabilités XIX, Lect. Notes in Math.*, Vol. 1123, Springer, Berlin, pp. 179–206.
4. Beckner, W. (1989). A generalized Poincaré inequality for Gaussian measures. *Proc. of the AMS*, **105**(2), 397–400.
5. Bobkov, S. G., and Götze, F. (1999). Exponential integrability and transportation cost related to logarithmic transportation inequalities. *J. Funct. Anal.* **163**, 1–28.
6. Bobkov, S. G., and Ledoux, M. (1997). Poincaré inequalities and Talagrand's measure concentration phenomenon for the exponential measure, *Prob. Th. Relat. Fields* **107**, 383–400.
7. Bobkov, S. G., and Ledoux, M. (1998). On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. *J. Funct. Anal.* **156**, 347–365.
8. Bobkov, S. G., and Tetali, P. (June 2003). Modified log-Sobolev inequalities, mixing and hypercontractivity. *Proc. of Symposium on Theory of computing*, San Diego, pp. 287–296.
9. Chen, G.-Y., and Shen, Y.-C. (2003). On the log-Sobolev constant for the simple random walk on the n -cycle: the even cases. *J. Funct. Anal.* **202**, 473–485.
10. Chung, F. R. K. (1989). Diameters and Eigenvalues. *J. Amer. Math. Soc.* **2**, 187–196.
11. Chung, F. R. K., Grigor'yan, A., and Yau, S.-T. (2000). Higher eigenvalues and isoperimetric inequalities on Riemannian manifolds and graphs. *Comm. Anal. Geom.* **8**(5), 969–1026.
12. Cover, T. M., and Thomas, J. A. (1991). *Elements of Information Theory*, John Wiley & Sons, New York.
13. Dai Pra, P., Paganoni, A.M., and Posta, G. (2002). Entropy inequalities for unbounded spin systems. *Ann. Probab.* **30**, 1959–1976.
14. Diaconis, P. (1988). *Group representations in Probability and Statistics*, IMS, Hayward, CA.
15. Diaconis, P., and Saloff-Coste, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6**, 695–750.
16. Diaconis, P., and Shahshahani, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57**(2), 159–179.
17. Diaconis, P., and Shahshahani, M. (1987). Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.* **18**, 208–218.
18. Durrett, R. (1999). *Essentials of Stochastic Processes. Springer Texts in Statistics*, Springer-Verlag, New York Inc.
19. Frieze, A., and Kannan, R. (1998). *Log-Sobolev Inequalities and Sampling from Log-Concave Distributions*. Preprint.
20. Gao, F., and Quastel, J. (2003). Exponential decay of entropy in the Random Transposition and Bernoulli-Laplace models. *Ann. Appl. Probab.* **13**, 1591–1600.
21. Goel, S. (2004). Modified logarithmic Sobolev inequalities for some models of random walk. *Stoch. Proc. Their Appl.* **114**, 51–79.
22. Gross, L. (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.* **97**, 1060–1083.
23. Higuchi, Y., and Yoshida, N. (1995). Analytic conditions and phase transition for Ising models. *Lect. Notes in Japanese*, 1995.
24. Houdré, C., and Tetali, P. (2001). Concentration of measure for products of Markov kernels and graph products via functional inequalities. *Comb. Probab. Comp.* **10**, 1–28.

25. Latala, R., and Oleszkiewicz, K. (2000). Between Sobolev and Poincaré. *Geometric Aspects Function. Anal. Lect. Notes Math.* **1745**, 147–168.
26. Ledoux, M. (1996). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.* **1**, 63–87.
27. Ledoux, M. (1999). Concentration of measure and logarithmic Sobolev inequalities and the phenomenon. *Séminaire de Probabilités XXXIII, Lecture Notes Math.*, **1709**, 120–216.
28. Ledoux, M. (2001). The concentration of measure phenomenon. *Mathematical Surveys and Monographs*, 89. American Mathematical Society, Providence, RI.
29. Lee, T. Y., and Yau, H. T. (1998). Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.* **26**(4), 1855–1873.
30. Lieb, E. H. (1975). Some convexity and subadditivity properties of entropy. *Bull. Amer. Math. Soc.* **81**, 1–13.
31. Miclo, L. (1996). Sur les problèmes de sortie discrets inhomogènes. *Ann. Appl. Probab.* **6**(4), 1112–1156.
32. Mohar, B. (1991). Eigenvalues, diameter, and mean distance in graphs. *Graphs Combin.* **7**, 53–64.
33. Rothaus, O. (1981). Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities. *J. Funct. Anal.* **42**, 102–109.
34. Saloff-Coste, L. (1997). Lectures on finite Markov chains. *Lect. Notes in Math.* Vol. 1665, Springer, Berlin, pp. 301–413.
35. Sammer, M., and Tetali, P. (2004). Concentration and Transportation. Preprint (2004); conference version in *SIAM Conf. on Discrete Math.* (June 2004).
36. Stark, D. (2002). Information loss in top to random shuffling. *Combin. Probab. & Comput.* **11**, 607–627.
37. Stroock, D. (1993). Logarithmic Sobolev inequalities for Gibbs measures. *Lect. Notes in Math.*, Vol. 1563, Springer, Berlin.