Chapter 9

Logarithmic Sobolev inequalities

We know that the spectral gap of the random walk on the n-dimensional cube is $\Theta(1/n)$, and that this entails an $O(n^2)$ bound on mixing time. This quadratic bound is made up from a linear factor arising from the reciprocal of the spectral gap, and another linear factor expressing the dependency on the initial distribution. This dependency has the form $\log(1/\pi(x_0))$, assuming the walk starts at a fixed initial state x_0. Whereas the contribution from the inverse spectral gap seems inescapable, one suspects that the factor $\log(1/\pi(x_0))$ might exaggerate the penalty for starting at a point-mass initial distribution. The logarithmic Sobolev constant introduced in this chapter is a parameter that in a sense incorporates more information than spectral gap, allowing one in favourable circumstances to replace $\log(1/\pi(x_0))$ by $\log\log(1/\pi(x_0))$. Sometimes, as in the case of the random walk on the cube, this improvement leads to a tight bound on mixing time.

The seminal work on logarithmic Sobolev inequalities was done by Gross [40]. The important role of logarithmic Sobolev inequalities in the analysis of the mixing time of MCs was revealed in an expository paper of Diaconis and Saloff-Coste [20]. An early algorithmic application was presented by Frieze and Kannan [35]. Much of this chapter, up to the end of §9.3, is plundered from Guionnet and Zegarlinski's lecture notes [41].

The key idea is to replace variance, which played a leading role in Chapter 8, with the entropy-like quantity

$$\mathcal{L}_\pi(f) := \mathbb{E}_\pi\left[f^2\big(\ln f^2 - \ln(\mathbb{E}_\pi f^2)\big)\right].$$

A logarithmic Sobolev inequality (c.f. (5.7)) has the form

$$\mathcal{E}_P(f,f) \geq \alpha\,\mathcal{L}_\pi(f), \quad \text{for all } f : \Omega \to \mathbb{R}, \tag{9.1}$$

where $\alpha > 0$ is the *logarithmic Sobolev constant* ("log-Sobolev" constant).

For a function $f : \Omega \to \mathbb{R}^+$, we use $\|f\|_{\pi,q}$ to denote

$$\|f\|_{\pi,q} = \left[\sum_{x\in\Omega} \pi(x) f(x)^q\right]^{1/q},$$

so that $\mathbb{E}_\pi f^q = \|f\|_{\pi,q}^q$. Observe that the substitution $f \to |f|$ leaves the r.h.s. of (9.1) unchanged, and does not increase the l.h.s. Therefore, condition (9.1) is equivalent to one in which the quantification is over non-negative functions $f : \Omega \to \mathbb{R}^+$. Then, by substituting $f^{q/2}$ for f, we see that (9.1) is equivalent to

$$\mathcal{E}_P(f^{q/2}, f^{q/2}) \geq \alpha q \, \mathbb{E}_\pi \left[f^q \ln \frac{f}{\|f\|_{\pi,q}} \right], \quad \text{for all } f : \Omega \to \mathbb{R}^+, \tag{9.2}$$

for any $q > 0$.

9.1 The relationship between logarithmic Sobolev and Poincaré inequalities

Before considering the relationship between the logarithmic Sobolev constant α and mixing time, it is instructive to compare α directly with the familiar Poincaré constant λ.

Theorem 9.1. *Denote by α and λ the optimal logarithmic Sobolev and Poincaré constants for some MC with transition matrix P. Then $\lambda \geq 2\alpha$.*

Proof. The proof is due to Rothaus [69].

Let $f : \Omega \to \mathbb{R}$ be an arbitrary function with $\mathbb{E}_\pi f = 0$. By the logarithmic Sobolev inequality,

$$\begin{aligned} \varepsilon^2 \mathcal{E}_P(f, f) &= \mathcal{E}_P(1 + \varepsilon f, 1 + \varepsilon f) \\ &\geq \alpha \, \mathbb{E}_\pi \Big[(1 + \varepsilon f)^2 \big\{ \ln((1 + \varepsilon f)^2) - \ln \mathbb{E}_\pi[(1 + \varepsilon f)^2] \big\} \Big], \end{aligned} \tag{9.3}$$

for all $\varepsilon > 0$. When ε is sufficiently small, $1 + \varepsilon f$ is a strictly positive function, and we may expand (9.3) as a Taylor series in ε:

$$\begin{aligned} \varepsilon^2 \mathcal{E}_P(f, f) &\geq \alpha \, \mathbb{E}_\pi \Big[(1 + \varepsilon f)^2 \big\{ 2\varepsilon f - \varepsilon^2 f^2 - \varepsilon^2 \, \mathbb{E}_\pi f^2 + O(\varepsilon^3) \big\} \Big] \\ &= \alpha \, \mathbb{E}_\pi \big[2\varepsilon f + 3\varepsilon^2 f^2 - \varepsilon^2 \, \mathbb{E}_\pi f^2 + O(\varepsilon^3) \big] \\ &= 2\varepsilon^2 \alpha \, \mathbb{E}_\pi f^2 + O(\varepsilon^3) \\ &= 2\varepsilon^2 \alpha \operatorname{Var}_\pi f + O(\varepsilon^3). \end{aligned}$$

Letting $\varepsilon \to 0$, we see that $\lambda \geq 2\alpha$. □

The advantage of the logarithmic Sobolev constant over spectral gap, as we shall see in §9.3, is that α is more tightly related to mixing time than λ. The main disadvantage is that the inequality assured by Theorem 9.1 is not always tight, and even when it is, α may be harder to calculate than λ. It is natural to ask how big the gap can be between α and λ, but we do not pause to consider that question here. Those seeking an answer are directed to Diaconis and Saloff-Coste [20, Cor. A.4].

9.2 Hypercontractivity

Just as spectral gap is related to decay of variance, so the logarithmic Sobolev constant is related to a more powerful phenomenon known as "hypercontractivity". For conciseness, we write f_t for $P^t f$, where, as usual, $P^t f : \Omega \to \mathbb{R}$ denotes the function defined by

$$[P^t f](x) = \sum_{y \in \Omega} P^t(x, y) f(y), \quad \text{for all } x \in \Omega.$$

For convenience, we'll work in continuous time (refer to §5.5). Recall that $P^t = \exp(Qt)$ where $Q = P - I$, and that $\frac{d}{dt} f_t = Qf_t$.

Lemma 9.2. *Let $q(t) = 1 + e^{2\alpha t}$, where α satisfies (9.1), and let $f : \Omega \to \mathbb{R}^+$ be any non-negative function. Then, for all $t \geq 0$,*

$$\frac{d}{dt} \|f_t\|_{\pi,q(t)} \leq 0.$$

Remark 9.3. Recall, from §5.5, the analogous statement for spectral gap λ, which in the notation of the current section could be written

$$\frac{d}{dt} \|f_t\|_{\pi,2}^2 \leq -2\lambda \|f_t\|_{\pi,2}^2,$$

assuming f is normalised so that $\mathbb{E}_\pi f = 0$. In that section, we fixed $q = 2$ and investigated the the decay of $\|f_t\|_{\pi,q}$ with time. In contrast, in Lemma 9.2 we set a fixed bound for $\|f_t\|_{\pi,q(t)}$ but arrange for $q(t)$ to increase with time t, so that the variation of f_t is being measured with respect to an ever more demanding norm. Since $q(t)$ increases exponentially fast with t, the norm we are working with soon comes "close" to the ℓ_∞ norm. Thus Lemma 9.2 makes a powerful statement about $f_t(x)$ at every point x, and in particular when x is the initial state.

The proof of Lemma 9.2 may be clarified by introducing the general Dirichlet form $\mathcal{E}_P(f,g)$. Until now, we have encountered the Dirichlet form only the special case $f = g$, and this allowed us the luxury of being able to use various expressions for $\mathcal{E}_P(f,f)$ interchangeably. It is important to note that these equivalent definitions do not remain equivalent when generalised, in the natural way, to the situation $f \neq g$, at least when P is not time-reversible. Since in this chapter we sometimes want to allow $f \neq g$, while at the same time not restricting ourselves to the time-reversible case, it is important for us to use the "correct" definition, which is

$$\mathcal{E}_P(f,g) = -\mathbb{E}_\pi[fQg] = -\sum_x \pi(x)f(x)[Qg](x) = -\sum_{x,y} \pi(x)f(x)Q(x,y)g(y),$$

where, as usual, $Q = P - I$. Note, in particular, that the above expression may not be equal to

$$\frac{1}{2}\sum_{x,y} \pi(x)P(x,y)(f(y) - f(x))(g(y) - g(x)) \tag{9.4}$$

when $f \neq g$ and P is not time reversible.

Exercise 9.4. Show that (9.4) is equal to $\mathcal{E}_P(f,g)$ when either $f = g$ or P is time reversible, and provide a counterexample to the equivalence in general.

The proof of Lemma 9.2 follows a preparatory lemma.

Lemma 9.5.

$$\mathcal{E}_P(f^{q-1}, f) \geq \frac{2}{q}\mathcal{E}_P(f^{q/2}, f^{q/2}),$$

for all non-negative functions f, and all $q \geq 2$.

Proof. The proofs in this section are largely based on Guionnet and Zegarlinski [41], but the calculation is modified to avoid their assumption that P is time-reversible. In order to achieve this, we have to give away a factor of 2 in the rate of convergence. The possibility of proving Lemma 9.2 without assuming time-reversibility was noted by Diaconis and Saloff-Coste [20, Thm 3.5], who credit Bakry as their source.

First note the inequality

$$z^q - qz + (q-1) \ge (z^{q/2}-1)^2, \quad \text{for all } q \ge 2 \text{ and } z \ge 0. \tag{9.5}$$

To see this, write $h(z) := z^q - qz + (q-1) - (z^{q/2}-1)^2$, and note that $h(1) = 0$, $h'(1) = 0$, and $h''(z) \ge 0$ for all $z \ge 0$ (provided $q \ge 2$), where prime signifies derivative with respect to z. Then, provided $f \ge 0$ and $q \ge 2$,

$$\begin{aligned}
\mathcal{E}_P(f^{q-1}, f) &= -\mathbb{E}_\pi[f^{q-1}Qf] \\
&= \sum_{x,y} \pi(x) f(x)^{q-1} \big(I(x,y) - P(x,y)\big) f(y) \\
&= \frac{q-1}{q} \sum_x \pi(x) f(x)^q + \frac{1}{q} \sum_y \pi(y) f(y)^q \\
&\qquad - \sum_{x,y} \pi(x) P(x,y) f(x)^{q-1} f(y) \\
&= \sum_{x,y} \pi(x) P(x,y) \left[\frac{q-1}{q} f(x)^q + \frac{1}{q} f(y)^q - f(x)^{q-1} f(y) \right] \\
&\ge \frac{1}{q} \sum_{x,y} \pi(x) P(x,y) \big[f(x)^{q/2} - f(y)^{q/2} \big]^2 \qquad (9.6)\\
&= \frac{2}{q}\, \mathcal{E}_P(f^{q/2}, f^{q/2}),
\end{aligned}$$

where inequality (9.6) uses (9.5). □

Proof of Lemma 9.2. With the groundwork out of the way, we are just left with a calculation akin to that in §5.5. Since $\ln z$ is an monotone increasing function, it is enough to show

$$\frac{d}{dt} \ln \|f_t\|_{\pi,q(t)} \le 0.$$

So with $q = q(t) = 1 + e^{2\alpha t}$,

$$\begin{aligned}
\frac{d}{dt} \ln \|f_t\|_{\pi,q} &= \frac{d}{dt} \left[\frac{1}{q} \ln(\mathbb{E}_\pi f_t^q) \right] \\
&= -\frac{q'}{q^2} \ln(\mathbb{E}_\pi f_t^q) + \frac{1}{q\, \mathbb{E}_\pi f_t^q}\, \mathbb{E}_\pi \left[f_t^q \left(q' \ln f_t + q \frac{f_t'}{f_t} \right) \right] \\
&= \frac{1}{\mathbb{E}_\pi f_t^q} \left\{ -\frac{q'}{q^2} (\mathbb{E}_\pi f_t^q) \ln(\mathbb{E}_\pi f_t^q) + \frac{q'}{q}\, \mathbb{E}_\pi[f_t^q \ln f_t] + \mathbb{E}_\pi[f_t^{q-1} Q f_t] \right\} \\
&= \frac{1}{\mathbb{E}_\pi f_t^q} \left\{ \frac{q'}{q}\, \mathbb{E}_\pi \left[f_t^q \ln \frac{f_t}{(\mathbb{E}_\pi f_t^q)^{1/q}} \right] - \mathcal{E}_P(f_t^{q-1}, f_t) \right\} \\
&\le \frac{1}{\mathbb{E}_\pi f_t^q} \left\{ 2\alpha\, \mathbb{E}_\pi \left[f_t^q \ln \frac{f_t}{\|f_t\|_{\pi,q}} \right] - \frac{2}{q}\, \mathcal{E}_P(f_t^{q/2}, f_t^{q/2}) \right\} \qquad (9.7)\\
&\le 0, \qquad (9.8)
\end{aligned}$$

where inequality (9.7) uses Lemma 9.5 and the fact that $q' \le 2\alpha q$, and (9.8) is from (9.2). □

9.3 Mixing

Remark 9.3, although couched in informal terms, strongly suggests that hypercontractivity might be the key to obtaining bounds on mixing time with much reduced dependence on the distribution of the initial state. We now make that idea precise.

Theorem 9.6. *Suppose (Ω, P) is an ergodic MC satisfying the logarithmic Sobolev inequality (9.1) with constant α. Then, for any $\varepsilon > 0$,*

$$\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \le \varepsilon,$$

whenever $t \ge \alpha^{-1}[\ln\ln\pi(x)^{-1} + 2\ln\varepsilon^{-1} + \ln 4]$. *(To avoid pathologies, interpret* $\ln\ln\pi(x)^{-1}$ *as zero when* $\pi(x) > e^{-1}$.*)*

Proof. Let $A \subset \Omega$ be arbitrary and define $f : \Omega \to \mathbb{R}$ to be the characteristic function of A. Recall that λ denotes spectral gap. Then, from §5.5,

$$\operatorname{Var}_\pi f_{t_1} \le e^{-2\lambda t_1} \operatorname{Var}_\pi f \le \tfrac{1}{4} e^{-2\lambda t_1} = \tfrac{1}{4}\varepsilon^2$$

where $t_1 = \lambda^{-1} \ln \varepsilon^{-1}$. It follows that

$$\|f_{t_1}\|_{\pi,2}^2 = (\mathbb{E}\, f_{t_1})^2 + \operatorname{Var}_\pi f_{t_1} \le \pi(A)^2 + \tfrac{1}{4}\varepsilon^2,$$

and hence

$$\|f_{t_1}\|_{\pi,2} \le \pi(A) + \tfrac{1}{2}\varepsilon.$$

Then, by Lemma 9.2

$$\|f_t\|_{\pi,q(t_2)} \le \pi(A) + \tfrac{1}{2}\varepsilon, \tag{9.9}$$

for any $t_2 \ge 0$ and $t = t_1 + t_2$. Set $t_2 = \frac{1}{2}\alpha^{-1}[\ln\ln\pi(x)^{-1} + \ln\varepsilon^{-1} + \ln 2]$. (We need $t_2 \ge 0$, so interpret $\ln\ln\pi(x)^{-1}$ as zero when $\pi(x) > e^{-1}$.) Then

$$\pi(x)^{1/q(t_2)} \ge \pi(x)^{\exp(-2\alpha t_2)} = e^{-\varepsilon/2} \ge 1 - \tfrac{1}{2}\varepsilon,$$

and hence

$$\|f_t\|_{\pi,q(t_2)} \ge \left[\pi(x) f_t(x)^{q(t_2)}\right]^{1/q(t_2)} \ge (1 - \tfrac{1}{2}\varepsilon) f_t(x) \ge f_t(x) - \tfrac{1}{2}\varepsilon. \tag{9.10}$$

Combining (9.9) and (9.10) yields $P^t(x, A) = f_t(x) \le \pi(A) + \varepsilon$. But A is arbitrary, so $\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \le \varepsilon$. Finally, observe that

$$t = t_1 + t_2 = \frac{1}{2\alpha}\Big[\ln\ln\pi(x)^{-1} + 2\ln\varepsilon^{-1} + \ln 2\Big],$$

where we have used Theorem 9.1 to eliminate λ in favour of α. □

Remark 9.7. Comparing Theorem 9.6 against Corollary 5.9 we appreciate the potential gain from using α in place of λ. Recall that the size of the state space, and hence $\pi(x)^{-1}$, is typically exponential in some reasonable measure of instance size.

9.4 The cube (again)

The analysis of random walk on the cube from Chapter 8 may readily be adapted from spectral gap to logarithmic Sobolev constant. This will lead directly to our first application of Theorem 9.6. A move convincing application will be provided by the bases-exchange walk. This section and the next is a reworking of Jerrum and Son [47].

For the time being, we'll take (Ω, P) to be an arbitrary time-reversible finite-state MC (Ω, P), and only later specialise it to the random walk on the cube. As in §8.1 we suppose a partition of the state space $\Omega = \Omega_0 \cup \Omega_1$ is given. For convenience we repeat here the formula expressing the decomposition of Dirichlet form:

$$\mathcal{E}_P(f,f) = \pi(\Omega_0)\mathcal{E}_{P_0}(f,f) + \pi(\Omega_1)\mathcal{E}_{P_1}(f,f) + \mathcal{C}, \tag{9.11}$$

where

$$\mathcal{E}_{P_b}(f,f) = \frac{1}{2}\sum_{x,y\in\Omega_b} \pi_b(x)P(x,y)(f(x)-f(y))^2, \qquad \text{for } b = 0,1$$

and

$$\mathcal{C} = \sum_{x\in\Omega_0, y\in\Omega_1} \pi(x)P(x,y)(f(x)-f(y))^2.$$

To proceed, we need an analogue of (8.1) (decomposition of variance) for the entropy-like quantity $\mathcal{L}_\pi(f)$. It is the following:

$$\mathcal{L}_\pi(f) = \pi(\Omega_0)\mathcal{L}_{\pi_0}(f) + \pi(\Omega_1)\mathcal{L}_{\pi_1}(f) + \mathcal{L}_\pi(\bar f\,), \tag{9.12}$$

where

$$\mathcal{L}_{\pi_b}(f) = \mathbb{E}_{\pi_b}\left[f^2\big(\ln f^2 - \ln(\mathbb{E}_{\pi_b} f^2)\big)\right]$$

and

$$\mathcal{L}_\pi(\bar f) = \sum_{b=0,1} \pi(\Omega_b)\big[(\mathbb{E}_{\pi_b} f^2)\big(\ln(\mathbb{E}_{\pi_b} f^2) - \ln(\mathbb{E}_\pi f^2)\big)\big]. \tag{9.13}$$

The use of the notation $\mathcal{L}_\pi(\bar f)$ for the expression on the right hand side of (9.13) is justified, provided we interpret $\bar f : \Omega \to \mathbb{R}^+$ as the function that is constant $\sqrt{\mathbb{E}_{\pi_b} f^2}$ on Ω_b, for $b = 0,1$.

Exercise 9.8. Verify identity (9.12). (The calculation is given at end of chapter.)

As in Chapter 8, we aim to exploit (9.11) and (9.12) to synthesise an inequality of the form $\mathcal{E}_P(f,f) \ge \alpha\,\mathcal{L}_\pi(f)$ from ones of the form $\mathcal{E}_{P_b}(f,f) \ge \alpha_b\,\mathcal{L}_{\pi_b}(f)$ and $\mathcal{C} \ge \bar\alpha\,\mathcal{L}_\pi(\bar f\,)$. Inequalities $\mathcal{E}_{P_b}(f,f) \ge \alpha_b\,\mathcal{L}_{\pi_b}(f)$ will clearly come from the inductive hypothesis, exactly as before. The derivation of $\mathcal{C} \ge \bar\alpha\,\mathcal{L}_\pi(\bar f\,)$ is by way of algebraic manipulation, similar in spirit to that used in Chapter 8, but of greater complexity. This increase in calculational complexity represents the main downside in using the logarithmic Sobolev constant.

In the following lemma, we take the first step in relating $\mathcal{C}$ to $\mathcal{L}_\pi(\bar f\,)$.

Lemma 9.9. *Let r and s be positive numbers with $r + s = 1$. Then*

$$r\xi^2 \ln\frac{\xi^2}{r\xi^2 + s\eta^2} + s\eta^2 \ln\frac{\eta^2}{r\xi^2 + s\eta^2} \le (\xi - \eta)^2,$$

for all $\xi, \eta \in \mathbb{R}$.

Proof. Applying the inequality $\ln a \le a - 1$, which is valid for all $a > 0$:

$$\begin{aligned}
r\xi^2 \ln \frac{\xi^2}{r\xi^2 + s\eta^2} + s\eta^2 \ln \frac{\eta^2}{r\xi^2 + s\eta^2} &\le r\xi^2 \frac{s(\xi^2 - \eta^2)}{r\xi^2 + s\eta^2} + s\eta^2 \frac{r(\eta^2 - \xi^2)}{r\xi^2 + s\eta^2} \\
&= \frac{rs(\xi^2 - \eta^2)^2}{r\xi^2 + s\eta^2} \\
&= \frac{rs(\xi + \eta)^2}{r\xi^2 + s\eta^2} (\xi - \eta)^2 \\
&\le (\xi - \eta)^2.
\end{aligned}$$

To verify the final inequality, first note that by scaling one may assume that $\xi + \eta = 1$; it is then easy to see (by calculus) that the extremal case is when $\xi = s$ and $\eta = r$. □

Corollary 9.10. *With $\mathcal{L}_\pi(\bar{f}\,)$ defined as in (9.13),*

$$\mathcal{L}_\pi(\bar{f}\,) \le \left(\sqrt{\mathbb{E}_{\pi_0} f^2} - \sqrt{\mathbb{E}_{\pi_1} f^2}\,\right)^2.$$

Remark 9.11. In view of our interpretation of $\bar{f}$, the right hand side of the inequality appearing in Corollary 9.10 may be written $\big(\bar{f}(\Omega_0) - \bar{f}(\Omega_1)\big)^2$. In other words, Corollary 9.10 may be regarded as providing a logarithmic Sobolev inequality for a two-state MC. In is natural to ask what is the optimal constant c such that

$$c\,\mathcal{L}_\pi(\bar{f}\,) \le \left(\sqrt{\mathbb{E}_{\pi_0} f^2} - \sqrt{\mathbb{E}_{\pi_1} f^2}\,\right)^2?$$

The question has been answered by Diaconis and Saloff-Coste [20, Theorem A.2], though it proves a surprisingly hard nut: Diaconis and Saloff-Coste refer to its resolution as "a tedious calculus exercise".

Given the crude approximations used in the proof of Lemma 9.9, we would expect our estimate $c = 1$ to be a long way off, and indeed it is when either $r = \pi(\Omega_0)$ or $s = \pi(\Omega_1)$ is close to zero. Nevertheless, when $r = s = \frac{1}{2}$, we lose only a factor 2. Fortunately, in our applications, little is gained by using more refined estimates for c. Better, then, to keep things simple!

Recall the random walk on the n-dimensional cube from the beginning of §8.1. Our partition of the state space in this instance is the natural one, namely $\Omega_b = \{x = x_0x_1 \dots x_{n-1} \in \Omega : x_0 = b\}$. Corollary 9.10 puts us neatly back on the track of our earlier calculation, where our goal was to bound the spectral gap.

Consider the r.v. $(G_0, G_1) \in \mathbb{R}^2$ defined by the following trial: select $z \in \{0,1\}^{n-1}$ u.a.r.; then let $(G_0, G_1) = (f(0z)^2, f(1z)^2) \in \mathbb{R}^2$. (Recall that bz denotes the element of Ω_b obtained by prefixing z by the bit b.) Then, using $\mathbb{E}_z$ to denote expectations with respect to a uniformly selected $z \in \{0,1\}^{n-1}$,

$$
\begin{aligned}
\mathcal{L}_\pi(\bar f\,) &\le \bigl(\sqrt{\mathbb{E}_{\pi_0} f^2} - \sqrt{\mathbb{E}_{\pi_1} f^2}\,\bigr)^2 && \text{from Cor. 9.10}\\
&= \bigl(\sqrt{\mathbb{E}_z G_0} - \sqrt{\mathbb{E}_z G_1}\,\bigr)^2\\
&\le \mathbb{E}_z\Bigl[\bigl(\sqrt{G_0} - \sqrt{G_1}\,\bigr)^2\Bigr] \qquad (9.14)\\
&= 2\sum_{z\in\{0,1\}^{n-1}} \pi(0z)\bigl(f(0z)-f(1z)\bigr)^2\\
&= \frac{2}{p}\sum_{z\in\{0,1\}^{n-1}} \pi(0z)P(0z,1z)\bigl(f(0z)-f(1z)\bigr)^2\\
&= \frac{2}{p}\,\mathcal{C},
\end{aligned}
$$

where (9.14) is by Lemma 8.1 (Jensen's inequality), noting the the function $(\mathbb{R}^+)^2 \to \mathbb{R}^+$ defined by $(\xi,\eta) \mapsto (\sqrt{\xi} - \sqrt{\eta}\,)^2$ is convex. Thus, by the same inductive argument as before $\alpha_{n,p} \ge p/2$, where $\alpha_{n,p}$ denotes the logarithmic Sobolev constant of the n-dimensional cube with constant transition probability p.

Remark 9.12. Where did we lose a factor 4 relative to the spectral gap calculation? A factor of 2 was lost to the sloppy estimate in Lemma 9.9. The loss of the other factor of 2 must, by Theorem 9.1, be inevitable.

Note that, by Theorem 9.6, our logarithmic Sobolev constant translates to an $O\bigl(n(\log n + \log \varepsilon^{-1})\bigr)$ upper bound on mixing time for the random walk on the n-dimensional cube.

9.5 The bases-exchange walk (again)

A convenient feature of the cube, as regards our analysis, is that transitions from Ω_0 to Ω_1 support a perfect matching. We saw, in the context of the spectral gap lower bound of Chapter 8, that it is enough for our purposes that the transitions support a *fractional* matching. The same is true here.

Recall the bases-exchange random walk from §8.3. From Lemma 8.12, we know that the transitions from Ω_0 to Ω_1 support a fractional matching $w : \Omega_0 \times \Omega_1 \to [0,1]$. As before, we regard $(\Omega_0 \times \Omega_1, w)$ as a probability space.

Let $(G_0, G_1) \in \mathbb{R}^2$ be the r.v. defined on $(\Omega_0 \times \Omega_1, w)$ as follows: select $(x,y) \in \Omega_0 \times \Omega_1$ according to the distribution $w(\cdot,\cdot)$ and return $(G_0,G_1) = (f(x)^2, f(y)^2)$. Then, using

$\mathbb{E}_w$ to denote expectations with respect to the sample space just described,

$$\begin{aligned}
\mathcal{L}_\pi(\bar{f}\,) &\le \left(\sqrt{\mathbb{E}_{\pi_0} f^2} - \sqrt{\mathbb{E}_{\pi_1} f^2}\,\right)^2 \\
&= \left(\sqrt{\mathbb{E}_w\, G_0} - \sqrt{\mathbb{E}_w\, G_1}\,\right)^2 \\
&\le \mathbb{E}_w\left[\left(\sqrt{G_0} - \sqrt{G_1}\,\right)^2\right] \\
&= \sum_{(x,y)\in\Omega_0\times\Omega_1} w(x,y)\big(f(x)-f(y)\big)^2 \\
&\le \sum_{(x,y):w(x,y)>0} \frac{\pi(x)}{\pi(\Omega_0)}\big(f(x)-f(y)\big)^2 \\
&\le \frac{1}{p\,\pi(\Omega_0)} \sum_{(x,y)\in\Omega_0\times\Omega_1} \pi(x)P(x,y)\big(f(x)-f(y)\big)^2 \\
&\le \frac{2}{p}\,\mathcal{C},
\end{aligned}$$

where we have assumed, by symmetry, that $\pi(\Omega_0) \ge \pi(\Omega_1)$ and hence $\pi(\Omega_0) \ge \frac{1}{2}$.

Exactly the same inductive argument as in the case of the cube yields $p/2$ as the logarithmic Sobolev constant for the bases-exchange walk.

Example 9.13. Consider again the walk on spanning trees of a graph described in Example 8.19. Applying Theorem 9.6 in place of 5.9, improves our bound on mixing time to from $O(mn^2 \log m)$ to $O(mn \log n)$.

Exercise 9.14. By exhibiting a suitable graph, show that the bound in Example 9.13 is of the correct order of magnitude, at least in some circumstances.

Remark 9.15. What we have done in this chapter can be viewed as a application of a more general "decomposition" approach to the analysis of MCs apparently introduced by Caracciolo, Pelissetto and Sokal [17], and exploited by authors such as Madras, Martin and Randall [61, 59]. See Jerrum, Son, Tetali and Vigoda [48] for a general treatment of decomposition along the lines of this chapter and the previous one.

9.6 An alternative point of view

In this section we explore an alternative approach to relating the logarithmic Sobolev constant α to mixing time. The idea is to measure closeness to stationarity in terms of the "Kullback-Leibler divergence", and show that convergence in this sense is exponential, at a rate determined by α.

First, another inequality in the same spirit as Lemma 9.5.

Lemma 9.16. *$\mathcal{E}_P(f, \ln f) \ge \mathcal{E}_P(\sqrt{f}, \sqrt{f}\,)$, and hence $\mathcal{E}_P(\ln f, f) \ge \mathcal{E}_P(\sqrt{f}, \sqrt{f}\,)$, for any $f : \Omega \to \mathbb{R}^+$.*

Proof. The key to the proof is the inequality

$$a^2(\ln a - \ln b) \ge a(a-b), \tag{9.15}$$

which is valid for all $a, b > 0$. (By homogeneity it is enough to verify (9.15) in the case $a = 1$, when it reduces to the well known $\ln b \le b - 1$.) The result now follows from the following sequence of inequalities:

$$\begin{aligned}
\mathcal{E}_P(f, \ln f) &= -\sum_{x,y} \pi(x) f(x) Q(x,y) \ln f(y) \\
&= \sum_x \pi(x) f(x) \Big[\ln f(x) - \sum_y P(x,y) \ln f(y) \Big] \\
&= 2 \sum_x \pi(x) f(x) \Big[\ln \sqrt{f(x)} - \sum_y P(x,y) \ln \sqrt{f(y)} \,\Big] \\
&\ge 2 \sum_x \pi(x) f(x) \Big[\ln \sqrt{f(x)} - \ln \Big\{ \sum_y P(x,y) \sqrt{f(y)} \,\Big\} \Big] \qquad (9.16) \\
&\ge 2 \sum_{x,y} \pi(x) \sqrt{f(x)} \Big[\sqrt{f(x)} - \sum_y P(x,y) \sqrt{f(y)} \,\Big] \qquad (9.17) \\
&= -2 \sum_{x,y} \pi(x) \sqrt{f(x)}\, Q(x,y) \sqrt{f(y)} \\
&= 2\mathcal{E}_P(\sqrt{f}, \sqrt{f}\,),
\end{aligned}$$

where (9.16) is Jensen's inequality (Lemma 8.1), and (9.17) uses inequality (9.15) with $a = \sqrt{f(x)}$ and $b = \sum_y P(x,y)\sqrt{f(y)}$.

To see that the inequality holds with f and $\ln f$ reversed, consider the time reversal P^* of P, defined by

$$\pi(x) P^*(x,y) = \pi(y) P(y,x), \quad \text{fall all } x, y \in \Omega.$$

Then

$$\mathcal{E}_P(\ln f, f) = \mathcal{E}_{P^*}(f, \ln f) \ge \mathcal{E}_{P^*}(\sqrt{f}, \sqrt{f}\,) = \mathcal{E}_P(\sqrt{f}, \sqrt{f}\,).$$

□

For probability distributions σ and π on Ω, define the *Kullback-Leibler divergence* of σ from π by

$$D(\sigma \| \pi) = \mathcal{L}_\pi\left(\sqrt{\frac{\sigma}{\pi}}\right) = \sum_{x \in \Omega} \sigma(x) \ln \frac{\sigma(x)}{\pi(x)}. \qquad (9.18)$$

The word "divergence" and the curious but conventional notation is supposed to emphasise the fact that $D(\cdot \,\|\, \cdot)$ is not a metric. (It is not symmetric, for one thing.)

Remark 9.17. In interpreting definition (9.18) we use the reasonable convention $0 \ln 0 = 0$. Since we only deal with ergodic MCs, we do not have to contemplate the possibility that $\pi(x) = 0$ for some $x \in \Omega$.

Exercise 9.18. Verify that $D(\sigma \| \tau)$ is non-negative, and that $D(\sigma \| \tau) = 0$ implies $\sigma = \tau$.

Denote by $\pi_t = \pi_0 P^t : \Omega \to [0,1]$ the distribution of X_t given that the initial distribution (that of X_0) is π_0. In long-hand,

$$\pi_t(x) = \sum_{y \in \Omega} \pi_0(y) P^t(y,x).$$

Note that $\frac{d}{dt}\pi_t = \pi_t Q$, where, as usual, $Q = P - I$ (c.f. §5.5). The alternative approach to bounding mixing time rests on exponential decay of Kullback-Leibler divergence.

Theorem 9.19. *$\frac{d}{dt}D(\pi_t\|\pi) \le -2\alpha D(\pi_t\|\pi)$, and hence $D(\pi_t\|\pi) \le e^{-2\alpha t}D(\pi_0\|\pi)$.*

Proof.

$$\begin{aligned}
\frac{d}{dt}D(\pi_t\|\pi) &= \frac{d}{dt}\sum_x \pi_t(x)\ln\frac{\pi_t(x)}{\pi(x)} \\
&= \sum_x [\pi_t Q](x)\ln\frac{\pi_t(x)}{\pi(x)} + \sum_x [\pi_t Q](x) \\
&= \sum_x [\pi_t Q](x)\ln\frac{\pi_t(x)}{\pi(x)} \\
&= \sum_{x,y} \pi(x)\ln\frac{\pi_t(x)}{\pi(x)}\, Q(x,y)\frac{\pi_t(y)}{\pi(y)} \\
&= -\mathcal{E}_P\Big(\ln\frac{\pi_t}{\pi}, \frac{\pi_t}{\pi}\Big). \qquad (9.19)
\end{aligned}$$

At this point we might decide to continue by defining a modified logarithmic Sobolev constant based on the Dirichlet form (9.19) in place of the usual one. (See Bobkov and Tetali [7].) Instead, we'll use Lemma 9.16 to bring us onto a more familiar path. Picking up from (9.19),

$$\begin{aligned}
\frac{d}{dt}D(\pi_t\|\pi) &= -\mathcal{E}_P\Big(\ln\frac{\pi_t}{\pi}, \frac{\pi_t}{\pi}\Big) \\
&\le -2\mathcal{E}_P\Big(\sqrt{\frac{\pi_t}{\pi}}, \sqrt{\frac{\pi_t}{\pi}}\Big) && \text{by Lemma 9.16} \\
&\le -2\alpha\,\mathcal{L}_\pi\Big(\sqrt{\frac{\pi_t}{\pi}}\Big) \\
&= -2\alpha\sum_x \pi(x)\frac{\pi_t(x)}{\pi(x)}\ln\frac{\pi_t(x)}{\pi(x)} \\
&= -2\alpha\, D(\pi_t\|\pi).
\end{aligned}$$

□

Suppose we start at a fixed state $X_0 = x$, so that π_0 is the distribution with all its mass at the state x. Then $D(\pi_0\|\pi) = \ln(\pi(x)^{-1})$. This is promising: compared to the decay of variance argument in §5.5, this relatively small initial value provides us with a head start. However, it is not immediately clear how Kullback-Leibler divergence relates to our familiar total variation distance. Fortunately, the two are tightly related (in the direction that concerns us here at any rate) by *Pinsker's inequality*:

$$2\|\sigma - \pi\|_{\mathrm{TV}}^2 \le D(\sigma\|\pi). \qquad (9.20)$$

A proof of Pinsker's inequality may be found in the appendix to this chapter (§9.7). (If you want to try to prove Pinsker's inequality for yourself at this point, be warned that it is surprisingly tricky!)

Putting the pieces together,

$$\|\pi_t - \pi\|_{\mathrm{TV}}^2 \le \frac{1}{2} D(\pi_t\|\pi) \le \frac{1}{2} e^{-2\alpha t} D(\pi_0\|\pi) = \frac{1}{2} e^{-2\alpha t} \ln \pi(x)^{-1}.$$

Thus we are assured that $\|\pi_t - \pi\|_{\mathrm{TV}}^2 \le \varepsilon$ provided

$$t \ge \frac{1}{2\alpha}\big[\ln\ln \pi(x)^{-1} + 2\ln \varepsilon^{-1} - \ln 2\big],$$

recovering Theorem 9.6.

As a proof of Theorem 9.6, the approach taken in this section is probably a little smoother than that of §9.3. For one thing, it avoids the two-stage argument of §9.3 which requires the ℓ_2-norm to be brought under control before the norm itself is sharpened. However, hypercontractivity is stronger than exponential convergence of Kullback-Leibler divergence, implying, for example, convergence in ℓ_2-norm and not just in total variation distance (ℓ_1-norm). In fact, the connection between the logarithmic Sobolev constant and convergence in ℓ_2-norm is surprisingly tight: refer to Diaconis and Saloff-Coste [20, Cor. 3.11] for details.

9.7 Appendix

Proof of identity (9.12). By appropriately scaling the function f, it is enough to establish (9.12) when $\mathbb{E}_\pi f^2 = 1$. With this simplification,

$$\mathcal{L}_\pi(f) = \mathbb{E}_\pi[f^2 \ln f^2] = \sum_{b=0,1} \pi(\Omega_b)\, \mathbb{E}_{\pi_b}[f^2 \ln f^2]$$

and

$$\mathcal{L}_\pi(\bar f\,) = \sum_{b=0,1} \pi(\Omega_b)(\mathbb{E}_{\pi_b} f^2) \ln(\mathbb{E}_{\pi_b} f^2)$$

Subtracting,

$$\begin{aligned}\mathcal{L}_\pi(f) - \mathcal{L}_\pi(\bar f\,) &= \sum_{b=0,1} \pi(\Omega_b)\, \mathbb{E}_{\pi_b}\big[f^2(\ln f^2 - \ln(\mathbb{E}_{\pi_b} f^2)\big] \\ &= \sum_{b=0,1} \pi(\Omega_b) \mathcal{L}_{\pi_b}(f),\end{aligned}$$

as required. □

Proof of Pinsker's inequality (9.20). Our starting point is the inequality

$$u \ln u - u + 1 \ge 0, \qquad \text{for all } u > 0, \tag{9.21}$$

whose validity is easy to check. From there we bootstrap to the inequality

$$3(u-1)^2 \le (2u+4)(u \ln u - u + 1), \qquad \text{for all } u > 0. \tag{9.22}$$

To verify (9.22), define $h(u) = (2u+4)(u \ln u - u + 1) - 3(u-1)^2$, and observe that $h(1) = h'(1) = 0$, and $h''(u) = 4u^{-1}(u \ln u - u + 1) \ge 0$, where we have used (9.21). It follows that $h(u) \ge 0$ for all $u > 0$.

Pinsker's inequality itself follows from the following sequence of (in)equalities, where $u(x) = \sigma(x)/\pi(x)$:

$$
\begin{aligned}
\|\sigma - \pi\|_{\mathrm{TV}}^2 &= \frac{1}{4}\left[\sum_x \big|\sigma(x) - \pi(x)\big|\right]^2 \\
&= \frac{1}{2}\left[\sum_x \pi(x)\big|u(x) - 1\big|\right] \\
&\le \frac{1}{12}\left[\sum_x \pi(x)\sqrt{2u(x)+4}\,\sqrt{u(x)\ln u(x) - u(x) + 1}\right]^2 \qquad (9.23) \\
&\le \frac{1}{12}\left[\sum_x \pi(x)(2u(x)+4)\right]\left[\sum_x \pi(x)\big(u(x)\ln u(x) - u(x) + 1\big)\right] \qquad (9.24) \\
&= \frac{1}{2}\sum_x \pi(x)u(x)\ln u(x) \\
&= \frac{1}{2}D(\sigma\|\pi).
\end{aligned}
$$

Here, inequality (9.23) is from (9.22), and inequality (9.24) is Cauchy-Schwarz. $\square$