Abstract

We establish universal modified log-Sobolev inequalities for reversible Markov chains on the boolean lattice \( \{0, 1\}^n \), under the only assumption that the invariant law \( \pi \) satisfies a form of negative dependence known as the \textit{stochastic covering property}. This condition is strictly weaker than the strong Rayleigh property, and is satisfied in particular by all determinantal measures, as well as the uniform distribution over the set of bases of any balanced matroid. In the special case where \( \pi \) is \( k \)-homogeneous, our results imply the celebrated concentration inequality for Lipschitz functions due to Pemantle & Peres (2014). As another application, we deduce that the natural Monte-Carlo Markov Chain used to sample from \( \pi \) has mixing time at most \( kn \log \log \frac{1}{\pi(x)} \) when initialized in state \( x \). To the best of our knowledge, this is the first work relating negative dependence and modified log-Sobolev inequalities.

Contents

1 Introduction 2
  1.1 Functional inequalities for Markov chains ................................. 2
  1.2 Negative dependence for binary variables .................................. 4
  1.3 Results and implications ....................................................... 5

2 Proofs 9
  2.1 A general decomposition lemma .............................................. 9
  2.2 Proof of Theorem 3 .................................................................. 12
  2.3 Proof of Theorem 2 .................................................................. 14
  2.4 The Herbst argument ............................................................... 16
1 Introduction

1.1 Functional inequalities for Markov chains

Consider a reversible Markov generator $Q$ with respect to some probability distribution $\pi$ on a finite state space $\Omega$. In other words, $Q$ is a $\Omega \times \Omega$ matrix with non-negative off-diagonal entries, with each row summing up to 0, and satisfying the \textit{local balance} equations

$$\pi(x)Q(x,y) = \pi(y)Q(y,x), \quad (x,y) \in \Omega^2. \tag{1}$$

When $Q$ is irreducible, the underlying Markov semi-group $p_t = e^{tQ}$ \textit{mixes}, in the sense that

$$p_t(x,y) \xrightarrow{t \to \infty} \pi(y), \tag{2}$$

for all states $x, y \in \Omega$. The time-scale on which this convergence occurs is traditionally measured by the so-called \textit{mixing times}, defined for any initial state $x \in \Omega$ and any precision $\varepsilon \in (0, 1)$ by

$$t_{\text{mix}}(x; \varepsilon) := \min \{ t \geq 0 : \| p_t(x, \cdot) - \pi \|_{\text{TV}} \leq \varepsilon \}, \tag{3}$$

where $\| \mu - \nu \|_{\text{TV}} = \max_{A \subseteq \Omega} | \mu(A) - \nu(A) |$. Estimating this fundamental parameter is a fascinating theoretical problem with many practical applications, see e.g. the books \cite{29, 27}.

Throughout the paper, we will use the standard notation $E_\pi [\cdot], \text{Var}_\pi (\cdot)$ and $\text{Ent}_\pi (\cdot)$ for expectation, variance and entropy over the probability space $(\Omega, \mathcal{P}(\Omega), \pi)$, i.e.

$$E_\pi [f] := \sum_{x \in \Omega} \pi(x)f(x), \tag{4}$$

$$\text{Var}_\pi (f) := E_\pi [f^2] - E_\pi^2 [f], \tag{5}$$

$$\text{Ent}_\pi (f) := E_\pi [f \log f] - E_\pi [f] \log E_\pi [f]. \tag{6}$$

All logarithms appearing here are natural logarithms. Some of the most powerful controls on mixing times are obtained by establishing appropriate functional inequalities for the \textit{Dirichlet form}:

$$E_\pi (f, g) := -E_\pi [f Q g]. \tag{7}$$

In particular, the \textit{Poincaré constant} $\lambda(Q)$, the \textit{modified log-Sobolev constant} $\alpha(Q)$, and the \textit{log-Sobolev constant} $\rho(Q)$ are respectively defined as the largest numbers $\lambda, \alpha, \rho \geq 0$ such that

$$E_\pi (f, f) \geq \lambda \text{Var}_\pi (f); \tag{8}$$

$$E_\pi (f, \log f) \geq \alpha \text{Ent}_\pi (f); \tag{9}$$

$$E_\pi (\sqrt{f}, \sqrt{f}) \geq \rho \text{Ent}_\pi (f). \tag{10}$$
for all positive observables $f : \Omega \to (0, \infty)$. It is classical that these inequalities are increasing in strength, in the precise sense that one always has

$$2\lambda(Q) \geq \alpha(Q) \geq 4\rho(Q). \quad (11)$$

The constants $\lambda(Q), \alpha(Q), \rho(Q)$ provide fundamental controls on the underlying Markov semi-group, from transport and hypercontractivity to quantitative rates of convergence with respect to stronger metrics than total-variation distance. We refer the reader to the survey papers [18, 8] for details. Let us here only mention the following classical implications in terms of mixing times:

$$t_{\text{mix}}(x; \varepsilon) \leq \frac{1}{2\lambda(Q)} \left( \log \frac{1}{\pi(x)} + \log \frac{1}{4\varepsilon^2} \right), \quad (12)$$

$$t_{\text{mix}}(x; \varepsilon) \leq \frac{1}{\alpha(Q)} \left( \log \log \frac{1}{\pi(x)} + \log \frac{1}{2\varepsilon^2} \right), \quad (13)$$

$$t_{\text{mix}}(x; \varepsilon) \leq \frac{1}{4\rho(Q)} \left( \log \log \frac{1}{\pi(x)} + \log \frac{1}{2\varepsilon^2} \right). \quad (14)$$

Another important motivation for establishing functional inequalities is the concentration-of-measure phenomenon [10, 26]. For example, a now-classical use of the Herbst argument (see Section 2.4) yields the following sub-Gaussian tail estimate: for all $f : \Omega \to \mathbb{R}$ and $a \geq 0$,

$$\pi \left( f \geq \mathbb{E}_\pi(f) + a \right) \leq \exp \left( -\frac{\alpha(Q)a^2}{4v(f)} \right), \quad (15)$$

where $v(f)$ denotes the maximal one-sided quadratic variation of $f$ under $Q$, i.e.

$$v(f) := \max_{x \in \Omega} \left\{ \sum_{y \in \Omega} Q(x, y) [f(y) - f(x)]^2 \right\}. \quad (16)$$

In view of the above estimates, it is clear that the modified log-Sobolev inequality (MLSI) [9] is much more powerful than a Poincaré inequality [8] with the same constant. Unfortunately, establishing sharp lower bounds on the modified log-Sobolev constant of practical-purpose Markov chains remains a notoriously challenging task. The aim of the present paper is to provide simple, universal functional-analytic estimates in the case where the state space $\Omega$ is a subset of the boolean lattice $\{0, 1\}^n$, and the probability law $\pi$ satisfies an appropriate form of negative dependence known as the stochastic covering property (SCP) [32]. The precise definition will be recalled in the next section, but let us already point out that the SCP is strictly weaker than the well-known strong Rayleigh property, and that it is satisfied in particular by all determinantal measures, as well as the uniform distribution over the set of bases of any balanced matroid.

To the best of our knowledge, this is the first work relating negative dependence and modified log-Sobolev inequalities. Among other consequences, our results imply the celebrated sub-Gaussian
concentration estimate for Lipschitz functions due to Pemantle & Peres (2014) \[32\]. In addition, we
develop a new general framework for proving discrete functional inequalities using couplings, which
should be applicable to many other settings. Since the first appearance of our paper online, we have
used it in \[21\] to confirm a prediction of Caputo and Posta regarding the validity of the dimension-
free MLSI for the mean-field zero-range process \[13\] with increasing rates. Also, a matrix version
of our strategy was recently developed in \[7\] to establish matrix Poincaré inequalities, see also
\[24\]. Finally, the recent preprint \[25\] uses our decomposition method to analyze a Markov chain on
trees, while the even more recent paper \[1\] builds on our framework to obtain refined concentration
inequalities in many interesting examples.

1.2 Negative dependence for binary variables

Among the plethora of possible notions of negative dependence for binary variables \(X_1, \ldots, X_n\)
(see, e.g., the survey \[31\]), the most powerful one is arguably the strong Rayleigh property (SRP)
introduced by Borcea, Branden and Liggett \[9\]: a probability measure \(\pi\) on \(\\{0,1\}^n\) has the SRP
if its generating polynomial

\[
(z_1, \ldots, z_n) \mapsto -\mathbb{E}_\pi \left[ \prod_{i=1}^n z_i^{X_i} \right]
\]

has no root \((z_1, \ldots, z_n) \in \mathbb{C}^n\) such that \(\text{Im}(z_i) > 0\) for all \(1 \leq i \leq n\). This fundamental property
is robust under all the natural “closure” operations – products, projections, conditioning, external
fields, symmetrization, truncations – and is satisfied by many distributions arising in a variety of
contexts. Important examples include determinantal measures, product measures conditioned on
their sum, conditional balls-and-bins samples \[12\], and measures obtained by running the exclusion
dynamics from a deterministic state \[9\].

In 2011, Pemantle and Peres \[32\] put forward a strictly weaker form of negative dependence
called the stochastic covering property (SCP), which will suffice for our purposes. Its definition
requires some notation. Let \((e_i)_{1 \leq i \leq n}\) denote the canonical \(n\)-dimensional basis. For \(x, y \in \{0,1\}^n\),
we write \(x \triangleright y\) if \(x\) can be obtained from \(y\) by increasing at most one coordinate from 0 to 1, i.e.

\[
x = y \quad \text{or} \quad \exists i \in [n], x = y + e_i.
\]

We then “lift” this covering relation to probability measures on \(\{0,1\}^n\) in the usual way, writing \(\pi \triangleright \pi'\)
whenever there is a coupling of \(\pi\) and \(\pi'\) which is supported on the set \(\{(x,y) : x \triangleright y\}\).
For \(S \subseteq [n]\) and \(x \in \{0,1\}^S\), we abusively write \(\pi(X_{S^c} = \cdot | X_S = x)\) for the conditional law of
\((X_i : i \in [n] \setminus S)\) given \(\forall i \in S, X_i = x_i\). Finally, we say that \(\pi\) satisfies the SCP if the implication

\[
x \triangleright y \implies \pi(X_{S^c} = \cdot | X_S = y) \triangleright \pi(X_{S^c} = \cdot | X_S = x)
\]

4
holds for all choices of $S \subseteq [n]$ and $x, y \in \{0, 1\}^S$ such that $\pi(X_S = x), \pi(X_S = y) > 0$. This property is strictly weaker than the SRP. More precisely, the following facts are known.

1. Any measure with the SRP has the SCP [32, Proposition 2.2].

2. The uniform measure over the bases of a balanced matroid has the SCP [20, Corollary 3.3].

3. There are examples of balanced matroids for which the SRP fails [15].

The definition of a balanced matroid is not required for understanding the statements of our main results, and we simply refer the unfamiliar reader to [30, 20, 16] for more details. Let us here note that a different notion of negative dependence called strong log-concavity (SLC) has very recently been developed [13, 11, 6, 4, 5]. The latter is not comparable to the SCP, in the sense that neither of them implies the other. See Remark 1 below for more details.

### 1.3 Results and implications

Throughout this section, we consider a probability distribution $\pi$ on $\{0, 1\}^n$, and we let

$$\Omega := \{x \in \{0, 1\}^n : \pi(x) > 0\},$$

(20)

denote its support. We say that $\pi$ is $k$--homogeneous $(0 \leq k \leq n)$ if

$$\Omega \subseteq \{x \in \{0, 1\}^n : x_1 + \cdots + x_n = k\}.$$  

(21)

Our starting point is the following celebrated sub-Gaussian concentration inequality for homogeneous measures with the SCP, due to Pemantle & Peres [32]. We turn the hypercube $\{0, 1\}^n$ into a metric space by equipping it with the usual Hamming distance.

**Theorem 1** (Pemantle & Peres [32]). Let $\pi$ be any $k$–homogeneous distribution on $\{0, 1\}^n$ with the SCP, and let $f : \{0, 1\}^n \to \mathbb{R}$ be any $1$–Lipschitz function. Then, for any $a \geq 0$, we have

$$\pi(f \geq E_\pi[f] + a) \leq \exp \left( -\frac{a^2}{8k} \right).$$

(22)

In light of [15], it is natural to hope that a much deeper property underlies this remarkable concentration-of-measure phenomenon: namely, that $k$–homogeneous measures with the SCP satisfy a universal modified log-Sobolev inequality with constant $\frac{1}{k}$ w.r.t. some intrinsic local dynamics on the boolean lattice. Our first main contribution consists in showing that this is indeed the case. To formalize the notion of a “local dynamics”, let us define the relation $\sim$ on $\{0, 1\}^n$ as follows:

$$x \sim y \iff x, y \text{ differ by a flip or a swap},$$

(23)
where \( x, y \) differ by a *flip* if \( x = y \pm \epsilon_j \) for some \( 1 \leq j \leq n \), and by a *swap* if \( y = x + \epsilon_j - \epsilon_k \) for some distinct \( 1 \leq j \neq k \leq n \) (flips are only relevant for non-homogeneous measures). An *exchange walk* for \( \pi \) is a Markov generator \( Q \) on \( \Omega \) which is reversible under \( \pi \) and such that

\[
Q(x,y) > 0 \implies x \sim y \quad (x,y \in \Omega).
\]  

(24)

We let \( \Delta(Q) := \max \{-Q(x,x) : x \in \Omega\} \) denote the maximal rate at which changes occur under \( Q \), and we say that \( Q \) is *normalized* whenever \( \Delta(Q) \leq 1 \). This means that \( Q \) may be written as

\[
Q = P - I,
\]  

(25)

for some stochastic matrix \( P \), allowing us to encompass the (more standard) discrete-time setting.

With this terminology in hands, our main result asserts that every \( k \)-homogeneous measure with the SCP admits a normalized exchange walk with modified log-Sobolev constant \( \frac{1}{2k} \).

**Theorem 2** (Intrinsic modified log-Sobolev inequality for SCP measures). *If \( \pi \) is a \( k \)-homogeneous measure with the SCP, then there is a normalized exchange walk \( Q \) for \( \pi \) such that

\[
\lambda(Q), \alpha(Q) \geq \frac{1}{2k}.
\]  

(26)

Furthermore, the conclusion remains valid without the homogeneity assumption, with \( k := \frac{n}{2} \).

By virtue of (15), this immediately implies Theorem 1 albeit with the constant 8 replaced by 32. As already explained, modified log-Sobolev inequalities are of intrinsic interest beyond the sole fact that they imply sub-Gaussian concentration. For example, another notable consequence of Theorem 2 concerns sampling complexity: starting from an arbitrary initial state \( x \in \Omega \), our exchange walk will – by virtue of (13) – produce an \( \varepsilon \)-approximate sample from \( \pi \) in time

\[
t_{\text{mix}}(x; \varepsilon) \leq 2k \left( \log \log \frac{1}{\pi(x)} + \log \frac{1}{2\varepsilon^2} \right).
\]  

(27)

Here again, the conclusion remains valid without the homogeneity assumption, provided we set \( k := \frac{n}{2} \). Note that this bound – and hence our modified log-Sobolev estimate – is sharp up to a factor of 4 only, as can be seen by considering the basic case where \( \pi \) is uniform on \( \{0,1\}^n \) [19].

In practical situations however, one might be constrained to use more tractable rates than those of the exchange walk of Theorem 2. A canonical choice is the Monte Carlo Markov Chain (MCMC)

\[
Q(x,y) := \frac{1}{2kn} \min \left\{ \frac{\pi(y)}{\pi(x)}, 1 \right\} \quad \text{if } x \sim y.
\]  

(28)

Notice that only the positive entries have here been made explicit: all other off-diagonal entries should be interpreted as zero, and the diagonal adjusted so that the rows sum up to zero. The
normalization ensures that $\Delta(Q) \leq \frac{1}{2}$, so that $Q$ corresponds to a lazy stochastic matrix $P$ via (25). The strategy that we develop to establish Theorem 2 is sufficiently robust to yield sharp performance guarantees for this chain too, and many others. For any generator $Q$ on $\Omega$, define

$$ m(Q) := \min_{x, y \in \Omega, x \sim y} Q(x, y), \quad (29) $$

$$ \overline{m}(Q) := \min_{x, y \in \Omega, x \sim y} \max \{Q(x, y), Q(y, x)\}. \quad (30) $$

Note that $m(Q) \leq \overline{m}(Q)$. It turns out that these simple statistics provide universal functional-analytic estimates whenever $Q$ is reversible w.r.t. a measure with the SCP.

**Theorem 3** (Universal functional inequalities for SCP measures). Consider a probability distribution $\pi$ with the SCP, and let $Q$ be any reversible Markov generator with respect to $\pi$. Then,

$$ \lambda(Q) \geq \max \{\overline{m}(Q), 2m(Q)\}; \quad (31) $$

$$ \alpha(Q) \geq \max \{\overline{m}(Q), 4m(Q)\}; \quad (32) $$

$$ \rho(Q) \geq m(Q). \quad (33) $$

This result unifies, extends or strengthens various powerful estimates obtained over the past decades. In particular, in an inspiring work, Jerrum & Son [22] considered the special case where $\Omega$ is the set of bases of a balanced matroid of rank $k$, and $\pi$ is the uniform distribution on it. In this setting, the MCMC (28) reduces to the so-called bases-exchange walk

$$ Q(x, y) := \frac{1}{2kn} \quad \text{if} \ x \sim y. \quad (34) $$

For the latter, Jerrum & Son established the Poincaré and log-Sobolev inequalities

$$ \lambda(Q) \geq \frac{1}{kn}, \quad \rho(Q) \geq \frac{1}{4kn}. \quad (35) $$

Both estimates readily follow from Theorem 3, the second one being even improved by a factor 2.

Note that the above setting is rather specific, since $\pi$ is uniform. In a recent breakthrough, Anari, Oveis Gharan & Rezaei [2] managed to relax this constraint: under the assumption that $\pi$ is $k$–homogeneous with the SRP, they showed that the chain (28) satisfies the Poincaré inequality

$$ \lambda(Q) \geq \frac{1}{2kn}. \quad (36) $$

By virtue of (12), this immediately implies the mixing-time bound

$$ t_{\text{mix}}(x; \varepsilon) \leq kn \left( \log \frac{1}{\pi(x)} + \log \frac{1}{4\varepsilon^2} \right). \quad (37) $$
This result is again covered by Theorem 3, under the weaker condition that $\pi$ has the SCP. More importantly, Theorem 3 provides a new modified log-Sobolev inequality with the same constant, leading to a much tighter control on the convergence rate. Let us state this as a corollary.

**Corollary 1** (Modified log-Sobolev inequalities for MCMC). Consider a $k$-homogeneous measure $\pi$ on $\{0,1\}^n$ with the SCP, and let $Q$ be the MCMC defined at (28). Then,

$$\alpha(Q) \geq \frac{1}{2kn}. \quad (38)$$

In particular, for any initial state $x \in \Omega$ and any precision $\varepsilon \in (0,1)$, we have

$$t_{\text{mix}}(x; \varepsilon) \leq 2kn \left( \log \log \frac{1}{\pi(x)} + \log \frac{2}{\varepsilon^2} \right), \quad (39)$$

which offers a considerable improvement upon (37).

We note that resorting to the modified log-Sobolev constant is crucial here: when the measure $\pi$ is not uniform, the log-Sobolev constant $\rho(Q)$ can not be bounded from below by a function of $n,k$ only, as in (35). Indeed, it follows from the definitions that any normalized generator $Q$ satisfies

$$\rho(Q) \leq \min_{x \in \Omega} \left\{ \left( \log \frac{1}{\pi(x)} \right)^{-1} \right\}, \quad (40)$$

and the right-hand side can be arbitrarily small (e.g., take $\pi$ to be the law of $n$ independent Bernoulli variables with very different parameters, conditioned on their sum being equal to $k$).

**Remark 1** (Strong log-concavity). Very recently, a powerful notion of negative dependence for $k$-homogenous measures on $\{0,1\}^n$ called strong log-concavity (SLC) has been developed in two impressive series of papers by Anari, Liu, Oveis-Gharan & Vinzant [6, 4, 5, 3] and by Brändén and Huh [13, 11]. In particular, three months before the appearance of our paper online, a Poincaré inequality with constant $1/k$ was established by Anari, Liu, Oveis-Gharan & Vinzant [4] for sampling from such measures, and a month after the appearance of our paper online, this Poincaré inequality was improved to a MLSI with constant $1/k$ by Cryan, Guo & Mousa [17]. Finally, a year later, a sharp mixing time bound of $k \log k$ was established by Anari, Liu, Oveis Gharian, Vinzant, Vuong [3]. We note, however, that these two powerful (posterior) results are not directly comparable to ours, for several reasons. First, the negative dependence assumptions SLC and SCP are not comparable, in the sense that neither of them implies the other, as observed in [17]. Second, the MCMC considered in [4, 17] is actually an accelerated version of (28), in which each iteration may require more than constant time to implement in practice. Third, we recall that neither homogeneity, nor the specific form (28) is required by our methods: Theorem 3 applies to any measure $\pi$ with the SCP, and any reversible generator $Q$ w.r.t. $\pi$. 

2 Proofs

2.1 A general decomposition lemma

Our starting point is a well-known recursive strategy for establishing functional inequalities for reversible chains. The method was invented by Lu & Yau in the context of interacting particle systems [28], and developed further by Jerrum & Son [22] and Jerrum, Son, Tetali & Vigoda [23].

Let $Q$ be any reversible Markov generator with respect to some (fully-supported) probability distribution $\pi$ on a finite set $\Omega$. Consider an arbitrary partition of the state space:

$$\Omega = \bigcup_{i \in I} \Omega_i. \quad (41)$$

The projection chain induced by this partition is the Markov chain whose state space is $I$ and whose Markov generator $\hat{Q}$ is defined as follows: for all $(i, j) \in \Omega^2$,

$$\hat{Q}(i, j) := \frac{1}{\hat{\pi}(i)} \sum_{x \in \Omega_i} \sum_{y \in \Omega_j} \pi(x)Q(x, y), \quad \text{where} \quad \hat{\pi}(i) := \sum_{x \in \Omega_i} \pi(x). \quad (42)$$

Note that $\hat{\pi}$ is a probability distribution on $I$, and that it is reversible under $\hat{Q}$, i.e.

$$\hat{\pi}(i)\hat{Q}(i, j) = \hat{\pi}(j)\hat{Q}(j, i). \quad (43)$$

For each $i \in I$, the restriction chain on $\Omega_i$ is the Markov chain whose state space is $\Omega_i$ and whose Markov generator $Q_i$ is defined as follows: for any distinct states $x, y$ in $\Omega_i$,

$$Q_i(x, y) = Q(x, y), \quad (44)$$

with the diagonal entries being adjusted so that the rows of $Q_i$ sum up to zero. It is then clear that the probability measure $\pi_i$ defined on $\Omega_i$ by

$$\pi_i(x) := \frac{\pi(x)}{\hat{\pi}(i)} \quad (45)$$

is reversible under $Q_i$. Now, suppose that for each $(i, j) \in I^2$ with $\hat{Q}(i, j) > 0$, we are given a coupling $\kappa_{ij} : \Omega_i \times \Omega_j \to [0, 1]$ of the probability distributions $\pi_i$ and $\pi_j$, i.e.

$$\forall x \in \Omega_i, \quad \sum_{y \in \Omega_j} \kappa_{ij}(x, y) = \pi_i(x), \quad (46)$$

$$\forall y \in \Omega_j, \quad \sum_{x \in \Omega_i} \kappa_{ij}(x, y) = \pi_j(y). \quad (47)$$

For reasons that will soon become clear, we measure the quality of these couplings by the quantity

$$\chi := \min \left\{ \frac{\pi(x)Q(x, y)}{\hat{\pi}(i)\hat{Q}(i, j)\kappa_{ij}(x, y)} \right\}, \quad (48)$$
where the minimum runs over all \((x, y, i, j)\) such that the denominator is positive. Note that \(\chi = 0\) if 
\[
\min_{x \in i} \sum_{y \in j} Q(x, y) = 0 \text{ for some } i, j \in I.
\]
Hence such partitions should be avoided when applying Lemma 1 below.

**Lemma 1** (Recursive functional inequalities). With the above notations, we have
\[
\lambda(Q) \geq \min \left\{ \chi \lambda(\hat{Q}), \min_{i \in I} \lambda(Q_i) \right\};
\]
\[
\alpha(Q) \geq \min \left\{ \chi \alpha(\hat{Q}), \min_{i \in I} \alpha(Q_i) \right\};
\]
\[
\rho(Q) \geq \min \left\{ \chi \rho(\hat{Q}), \min_{i \in I} \rho(Q_i) \right\}.
\]

**Proof.** Let \(f : \Omega \to (0, \infty)\). The left-hand sides of (49),(50),(51) take the generic form
\[
L_{\pi}(f) := \frac{1}{2} \sum_{(x, y) \in \Omega^2} \pi(x)Q(x, y)\Psi(f(x), f(y)),
\]
with the function \(\Psi : (0, \infty)^2 \to [0, \infty)\) being given by
\[
\Psi(u, v) := \begin{cases} 
(u - v)^2 & \text{in the Poincaré case} \\
(u - v)(\log u - \log v) & \text{in the modified log-Sobolev case} \\
(\sqrt{u} - \sqrt{v})^2 & \text{in the log-Sobolev case.}
\end{cases}
\]
The quantity \(L_{\pi}(f)\) measures the average local variation of \(f\) along the transitions of the chain. We wish to compare it to the average global variation of \(f\) across the whole space, as measured by
\[
R_{\pi}(f) := \begin{cases} 
\text{Var}_{\pi}(f) & \text{in the Poincaré case} \\
\text{Ent}_{\pi}(f) & \text{in the log-Sobolev and modified log-Sobolev cases.}
\end{cases}
\]
To this end, we start by decomposing \(R_{\pi}(f), L_{\pi}(f)\) according to the partition (41). Define a projected observable \(\hat{f} : I \to (0, \infty)\) by \(\hat{f}(i) := E_{\pi_i}(f)\). It is immediate to check that
\[
R_{\pi}(f) = \sum_{i \in I} \pi_i R_{\pi_i}(f) + R_{\tilde{\pi}}(\hat{f});
\]
\[
L_{\pi}(f) = \sum_{i \in I} \pi_i L_{\pi_i}(f) + \frac{1}{2} \sum_{i \neq j} \sum_{(x, y) \in \Omega_i \times \Omega_j} \pi(x)Q(x, y)\Psi(f(x), f(y)).
\]
In light of the similarity between the right-hand sides, it is then natural to hope for a term-by-term comparison. The problematic quantity is of course the second sum on the right-hand side of (56), which we would like to replace by \(L_{\tilde{\pi}}(\hat{f})\). To this end, let us fix a pair \((i, j) \in I^2\) such that \(\hat{Q}(i, j) > 0\). Since \(\kappa_{ij}\) is a coupling of \(\pi_i\) and \(\pi_j\), we have
\[
\hat{f}(i) = \sum_{(x, y) \in \Omega_i \times \Omega_j} \kappa_{ij}(x, y)f(x),
\]
\[
\hat{f}(j) = \sum_{(x, y) \in \Omega_i \times \Omega_j} \kappa_{ij}(x, y)f(y).
\]
Thus, the pair \((\hat{f}(i), \hat{f}(j))\) is the average of the function \((x, y) \mapsto (f(x), f(y))\) under the probability distribution \(\kappa_{ij}\). The crucial observation is that the bivariate function \(\Psi\) defined at (53) is convex, as can be easily checked by verifying that the hessian matrix
\[
H := \begin{bmatrix}
\partial_{uu} \Psi & \partial_{uv} \Psi \\
\partial_{uv} \Psi & \partial_{vv} \Psi
\end{bmatrix},
\]
is positive semi-definite in each case. Therefore, Jensen’s inequality ensures that
\[
\sum_{(x, y) \in \Omega_i \times \Omega_j} \kappa_{ij}(x, y) \Psi(f(x), f(y)) \geq \Psi\left(\hat{f}(i), \hat{f}(j)\right).
\]
Multiplying through by \(\chi \hat{\pi}(i) \hat{Q}(i, j)\) and recalling the definition of \(\chi\), we deduce that
\[
\sum_{(x, y) \in \Omega_i \times \Omega_j} \pi(x) Q(x, y) \Psi(f(x), f(y)) \geq \chi \hat{\pi}(i) \hat{Q}(i, j) \Psi\left(\hat{f}(i), \hat{f}(j)\right).
\]
Although this was established under the assumption that \(\hat{Q}(i, j) > 0\), the conclusion remains trivially valid when \(\hat{Q}(i, j) = 0\). Summing over all pairs \((i, j) \in \mathcal{I}^2\) with \(i \neq j\) and plugging the resulting estimate back into (56), we arrive at
\[
\mathcal{L}_\pi(f) \geq \sum_{i \in \mathcal{I}} \hat{\pi}(i) \mathcal{L}_{\pi_i}(f) + \chi \hat{\pi}(\hat{f}).
\]
Comparing this with (55), we immediately deduce that for any \(\hat{\kappa}, \kappa_i \geq 0\),
\[
\left\{ \begin{array}{l}
\mathcal{L}_{\hat{\pi}}(\hat{f}) \geq \hat{\kappa} \mathcal{R}_{\hat{\pi}}(\hat{f}) \\
\forall i \in \mathcal{I}, \mathcal{L}_{\pi_i}(f) \geq \kappa_i \mathcal{R}_{\pi_i}(f)
\end{array} \right. \implies \mathcal{L}_\pi(f) \geq \min_{i \in \mathcal{I}} \left\{ \chi \hat{\kappa}, \min_{i \in \mathcal{I}} \kappa_i \right\} \mathcal{R}(f).
\]
Since this implication holds for all observables \(f : \Omega \to (0, \infty)\), the three claims follow.

We end this section with a crude estimate on the ratio appearing in the definition of \(\chi\).

**Lemma 2** (Crude lower-bound on \(\chi\)). We always have
\[
\frac{\pi(x) Q(x, y)}{\hat{\pi}(i) \hat{Q}(i, j) \kappa_{ij}(x, y)} \geq \max \left\{ \frac{Q(x, y)}{\hat{Q}(i, j)}, \frac{Q(y, x)}{\hat{Q}(j, i)} \right\},
\]
provided the denominator on the left-hand side is positive.

**Proof.** Since \(\kappa_{ij}\) is a coupling of \(\pi_i\) and \(\pi_j\), we have \(\kappa_{ij}(x, y) \leq \pi_i(x)\) and \(\kappa_{ij}(x, y) \leq \pi_j(y)\). Thus,
\[
\frac{\pi(x) Q(x, y)}{\hat{\pi}(i) \hat{Q}(i, j) \kappa_{ij}(x, y)} \geq \frac{Q(x, y)}{\hat{Q}(i, j)}, \quad \frac{\pi(y) Q(y, x)}{\hat{\pi}(j) \hat{Q}(j, i) \kappa_{ij}(x, y)} \geq \frac{Q(y, x)}{\hat{Q}(j, i)}.
\]
The claim follows by noting that the left-hand sides of these two lines are equal, by reversibility. \(\Box\)
2.2 Proof of Theorem 3

We are now in position to prove Theorem 3 by induction over the number of states $|\Omega|$. The claim is trivial when $|\Omega| = 1$. Now, consider a probability distribution $\pi$ on $\{0, 1\}^n$ whose support $\Omega$ has at least two elements. This means that there is an index $\ell \in \{1, \ldots, n\}$ such that the subsets

$$\Omega_0 := \{x \in \Omega : x_\ell = 0\}, \quad \Omega_1 := \{x \in \Omega : x_\ell = 1\},$$

are both non-empty. Let $Q$ be any Markov generator on $\Omega$ under which $\pi$ is reversible, and let $(Q_0, \Omega_0, \pi_0)$, $(Q_1, \Omega_1, \pi_1)$ and $(\tilde{Q}, \{0, 1\}, \tilde{\pi})$ be the restriction and projection chains induced by the partition $\Omega = \Omega_0 \cup \Omega_1$. Using the short-hands $a := \tilde{Q}(0, 1)$ and $b := \tilde{Q}(1, 0)$, we have

$$\tilde{Q} = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix}. \quad (68)$$

Functional-analytic estimates for such two-state chains can easily be found in the literature.

**Lemma 3** (Two-state chains, see, e.g. [8]). For any choice of the rates $a, b \geq 0$, we have

$$\lambda(\tilde{Q}) = a + b, \quad \alpha(\tilde{Q}) \in [a + b, 2(a + b)], \quad \rho(\tilde{Q}) = \begin{cases} \frac{a-b}{\log a - \log b} & \text{if } a \neq b \\ a & \text{if } a = b. \end{cases} \quad (69)$$

Since $|a - b| \geq \min(a, b) |\log a - \log b|$, the last estimate implies in particular $\rho(\tilde{Q}) \geq \min(a, b)$.

In order to apply Lemma 1 it now remains to construct an appropriate coupling of $\pi_0$ and $\pi_1$. This is provided by the following lemma, which is where the SCP comes into play.

**Lemma 4** (Exploiting the SCP). If $\pi$ has the SCP, there is a coupling $\kappa$ of $\pi_0, \pi_1$ supported on

$$\{(x, y) \in \Omega_0 \times \Omega_1 : x \sim y\}. \quad (70)$$

In particular, taking $\kappa_{01}$ and $\kappa_{10}$ to be $\kappa$ and its transpose, respectively, we obtain

$$\chi \geq \max \left\{ \frac{m(Q)}{\max (a, b)}, \frac{m(Q)}{\min (a, b)} \right\}. \quad (71)$$

**Proof.** For ease of notation, let us here assume that $\ell = n$. The SCP ensures that we can construct, on a common probability space, two random vectors $(X_1, \ldots, X_{n-1})$ and $(Y_1, \ldots, Y_{n-1})$ such that

1. $(X_1, \ldots, X_{n-1}, 0)$ has law $\pi_0$;
2. $(Y_1, \ldots, Y_{n-1}, 1)$ has law $\pi_1$;
3. $(X_1, \ldots, X_{n-1}) \Rightarrow (Y_1, \ldots, Y_{n-1})$ with probability 1.
By definition of $\triangleright$ and $\sim$, the property $(X_1, \ldots, X_{n-1}) \triangleright (Y_1, \ldots, Y_{n-1})$ deterministically implies
\[(X_1, \ldots, X_{n-1}, 0) \sim (Y_1, \ldots, Y_{n-1}, 1). \quad (72)\]

The joint law $\kappa$ of these two vectors is thus a coupling of $\pi_0$ and $\pi_1$ satisfying the claim. To estimate the resulting constant $\chi$, we simply invoke Lemma 2 to get
\[\chi \geq \min \left\{ \max \left\{ \frac{Q(x, y)}{a}, \frac{Q(y, x)}{b} \right\} : (x, y) \in \Omega_0 \times \Omega_1, x \sim y \right\}. \quad (73)\]

The second claim now readily follows from this and the definitions of $\underline{m}(Q), \overline{m}(Q)$. \hfill \Box

We now have all we need to complete our induction step. Using Lemmas 3 and 4, we find
\[\chi \lambda(\hat{Q}) \geq \overline{m}(Q), \quad (74)\]
\[\chi \alpha(\hat{Q}) \geq \underline{m}(Q), \quad (75)\]
\[\chi \rho(\hat{Q}) \geq \underline{m}(Q). \quad (76)\]

Applying Lemma 1 we deduce that
\[\lambda(Q) \geq \min \{ \underline{m}(Q), \lambda(Q_0), \lambda(Q_1) \}; \quad (77)\]
\[\alpha(Q) \geq \min \{ \underline{m}(Q), \alpha(Q_0), \alpha(Q_1) \}; \quad (78)\]
\[\rho(Q) \geq \min \{ \underline{m}(Q), \rho(Q_0), \rho(Q_1) \}. \quad (79)\]

Now, observe that the SCP is closed under conditioning, so that $\pi_0$ and $\pi_1$ inherit it from $\pi$. Thus, the induction hypothesis applies to the restriction chains $(\Omega_i, \pi_i, Q_i), i \in \{0, 1\}$, yielding
\[\lambda(Q_i) \geq \overline{m}(Q_i), \quad (80)\]
\[\alpha(Q_i) \geq \underline{m}(Q_i), \quad (81)\]
\[\rho(Q_i) \geq \underline{m}(Q_i). \quad (82)\]

Since we obviously have $\overline{m}(Q_i) \geq \overline{m}(Q)$ and $\underline{m}(Q_i) \geq \underline{m}(Q)$, we conclude that
\[\lambda(Q) \geq \overline{m}(Q), \quad (83)\]
\[\alpha(Q) \geq \underline{m}(Q), \quad (84)\]
\[\rho(Q) \geq \underline{m}(Q). \quad (85)\]

Finally, the slight improvement appearing in (31)-(32) simply follows from (11).
2.3 Proof of Theorem 2

We now prove Theorem 2 by induction over the dimension $n$. It will actually be slightly more convenient to change the normalization as follows. For any distribution $\pi$ on $\{0,1\}^n$ with the SCP, we will prove the existence of a exchange walk $Q$ for $\pi$ such that $\lambda(Q), \alpha(Q) \geq 1$ and

$$\Delta(Q) \leq \begin{cases} n & \text{always} \\ 2k & \text{if } \pi \text{ is } k-\text{homogeneous.} \end{cases} \quad (86)$$

The original claim is obtained by dividing all entries by $\Delta(Q)$. The base case $n = 1$ is trivial, and we henceforth assume that $n \geq 2$. Fix a coordinate $\ell \in \{1, \ldots, n\}$. We may assume that the sets

$$\Omega_0 := \{x \in \Omega: x_\ell = 0\}, \quad \Omega_1 := \{x \in \Omega: x_\ell = 1\},$$

are both non-empty: otherwise, $\pi$ can be regarded as a measure on $\{0,1\}^{n-1}$, and the claim trivially follows from the induction hypothesis. Now, consider the corresponding projection and restriction measures $\hat{\pi}, \pi_0, \pi_1$, as defined in Section 2.1. For $i \in \{0,1\}$, $\pi_i$ may be regarded as a measure on $\{0,1\}^{n-1}$, and it inherits the SCP from $\pi$. Notice that if $\pi$ is $k$-homogeneous, then so is $\pi_0$, while $\pi_1$ is $(k-1)$-homogeneous (when regarded as a measure on $\{0,1\}^{n-1}$). Thus, for $i \in \{0,1\}$, the induction hypothesis provides a exchange walk $Q_i$ for $\pi_i$ satisfying $\lambda(Q_i), \alpha(Q_i) \geq 1$ and

$$\Delta(Q_i) \leq \begin{cases} n - 1 & \text{always} \\ 2(k-i) & \text{if } \pi \text{ is } k-\text{homogeneous.} \end{cases} \quad (88)$$

On the other hand, Lemma 4 provides a coupling $\kappa$ of $\pi_0$ and $\pi_1$ supported on the set

$$\{(x,y) \in \Omega_0 \times \Omega_1: x \sim y\}. \quad (89)$$

With these ingredients in hands, we define an exchange walk $Q$ for $\pi$ as follows: for $x \neq y \in \Omega$,

$$Q(x,y) := \begin{cases} Q_0(x,y) & \text{if } (x,y) \in \Omega_0 \times \Omega_0 \\ Q_1(x,y) & \text{if } (x,y) \in \Omega_1 \times \Omega_1 \\ \frac{\hat{\pi}(0)\hat{\pi}(1)\kappa(x,y)}{\pi(x)} & \text{if } (x,y) \in \Omega_0 \times \Omega_1 \\ \frac{\hat{\pi}(0)\hat{\pi}(1)\kappa(y,x)}{\pi(x)} & \text{if } (x,y) \in \Omega_1 \times \Omega_0. \end{cases} \quad (90)$$

Here again, the diagonal is implicitly adjusted so that the rows sum up to 0. In order to estimate $\lambda(Q), \alpha(Q)$, we will now use Lemma 11 with the partition being $\Omega = \Omega_0 \cup \Omega_1$ and the couplings $\kappa_{01}$ and $\kappa_{10}$ being $\kappa$ and its transpose, respectively. By Lemma 3, the projection chain $\hat{Q}$ satisfies

$$\lambda(\hat{Q}), \alpha(\hat{Q}) \geq \hat{Q}(0,1) + \hat{Q}(1,0). \quad (91)$$
On the other hand, for any \((x,y) \in \Omega_0 \times \Omega_1\) with \(\kappa(x,y) > 0\), we have by construction,

\[
\frac{\pi(x)Q(x,y)}{\tilde{\pi}(0)Q(0,1)\kappa_{01}(x,y)} = \frac{\tilde{\pi}(1)}{Q(1,1)} = \frac{\tilde{\pi}(0)}{Q(1,0)} = \frac{\pi(y)Q(y,x)}{\tilde{\pi}(1)Q(1,0)\kappa_{10}(y,x)},
\]

where the equality in the middle is nothing more than reversibility. Consequently, we see that

\[
\chi = \frac{\tilde{\pi}(0)}{Q(1,0)} = \frac{\tilde{\pi}(1)}{Q(0,1)}.
\]

Combining this with (91), we deduce that

\[
\chi \lambda(\tilde{Q}), \chi \alpha(\tilde{Q}) \geq \tilde{\pi}(0) + \tilde{\pi}(1) = 1.
\]

Recalling that \(\lambda(Q_i), \alpha(Q_i) \geq 1\), we may finally invoke Lemma 1 to conclude that

\[
\lambda(Q), \alpha(Q) \geq 1.
\]

It now remains to estimate the diagonal entries. For \(x \in \Omega_0\), we have by construction

\[
-Q(x,x) = \sum_{y \in \Omega_0 \setminus \{x\}} Q_0(x,y) + \frac{\tilde{\pi}(0)\tilde{\pi}(1)}{\pi(x)} \sum_{y \in \Omega_1} \kappa(x,y).
\]

Using the definition of \(\Delta(Q_0)\) and the fact that the first marginal of \(\kappa\) is \(\pi_0\), we deduce that

\[
-Q(x,x) \leq \Delta(Q_0) + \tilde{\pi}(1) = \Delta(Q_0) + \mathbb{E}_x[X_\ell].
\]

Similarly, if \(x \in \Omega_1\), we have

\[
-Q(x,x) \leq \Delta(Q_1) + \tilde{\pi}(0) \leq \Delta(Q_1) + x_\ell,
\]

because \(x_\ell = 1\) in this case. Using (88), we deduce that for all \(x \in \Omega\),

\[
-Q(x,x) \leq \left\{ \begin{array}{ll}
    n & \text{always} \\
    2k + \mathbb{E}_x[X_\ell] - x_\ell & \text{if } \pi \text{ is } k-\text{homogeneous},
\end{array} \right.
\]

Unfortunately, the second line is not bounded from above by \(2k\). To fix this, our last step will consist in averaging over the choice of the coordinate \(\ell\) used to define the partition (87). We henceforth write \(Q = Q^{(\ell)}\) to indicate explicitly the dependency upon \(\ell\). For each \(\ell \in \{1, \ldots, n\}\), the construction (90) produces an exchange walk \(Q^{(\ell)}\) for \(\pi\) satisfying \(\lambda(Q^{(\ell)}), \alpha(Q^{(\ell)}) \geq 1\) and \(\Delta(Q^{(\ell)}) \leq n\). Since all these properties are preserved under convex combinations, the generator

\[
Q^* := \frac{1}{n} \sum_{\ell=1}^{n} Q^{(\ell)}
\]

automatically inherits them. Now, if \(\pi\) is \(k\)-homogeneous, then for each \(x \in \Omega\), we have

\[
-Q^*(x,x) = -\frac{1}{n} \sum_{\ell=1}^{n} Q^{(\ell)}(x,x) \leq \frac{1}{n} \sum_{\ell=1}^{n} (2k + \mathbb{E}_x[X_\ell] - x_\ell) = 2k.
\]

where the inequality follows from (99), and the equality from the definition of \(k\)-homogeneity. Thus, the generator \(Q^*\) enjoys all the desired properties, and this completes the induction step.
2.4 The Herbst argument

For completeness, we finally explicitate the Herbst argument used to turn any modified log-Sobolev inequality into a sub-Gaussian concentration estimate, as claimed at (15). We emphasize that the approach is standard, and has been used in various settings, see e.g. \([26, 10, 8]\). We also note that similar concentration results can be derived via transportation inequalities, as observed by Sammer [34, Prop. 3.1.3 and §1.4.1] and by Sammer and Tetali [33].

**Lemma 5** (Modified log-Sobolev implies Gaussian concentration). Let \(Q\) be a reversible Markov generator with respect to some probability distribution \(\pi\) on a finite set \(\Omega\). Then,

\[
\pi( f \geq \mathbb{E}_\pi[f] + a) \leq \exp\left(-\frac{\alpha(Q)a^2}{4v(f)}\right),
\]

for any observable \(f\): \(\Omega \rightarrow \mathbb{R}\) and any \(a \geq 0\), where

\[
v(f) := \max_{x \in \Omega} \left\{ \sum_{y \in \Omega} Q(x,y) [f(y) - f(x)]^2_+ \right\}.
\]

**Proof.** To lighten notations, we drop the subscript \(\pi\). For \(t \in (0, \infty)\) and \(x \in \Omega\), define

\[
F_t(x) := e^{tf(x) - ct^2},
\]

where \(c > 0\) will be adjusted later. Using the reversibility \(\pi(x)Q(x,y) = \pi(y)Q(y,x)\), we have

\[
\mathcal{E}(F_t, \log F_t) = \frac{t}{2} \sum_{(x,y) \in \Omega^2} \pi(x)Q(x,y)(F_t(x) - F_t(y))(f(x) - f(y))
\]

\[
= t \sum_{(x,y) \in \Omega^2} \pi(x)Q(x,y)F_t(x) \left(1 - e^{-t(f(x) - f(y))}\right) [f(x) - f(y)]^2_+
\]

\[
\leq t^2 \sum_{(x,y) \in \Omega^2} \pi(x)Q(x,y)F_t(x) [f(x) - f(y)]^2_+
\]

\[
\leq t^2 v(f) \mathbb{E}[F_t],
\]

thanks to the bound \(1 - e^{-u} \leq u\) and \((103)\). Recalling the definition of \(\alpha(Q)\), we deduce that

\[
\text{Ent}(F_t) \leq \frac{v(f)}{\alpha(Q)} t^2 \mathbb{E}[F_t].
\]

On the other hand, for any \(t > 0\), we easily compute

\[
\frac{d}{dt} \left\{ \log \mathbb{E}[F_t] \right\} = \frac{\text{Ent}[F_t] - ct^2 \mathbb{E}[F_t]}{t^2 \mathbb{E}[F_t]}.
\]

Choosing \(c := \frac{v(f)}{\alpha(Q)}\), we see that \(t \mapsto \frac{\log \mathbb{E}[F_t]}{t}\) is non-increasing on \((0, \infty)\). In particular,

\[
\frac{\log \mathbb{E}[F_t]}{t} \leq \lim_{h \to 0} \frac{\log \mathbb{E}[F_h]}{h} = \mathbb{E}[f],
\]

16
or equivalently, $E[e^{tf}] \leq e^{ctf + ct^2}$. Using Chernov’s bound, we deduce that for all $a, t \geq 0,$

$$
\pi(f \geq E[f] + a) \leq e^{at}.
$$

(112)

The claim now follows by setting $t = \frac{a}{2c}$, so as to minimize the right-hand side.

\[\square\]

**Acknowledgments**

We thank Prasad Tetali and Radoslaw Adamczak for useful discussions and for bringing some relevant references to our attention. We also thank two anonymous referees for their many remarks.

**References**


