The Annals of Probability
2021, Vol. 49, No. 1, 506–525
https://doi.org/10.1214/20-AOP1453

MODIFIED LOG-SOBOLEV INEQUALITIES FOR STRONGLY LOG-CONCAVE DISTRIBUTIONS

BY MARY CRYAN*, HENG GUO† AND GIORGOS MOUSA‡

School of Informatics, University of Edinburgh, *mcryan@inf.ed.ac.uk; †hguo@inf.ed.ac.uk; ‡g.mousa@ed.ac.uk

We show that the modified log-Sobolev constant for a natural Markov chain which converges to an r-homogeneous strongly log-concave distribution is at least $1/r$. Applications include a sharp mixing time bound for the bases-exchange walk for matroids, and a concentration bound for Lipschitz functions over these distributions.

1. Introduction. Let $\pi : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a discrete distribution, where $[n] = \{1, \ldots, n\}$. Consider the generating polynomial of π:

$$g_\pi(\mathbf{x}) = \sum_{S \subseteq [n]} \pi(S) \prod_{i \in S} x_i.$$

We call a polynomial *log-concave* if its logarithm is concave, and *strongly log-concave* (**SLC**) if it is log-concave at the all-ones vector $\mathbf{1}$ after taking any sequence of partial derivatives. The distribution π is homogeneous and strongly log-concave if g_π is.

An important example of homogeneous strongly log-concave distributions is the uniform distribution over the bases of a matroid (Anari et al. (2019), Brändén and Huh (2019)).[1] This discovery leads to the breakthrough result that the exchange walk over the bases of a matroid is rapidly mixing (Anari et al. (2019)), which implies the existence of a fully polynomial-time randomised approximation scheme (FPRAS) for the number of bases of any matroid (given by an independence oracle).

The bases-exchange walk, denoted by P_{BX}, is defined as follows. In each step, we remove an element from the current basis uniformly at random to get a set S. Then we move to a basis containing S uniformly at random.[2] This chain is irreducible and it converges to the uniform distribution over the bases of a matroid. Brändén and Huh (2019) showed that the support of an r-homogeneous strongly log-concave distribution π must be the set of bases of a matroid. Thus, to sample from π, we may use a random walk $P_{\mathrm{BX},\pi}$ similar to the above. The only change required is that in the second step we move to a basis $B \supset S$ with probability proportional to $\pi(B)$.

Let P be a Markov chain over a state space Ω, and π be its stationary distribution. To measure the convergence rate of P, we use the total variation mixing time,

$$t_{\mathrm{mix}}(P, \varepsilon) := \min_t \{ t \mid \| P^t(x_0, \cdot) - \pi \|_{\mathrm{TV}} \leq \varepsilon \},$$

where $x_0 \in \Omega$ is the initial state and the subscript TV denotes the total variation distance between two distributions. The main goal of this paper is to show that for any r-homogeneous

Received September 2019.
MSC2020 subject classifications. 60J10, 05B35.
Key words and phrases. Markov chains, matroids, log-Sobolev inequality, bases exchange walk.

[1]For other examples, such as the determinantal point process and its variants, see Anari et al. (2019).

[2]Notice that to implement this step it may require more than constant time. The chain considered here is sometimes called the modified bases-exchange walk. A common alternative in the literature is to randomly propose an element and then apply a rejection filter.

strongly log-concave distribution π,

$$t_{\mathrm{mix}}(P_{\mathrm{BX},\pi}, \varepsilon) \le r\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right), \tag{1}$$

where $\pi_{\min} = \min_{x\in\Omega}\pi(x)$. This will improve upon the previous bound $t_{\mathrm{mix}}(P_{\mathrm{BX},\pi}, \varepsilon) \le r(\log\frac{1}{\pi_{\min}} + \log\frac{1}{\varepsilon})$ due to Anari et al. (2019). Since $\pi_{\min}$ is most commonly exponentially small in the input size (e.g., when π is the uniform distribution), the improvement is usually a polynomial factor. Our upper bound is sharp, as it is achieved (up to constant factors) when π is the uniform distribution over the bases of some matroids (Jerrum (2003)).[3]

Our main improvement is a modified log-Sobolev inequality (mLSI) for π and $P_{\mathrm{BX},\pi}$. To introduce this inequality, we define the *Dirichlet form* of a reversible Markov chain P, over state space Ω, as

$$\mathcal{E}_P(f, g) := \sum_{x,y\in\Omega} \pi(x)f(x)[\mathbf{I} - P](x, y)g(y),$$

where f, g are two functions over Ω, and $\mathbf{I}$ denotes the identity matrix. Moreover, let the (normalized) relative entropy of $f : \Omega \to \mathbb{R}_{\ge 0}$ be

$$\mathrm{Ent}_\pi(f) := \mathbb{E}_\pi(f \log f) - \mathbb{E}_\pi f \log \mathbb{E}_\pi f,$$

where we follow the convention that $0\log 0 = 0$. If we normalize $\mathbb{E}_\pi f = 1$, then $\mathrm{Ent}_\pi(f)$ is the relative entropy (or Kullback–Leibler divergence) between $\pi(\cdot)f(\cdot)$ and $\pi(\cdot)$.

The modified log-Sobolev constant (Bobkov and Tetali (2006)) is defined as

$$\rho(P) := \inf\left\{\frac{\mathcal{E}_P(f, \log f)}{\mathrm{Ent}_\pi(f)} \;\middle|\; f : \Omega \to \mathbb{R}_{\ge 0}, \mathrm{Ent}_\pi(f) \ne 0\right\}.$$

Our main theorem is the following, which is a special case of Theorem 7.

THEOREM 1. *Let π be an r-homogeneous strongly log-concave distribution, and $P_{\mathrm{BX},\pi}$ is the corresponding bases-exchange walk. Then*

$$\rho(P_{\mathrm{BX},\pi}) \ge \frac{1}{r}.$$

Since $t_{\mathrm{mix}}(P, \varepsilon) \le \frac{1}{\rho(P)}(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2})$ (cf. Bobkov and Tetali (2006)), Theorem 1 directly implies the mixing time bound (1).

In fact, we show more than Theorem 1. Following Anari et al. (2019) and Kaufman and Oppenheim (2018), we stratify independent sets of the matroid $\mathcal{M}$ by their sizes, and define two random walks for each level, depending on whether they add or delete an element first. For instance, the bases-exchange walk $P_{\mathrm{BX},\pi}$ is the "delete-add" or "down-up" walk for the top level. We give lower bounds for the modified log-Sobolev constants of both random walks for all levels. For the complete statement, see Section 3 and Theorem 7.

The previous work of Anari et al. (2019), building upon (Kaufman and Oppenheim (2018)), focuses on the spectral gap of $P_{\mathrm{BX},\pi}$. It is well known that lower bounds of the modified log-Sobolev constant are stronger than those of the spectral gap. Thus, we need to seek a different approach. Our key lemma, Lemma 11, shows that the relative entropy decays by a factor of $1 - \frac{1}{k}$ when we go from level k to level $k - 1$. Theorem 1 is a simple

[3] One such example is the matroid defined by a graph which is similar to a path but with two parallel edges connecting every two successive vertices instead of a single edge. Equivalently, this can be viewed as the partition matroid where each block has two elements and each basis is formed by choosing exactly one element from every block. The Markov chain $P_{\mathrm{BX},\pi}$ in this case is just the 1/2-lazy random walk on the Boolean hypercube.

consequence of this lemma and Jensen's inequality. In order to prove this lemma, we used a decomposition idea to inductively bound the relative entropy. Although similar ideas have appeared before (Jerrum et al. (2004), Lee and Yau (1998), Morris (2009, 2013)) our approach does not seem to fall into any existing framework.

Prior to our work, similar bounds have been obtained mostly for strong Rayleigh distributions, which, introduced by Borcea, Brändén and Liggett (2009), are a proper subset of strongly log-concave distributions. Hermon and Salez (2019) showed a lower bound on the modified log-Sobolev constant for strong Rayleigh distributions,[4] improving upon the spectral gap bound of Anari, Oveis Gharan and Rezaei (2016). The work of Hermon and Salez (2019) builds upon the previous work of Jerrum et al. (2004) for balanced matroids (Feder and Mihail (1992)). All of these results follow an inductive framework inspired by Lee and Yau (1998), which is apparently difficult to carry out in the case of general matroids or strongly log-concave distributions. Our analysis of the relative entropy took a different path from this line of work.

By the standard Herbst argument (see, e.g., Boucheron, Lugosi and Massart (2013), Goel (2004), Sammer (2005)), Theorem 1 also implies the following concentration bound.

COROLLARY 2. *Let π be an r-homogeneous strongly log-concave distribution with support $\Omega \subset 2^{[n]}$, and $P_{\mathrm{BX},\pi}$ be the corresponding bases-exchange walk. For any observable function $f : \Omega \to \mathbb{R}$ and $a \geq 0$,*

$$\Pr_{x\sim\pi}\big(\big|f(x) - \mathbb{E}_\pi f\big| \geq a\big) \leq 2\exp\left(-\frac{a^2}{2rv(f)}\right),$$

where $v(f)$ is the maximum of one-step variances,

$$v(f) := \max_{x\in\Omega}\left\{\sum_{y\in\Omega} P_{\mathrm{BX},\pi}(x,y)\big(f(x) - f(y)\big)^2\right\}.$$

There have been a number of results concerning concentration inequalities for Lipshitz functions of negatively correlated variables. Pemantle and Peres (2014) showed concentration for variables satisfying the *stochastic covering property* (**SCP**), which includes strong Rayleigh distributions as special cases. (See also Hermon and Salez (2019).) Correcting an earlier proof of Dubhashi and Ranjan (1998), Garbe and Vondrák (2018) showed concentration for variables with *negative regression* (**NR**), a property even weaker than **SCP**.

For a c-Lipschitz function (under the graph distance in the bases-exchange graph), $v(f) \leq c^2$. Thus, Corollary 2 generalizes the concentration bound for Lipschitz functions in strong Rayleigh distributions. However, **SLC** is *not* a negative correlation property. We construct examples in the Appendix to show that **SCP** and **SLC** are in fact incomparable. Thus, Corollary 2 is incomparable to the results of Garbe and Vondrák (2018), Hermon and Salez (2019), Pemantle and Peres (2014). It is not clear whether there is a larger class of distributions, generalizing both **NR** and **SLC**, which retains this concentration bound.

It is an interesting open problem to extend our result to more general settings. **SLC** distributions are special cases of high-dimensional expanders, where all local spectral gaps are at least 1. For more general cases, "local-to-global" bounds for spectral gaps have been obtained (Alev and Lau (2020), Kaufman and Oppenheim (2018)), whereas local-to-global mLSI on high-dimensional expanders is still elusive. Another interesting setting is the uniform distribution over common bases or independent sets of two matroids. Is there a Markov chain that

[4]The result of Hermon and Salez (2019) in fact requires a weaker assumption, namely the *stochastic covering property* (**SCP**). We construct examples in the Appendix to show that **SCP** and **SLC** are in fact incomparable.

converges rapidly to such distributions? Note that this setting includes the important problem of sampling perfect matchings of bipartite graphs, where the only known efficient algorithm is through an annealing process and its running time is a polynomial with high exponent (Jerrum, Sinclair and Vigoda (2004)).

In Section 2, we introduce necessary notions and briefly review relevant background. In Section 3, we formally state our main results. In Section 4, we show the decay of relative entropy and modified log-Sobolev constant lower bounds for the "down-up" and "up-down" walks. In Section 5, we show the concentration bound. In the Appendix, we discuss stochastic covering property and strong log-concavity.

2. Preliminaries. In this section, we define and give some basic properties of Markov chains, strongly log-concave distributions and matroids.

2.1. *Markov chains*. Let Ω be a discrete state space and π be a distribution over Ω. Let $P : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ be the transition matrix of a Markov chain whose stationary distribution is π. Then $\sum_{y\in\Omega} P(x, y) = 1$ for any $x \in \Omega$. We say P is *reversible* with respect to π if

$$\pi(x)P(x, y) = \pi(y)P(y, x). \tag{2}$$

We adopt the standard notation of $\mathbb{E}_\pi$ for a function $f : \Omega \to \mathbb{R}$, namely

$$\mathbb{E}_\pi f = \sum_{x\in\Omega} \pi(x)f(x).$$

We also view the transition matrix P as an operator that maps functions to functions. More precisely, let f be a function $f : \Omega \to \mathbb{R}$ and P acting on f is defined as

$$Pf(x) := \sum_{y\in\Omega} P(x, y)f(y).$$

This is also called the *Markov operator* corresponding to P. We will not distinguish the matrix P from the operator P as it will be clear from the context. Note that $Pf(x)$ is the expectation of f with respect to the distribution $P(x, \cdot)$. We can regard a function f as a column vector in $\mathbb{R}^\Omega$, in which case Pf is simply matrix multiplication.

The Hilbert space $L_2(\pi)$ is given by endowing $\mathbb{R}^\Omega$ with the inner product

$$\langle f, g\rangle_\pi := \sum_{x\in\Omega} \pi(x)f(x)g(x),$$

where $f, g \in \mathbb{R}^\Omega$. In particular, the norm in $L_2(\pi)$ is given by $\|f\|_\pi := \sqrt{\langle f, f\rangle_\pi}$.

The adjoint operator P^* of P is defined as $P^*(x, y) = \frac{\pi(y)P(y,x)}{\pi(x)}$. This is the (unique) operator which satisfies $\langle f, Pg\rangle_\pi = \langle P^*f, g\rangle_\pi$. It is easy to verify that if P satisfies the detailed balanced condition (2) (so P is reversible), then P is *self-adjoint*, namely $P = P^*$.

The *Dirichlet form* is defined as

$$\mathcal{E}_P(f, g) := \langle(\mathbf{I} - P)f, g\rangle_\pi, \tag{3}$$

where $\mathbf{I}$ stands for the identity matrix of the appropriate size. Let the *Laplacian* $\mathcal{L} := \mathbf{I} - P$. Then

$$\begin{aligned}\mathcal{E}_P(f, g) &= \sum_{x,y\in\Omega} \pi(x)g(x)\mathcal{L}(x, y)f(y)\\ &= g^{\mathrm{T}} \operatorname{diag}(\pi)\mathcal{L}f,\end{aligned}$$

where in the last line we regard f, g and π as (column) vectors over Ω. In particular, if P is reversible, then $\mathcal{L}^* = \mathcal{L}$ and

$$\begin{aligned}\mathcal{E}_P(f,g) &= \langle \mathcal{L}f, g\rangle_\pi = \langle f, \mathcal{L}^* g\rangle_\pi = \langle f, \mathcal{L}g\rangle_\pi = \mathcal{E}_P(g,f) \\ &= f^{\mathrm{T}} \operatorname{diag}(\pi)\mathcal{L}g.\end{aligned} \tag{4}$$

In this paper, all Markov chains are reversible and we will most commonly use the form (4). Another common expression of the Dirichlet form for reversible P is

$$\mathcal{E}_P(f,g) = \frac{1}{2}\sum_{x\in\Omega}\sum_{y\in\Omega} \pi(x)P(x,y)\big(f(x)-f(y)\big)\big(g(x)-g(y)\big), \tag{5}$$

but we will not need this expression until Section 5. It is well known that the spectral gap of P, or equivalently the smallest positive eigenvalue of $\mathcal{L}$, controls the convergence rate of P. It also has a variational characterization. Let the variance of f be

$$\operatorname{Var}_\pi(f) := \mathbb{E}_\pi f^2 - (\mathbb{E}_\pi f)^2.$$

Then

$$\lambda(P) := \inf\left\{ \frac{\mathcal{E}_P(f,f)}{\operatorname{Var}_\pi(f)} \;\middle|\; f:\Omega\to\mathbb{R}, \operatorname{Var}_\pi(f)\neq 0\right\}.$$

The usefulness of $\lambda(P)$ is due to the fact that, if, say, all eigenvalues of P are nonnegative, then

$$t_{\mathrm{mix}}(P,\varepsilon) \le \frac{1}{\lambda(P)}\left(\frac{1}{2}\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon}\right), \tag{6}$$

where $\pi_{\min} = \min_{x\in\Omega}\pi(x)$; see, for example, Levin and Peres (2017), Theorem 12.4.

The (standard) log-Sobolev inequality relates $\mathcal{E}_P(\sqrt{f},\sqrt{f})$ with the following entropy-like quantity:

$$\operatorname{Ent}_\pi(f) := \mathbb{E}_\pi(f\log f) - \mathbb{E}_\pi f \log \mathbb{E}_\pi f, \tag{7}$$

for a nonnegative function f, where we follow the convention that $0\log 0 = 0$. Also, log always stands for the natural logarithm in this paper. The log-Sobolev constant is defined as

$$\alpha(P) := \inf\left\{ \frac{\mathcal{E}_P(\sqrt{f},\sqrt{f})}{\operatorname{Ent}_\pi(f)} \;\middle|\; f:\Omega\to\mathbb{R}_{\ge 0}, \operatorname{Ent}_\pi(f)\neq 0\right\}.$$

The constant $\alpha(P)$ gives a better control of the mixing time of P, as shown by Diaconis and Saloff-Coste (1996),

$$t_{\mathrm{mix}}(P,\varepsilon) \le \frac{1}{4\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right). \tag{8}$$

The saving seems modest comparing to (6), but it is quite common that $\pi_{\min}$ is exponentially small in the instance size, in which case the saving is a polynomial factor.

What we are interested in, however, is the following modified log-Sobolev constant introduced by Bobkov and Tetali (2006):

$$\rho(P) := \inf\left\{ \frac{\mathcal{E}_P(f,\log f)}{\operatorname{Ent}_\pi(f)} \;\middle|\; f:\Omega\to\mathbb{R}_{\ge 0}, \operatorname{Ent}_\pi(f)\neq 0\right\}.$$

Similar to (8), we have that

$$t_{\mathrm{mix}}(P,\varepsilon) \le \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right), \tag{9}$$

as shown by Bobkov and Tetali (2006), Corollary 2.8.

For reversible P, the following relationships among these constants are known:

$$2\lambda(P) \geq \rho(P) \geq 4\alpha(P).$$

See, for example, Bobkov and Tetali (2006), Proposition 3.6.

Thus, lower bounds on these constants are increasingly difficult to obtain. However, to get the best asymptotic control of the mixing time, one only needs to lower bound the modified log-Sobolev constant $\rho(P)$ instead of $\alpha(P)$ by comparing (8) and (9). Indeed, as observed by Hermon and Salez (2019), by taking the indicator function $\frac{1}{\pi(x)}\mathbb{1}_x$ for all $x \in \Omega$,

$$\alpha(P) \leq \min_{x\in\Omega}\left\{\frac{1}{-\log \pi(x)}\right\}.$$

In our setting of r-homogeneous strongly log-concave distributions, we cannot hope for a uniform bound for $\alpha(P)$ similar to Theorem 1, as the right-hand side of the above can be arbitrarily small for fixed r.

By (3) and (7), it is clear that if we replace f by cf for some constant $c > 0$, then both $\mathcal{E}_P(f, \log f)$ and $\mathrm{Ent}_\pi(f)$ increase by the same factor c. Thus, in order to bound ρ, we may further assume that $\mathbb{E}_\pi f = 1$. This assumption allows the simplification $\mathrm{Ent}_\pi(f) = \mathbb{E}_\pi(f \log f)$. In this case, $\pi(\cdot)f(\cdot)$ is a distribution, and $\mathrm{Ent}_\pi(f)$ is the relative entropy (or Kullback–Leibler divergence) between $\pi(\cdot)f(\cdot)$ and $\pi(\cdot)$.

2.2. *Strongly log-concave distributions.* We write ∂_i as shorthand for $\frac{\partial}{\partial x_i}$, and ∂_I for an index set $I = \{i_1, \dots, i_k\}$ as shorthand for $\partial_{i_1} \dots \partial_{i_k}$.

DEFINITION 3. A polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ with nonnegative coefficients is *log-concave* at $\mathbf{x} \in \mathbb{R}_{\geq 0}$ if its Hessian $\nabla^2 \log p$ is negative semidefinite at $\mathbf{x}$. We call p *strongly log-concave* if for any index set $I \subseteq [n]$, $\partial_I p$ is log-concave at the all-1 vector $\mathbf{1}$.

The notion of strong log-concavity was introduced by Gurvits (2009a, 2009b). There are also notions of *complete log-concavity* introduced by Anari, Oveis Gharan and Vinzant (2018), and *Lorentzian* polynomials introduced by Brändén and Huh (2019). It turns out that for homogeneous polynomials the three notions are equivalent (Brändén and Huh (2019), Theorem 5.3). (See also Anari et al. (2019).)

The following property of strongly log-concave polynomials is particularly useful (Anari, Oveis Gharan and Vinzant (2018), Brändén and Huh (2019)).

PROPOSITION 4. *If p is strongly log-concave, then for any $I \subseteq [n]$, the Hessian matrix $\nabla^2 \partial_I p(\mathbf{1})$ has at most one positive eigenvalue.*

In fact, when p is homogeneous, $\nabla^2 \partial_I p(\mathbf{1})$ having at most one positive eigenvalue is equivalent to $\nabla^2 \log \partial_I p(\mathbf{1})$ being negative semidefinite (Anari, Oveis Gharan and Vinzant (2018)), but we will only need the proposition above.

A distribution π is called *r-homogeneous* (or *strongly log-concave*) if g_π is.

2.3. *Matroids.* A matroid is a combinatorial structure that abstracts the notion of linear independence. We shall define it in terms of its independent sets, although many different equivalent definitions exist. Formally, a matroid $\mathcal{M} = (E, \mathcal{I})$ consists of a finite ground set E and a collection $\mathcal{I}$ of subsets of E (independent sets) that satisfy the following:

- $\varnothing \in \mathcal{I}$;
- if $S \in \mathcal{I}$, $T \subseteq S$, then $T \in \mathcal{I}$;
- if $S, T \in \mathcal{I}$ and $|S| > |T|$, then there exists an element $i \in S \setminus T$ such that $T \cup \{i\} \in \mathcal{I}$.

The first condition guarantees that $\mathcal{I}$ is nonempty, the second implies that $\mathcal{I}$ is downward closed, and the third is usually called the augmentation axiom. We direct the reader to Oxley (1992) for a reference book on matroid theory. In particular, the augmentation axiom implies that all the maximal independent sets have the same cardinality, namely the rank r of $\mathcal{M}$. The set of bases $\mathcal{B}$ is the collection of maximal independent sets of $\mathcal{M}$. Furthermore, we denote by $\mathcal{M}(k)$ the collection of independent sets of size k, where $0 \le k \le r$. If we dropped the augmentation axiom, the resulting structure would be a nonempty collection of subsets of E that is downward closed, known as an (abstract) *simplicial complex*.

Brändén and Huh (2019), Theorem 7.1, showed that the support of an r-homogeneous strongly log-concave distribution π is the set of bases of a matroid $\mathcal{M} = (E, \mathcal{I})$ of rank r. We equip $\mathcal{I}$ with a weight function $w(\cdot)$ recursively defined as follows:[5]

$$w(I) := \begin{cases} \pi(I)Z_r & \text{if } |I| = r, \\ \displaystyle\sum_{I' \supset I, |I'| = |I|+1} w(I') & \text{if } |I| < r, \end{cases}$$

for some normalization constant $Z_r > 0$. For example, we may choose $w(B) = 1$ for all $B \in \mathcal{B}$ and $Z_r = |\mathcal{B}|$, which corresponds to the uniform distribution over $\mathcal{B}$. It follows that

$$w(I) = \bigl(r - |I|\bigr)! \sum_{B \in \mathcal{B}, I \subseteq B} w(B).$$

Let π_k be the distribution over $\mathcal{M}(k)$ such that $\pi_k(I) \propto w(I)$ for $I \in \mathcal{M}(k)$. Thus $\pi = \pi_r$. For any $I \in \mathcal{M}(k)$, $\pi_k(I)$ is proportional to the probability of generating a superset of I under π. Let $Z_k = \sum_{I \in \mathcal{M}(k)} w(I)$ be the normalization constant of π_k. In fact, for any $0 \le k \le r$, $k! Z_k = Z_0 = w(\varnothing)$.

It is straightforward to verify that for any $I \in \mathcal{I}$,

$$\partial_I g_\pi(\mathbf{1}) = \sum_{B \in \mathcal{B}, I \subset B} \pi(B) = \frac{1}{Z_r} \sum_{B \in \mathcal{B}, I \subset B} w(B). \tag{10}$$

We also write $w(v)$ as shorthand for $w(\{v\})$ for any $v \in E$.

For an independent set $I \in \mathcal{I}$, the contraction $\mathcal{M}_I = (E \setminus I, \mathcal{I}_I)$ is also a matroid, where $\mathcal{I}_I = \{J \mid J \subseteq E \setminus I, J \cup I \in \mathcal{I}\}$. We equip $\mathcal{M}_I$ with a weight function $w_I(\cdot)$ such that $w_I(J) = w(I \cup J)$. We may similarly define distributions $\pi_{I,k}$ for $k \le r - |I|$ such that $\pi_{I,k}(J) \propto w_I(J)$ for $J \in \mathcal{M}_I(k)$. For convenience, instead of defining $\pi_{I,k}$ over $\mathcal{M}_I(k)$, we define it over $\mathcal{M}(k + |I|)$ such that for any $J \in \mathcal{M}(k + |I|)$,

$$\pi_{I,k}(J) := \begin{cases} \dfrac{k! w(J)}{w(I)} & \text{if } I \subset J; \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Notice that the normalizing constant $Z_{I,k} = \frac{w(I)}{k!}$.

If $|I| \le r - 2$, let W_I be the matrix such that $W_{uv} = w_I(\{u, v\})$ for any $u, v \in E \setminus I$. Then notice that

$$\begin{aligned} w_I\bigl(\{u, v\}\bigr) &= w\bigl(I \cup \{u, v\}\bigr) \\ &= \bigl(r - |I| - 2\bigr)! \sum_{B \in \mathcal{B}, I \cup \{u,v\} \subseteq B} w(B) \\ \text{(by (10))} \qquad &= \bigl(r - |I| - 2\bigr)! Z_r \cdot \partial_u \partial_v \partial_I g_\pi(\mathbf{1}). \end{aligned}$$

[5]One may define $w(I)$ to be a $\frac{k!}{r!}$ fraction of the current definition for $I \in \mathcal{M}(k)$. This alternative definition will eliminate many factorial factors in the rest of the paper. However, it is inconsistent with the literature (Anari et al. (2019), Kaufman and Oppenheim (2018)), so we do not adopt it.

In other words, W_I is $\nabla^2 \partial_I g_\pi$ multiplied by the scalar $(r - |I| - 2)! Z_r$. Thus, Proposition 4 implies the following.

PROPOSITION 5. *Let π be an r-homogeneous strongly log-concave distribution over $\mathcal{M} = (E, \mathcal{I})$. If $I \in \mathcal{I}$ and $|I| \leq r - 2$, then the matrix W_I has at most one positive eigenvalue.*

Proposition 5 implies the following bound for a quadratic form, which will be useful later.

LEMMA 6. *Let π be an r-homogeneous strongly log-concave distribution over $\mathcal{M} = (E, \mathcal{I})$, and let $I \in \mathcal{I}$ such that $|I| \leq r - 2$. Let $f : \mathcal{M}_I(1) \to \mathbb{R}$ be a function such that $\mathbb{E}_{\pi_{I,1}} f = 1$. Then*

$$f^{\mathrm{T}} W_I f \leq w(I).$$

PROOF. Let $\mathbf{w}_I = \{w_I(v)\}_{v \in E \setminus I}$. The constraint $\mathbb{E}_{\pi_{I,1}} f = 1$ implies that $\sum_{v \in E \setminus I} w_I(v) f(v) = w(I)$. Let $D = \operatorname{diag}(\mathbf{w}_I)$ and $A = D^{-1/2} W_I D^{-1/2}$. Then A is a real symmetric matrix. By Proposition 5, W_I has at most one positive eigenvalue, and thus so does A (see, e.g., Anari et al. (2019), Lemma 2.4). We may decompose A as

$$A = \sum_{i=1}^{|E \setminus I|} \lambda_i g_i g_i^{\mathrm{T}}, \tag{12}$$

where $\{g_i\}$ is an orthonormal basis and $\lambda_i \leq 0$ for all $i \geq 2$. Moreover, notice that $\sqrt{\mathbf{w}_I}$ is an eigenvector of A with eigenvalue 1. Thus, $\lambda_1 = 1$ and g_1 can be taken as $\sqrt{\pi_{I,1}}$.

The decomposition (12) directly implies that

$$W = \sum_{i=1}^{|E \setminus I|} \lambda_i h_i h_i^{\mathrm{T}},$$

where $h_i = D^{1/2} g_i$. In particular, $h_1 = \frac{1}{\sqrt{w(I)}} \mathbf{w}_I$.

The assumption $\sum_{v \in E \setminus I} w_I(v) f(v) = w(I)$ can be rewritten as $\langle h_1, f \rangle = \sqrt{w(I)}$. Thus,

$$f^{\mathrm{T}} W_I f = \sum_{i=1}^{|E \setminus I|} \lambda_i \langle h_i, f \rangle^2 \leq \langle h_1, f \rangle^2 = w(I),$$

where the inequality is due to the fact that $\lambda_1 = 1$ and $\lambda_i \leq 0$ for all $i \geq 2$. The lemma follows. $\square$

3. Main results. There are two natural random walks $P_k^\wedge$ and $P_k^\vee$ on $\mathcal{M}(k)$ by starting with adding or deleting an element and coming back to $\mathcal{M}(k)$. Given the current $I \in \mathcal{M}(k)$, the "up-down" random walk $P_k^\wedge$ first chooses $I' \in \mathcal{M}(k+1)$ such that $I' \supset I$ with probability proportional to $w(I')$, and then removes one element from I' uniformly at random. More formally, for $1 \leq k \leq r - 1$ and $I, J \in \mathcal{M}(k)$, we have that

$$P_k^\wedge(I, J) = \begin{cases} \dfrac{1}{k+1} & \text{if } I = J; \\ \dfrac{w(I \cup J)}{(k+1) w(I)} & \text{if } I \cup J \in \mathcal{M}(k+1); \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

The "down-up" random walk $P_k^{\vee}$ removes an element of I uniformly at random to get $I' \in \mathcal{M}(k-1)$, and then moves to J such that $J \in \mathcal{M}(k)$, $J \supset I'$ with probability proportional to $w(J)$. More formally, for $2 \le k \le r$,

$$P_k^{\vee}(I, J) = \begin{cases} \displaystyle\sum_{I' \in \mathcal{M}(k-1), I' \subset I} \frac{w(I)}{kw(I')} & \text{if } I = J; \\ \dfrac{w(J)}{kw(I \cap J)} & \text{if } |I \cap J| = k-1; \\ 0 & \text{if } |I \cap J| < k-1. \end{cases} \tag{14}$$

Thus, the bases-exchange walk $P_{\mathrm{BX},\pi}$ according to π is just $P_r^{\vee}$. The stationary distribution of both $P_k^{\wedge}$ and $P_k^{\vee}$ is $\pi_k(I) = \frac{w(I)}{Z_k} = \frac{k!w(I)}{r!Z_r}$.

THEOREM 7. *Let π be an r-homogeneous strongly log-concave distribution, and $\mathcal{M}$ the associated matroid. Let $P_k^{\vee}$ and $P_k^{\wedge}$ be defined as above on $\mathcal{M}(k)$. Then the following hold*:

- *for any $2 \le k \le r$, $\rho(P_k^{\vee}) \ge \frac{1}{k}$;*
- *for any $1 \le k \le r-1$, $\rho(P_k^{\wedge}) \ge \frac{1}{k+1}$.*

Theorem 7 is shown in Section 4. Interestingly, we do not know how to directly relate $\rho(P_k^{\wedge})$ with $\rho(P_{k+1}^{\vee})$, although it is straightforward to see that both walks have the same spectral gap (see (17) and (18) below).

By (9), we have the following corollary.

COROLLARY 8. *In the same setting as Theorem 7, we have that*:

- *for any $2 \le k \le r$, $t_{\mathrm{mix}}(P_k^{\vee}, \varepsilon) \le k(\log\log \pi_{k,\min}^{-1} + \log \frac{1}{2\varepsilon^2})$;*
- *for any $1 \le k \le r-1$, $t_{\mathrm{mix}}(P_k^{\wedge}, \varepsilon) \le (k+1)(\log\log \pi_{k,\min}^{-1} + \log \frac{1}{2\varepsilon^2})$.*

In particular, for the bases-exchange walk $P_{\mathrm{BX},\pi}$ according to π,

$$t_{\mathrm{mix}}(P_{\mathrm{BX},\pi}, \varepsilon) \le r\left(\log\log \pi_{\min}^{-1} + \log \frac{1}{2\varepsilon^2}\right).$$

Let $\mathcal{M}$ be a matroid of rank r with a ground set of size n. For the uniform distribution over the bases of $\mathcal{M}$, Corollary 8 implies that the mixing time of the bases-exchange walk is $O(r(\log r + \log\log n))$, which improves upon the $O(r^2 \log n)$ bound of Anari et al. (2019). The mixing time bound in Corollary 8 is sharp, as there are matroids where the upper bound is achieved (Example 9.14 Jerrum (2003)). As mentioned in the Introduction, one such example is the graphic matroid defined by a graph which is similar to a path but with two parallel edges connecting every two successive vertices instead of a single edge. Equivalently, this can be viewed as the partition matroid where each block has two elements and each basis is formed by choosing exactly one element from every block. The rank of this matroid is $r = n/2$, and $\pi_{\min} = \frac{1}{2^{n/2}}$. The Markov chain $P_{\mathrm{BX},\pi}$ in this case is just the 1/2-lazy random walk on the $n/2$-dimensional Boolean hypercube, which has mixing time $\Theta(n \log n)$, matching the upper bound in Corollary 8.

For more details on the concentration result, Corollary 2, see Section 5.

4. Decay of relative entropy. In this section and what follows, we always assume that the matroid $\mathcal{M}$ and the weight function $w(\cdot)$ correspond to an r-homogeneous strongly log-concave distribution $\pi = \pi_r$.

We first give some basic decompositions of $P_k^{\vee}$ and $P_k^{\wedge}$. Let A_k be a matrix whose rows are indexed by $\mathcal{M}(k)$ and columns by $\mathcal{M}(k+1)$ such that

$$A_k(I,J) := \begin{cases} 1 & \text{if } I \subset J; \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{w}_k$ be the vector of $\{w(I)\}_{I \in \mathcal{M}(k)}$. Moreover, let

$$P_k^{\uparrow} := \operatorname{diag}(\mathbf{w}_k)^{-1} A_k \operatorname{diag}(\mathbf{w}_{k+1}), \tag{15}$$

$$P_{k+1}^{\downarrow} := \frac{1}{k+1} A_k^{\mathrm{T}}. \tag{16}$$

Then

$$P_k^{\wedge} = P_k^{\uparrow} P_{k+1}^{\downarrow}, \tag{17}$$

$$P_{k+1}^{\vee} = P_{k+1}^{\downarrow} P_k^{\uparrow}. \tag{18}$$

Let $D_k = \operatorname{diag}(\pi_k)$. Using (15) and (16), we get that

$$D_{k+1} P_{k+1}^{\downarrow} = (P_k^{\uparrow})^{\mathrm{T}} D_k. \tag{19}$$

By multiplying equation (19) by the all-ones vector, we also get that

$$\pi_{k+1} P_{k+1}^{\downarrow} = \pi_k, \tag{20}$$

$$\pi_k P_k^{\uparrow} = \pi_{k+1}. \tag{21}$$

For $k \geq 2$ and a function $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\geq 0}$, define $f^{(i)} : \mathcal{M}(i) \to \mathbb{R}_{\geq 0}$ for $1 \leq i \leq k-1$ such that

$$f^{(i)} := \prod_{j=i}^{k-1} P_j^{\uparrow} f^{(k)}. \tag{22}$$

Intuitively, $f^{(i)}$ is the function $f^{(k)}$ "going down" to level i. The key lemma, namely Lemma 11, is that this operation contracts the relative entropy by a factor of $1 - \frac{1}{i}$ from level i to level $i-1$.

In fact, recall that if we normalize $\mathbb{E}_{\pi_k} f^{(k)} = 1$, then $(f^{(k)})^{\mathrm{T}} D_k$ is a distribution (viewed as a row vector). Then it is easy to verify that

$$(f^{(k-1)})^{\mathrm{T}} D_{k-1} = (f^{(k)})^{\mathrm{T}} D_k P_k^{\downarrow}. \tag{23}$$

Namely, the corresponding distribution of $f^{(k-1)}$ is that of $f^{(k)}$ after the random walk $P_k^{\downarrow}$.

We first establish some properties of $f^{(i)}$ for $i < k$.

LEMMA 9. *Let $k \geq 2$ and $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\geq 0}$ be a nonnegative function on $\mathcal{M}(k)$. Then we have the following*:

1. *for any $1 \leq i < k$, $J \in \mathcal{M}(i)$, $f^{(i)}(J) = \mathbb{E}_{\pi_{J,k-i}} f^{(k)}$*;
2. *for any $1 \leq i \leq k$, $\mathbb{E}_{\pi_i} f^{(i)} = \mathbb{E}_{\pi_k} f^{(k)}$.*

PROOF. For (1), first notice that

$$\begin{aligned} \delta_J^{\mathrm{T}} \prod_{j=i}^{k-1} P_j^{\uparrow} &= \delta_J^{\mathrm{T}} \prod_{j=i}^{k-1} [\operatorname{diag}(\mathbf{w}_j)^{-1} A_j \operatorname{diag}(\mathbf{w}_{j+1})] \\ &= \frac{\delta_J^{\mathrm{T}}}{w(J)} \prod_{j=i}^{k-1} A_j \operatorname{diag}(\mathbf{w}_k) = \pi_{J,k-i}, \end{aligned}$$

where δ_J is the Dirac vector that equals 1 at J and 0 elsewhere. The last equality holds due to the fact that the product of the adjacency matrices counts the paths from independent sets at level i to independent sets at level k. For every such pair of sets, the number of these paths is $(k-i)!$ if one is contained in the other, or 0 otherwise. It follows that

$$\mathbb{E}_{\pi_{J,k-i}} f^{(k)} = \pi_{J,k-i} f^{(k)} = \delta_J^{\mathrm{T}} \prod_{j=i}^{k-1} P_j^{\uparrow} f^{(k)} = \delta_J^{\mathrm{T}} f^{(i)} = f^{(i)}(J).$$

For (2), we have that

$$\begin{aligned} \mathbb{E}_{\pi_i} f^{(i)} &= \pi_i \prod_{j=i}^{k-1} P_j^{\uparrow} f^{(k)} \\ &= \pi_k f^{(k)} \quad \text{(by Equation (21))} \\ &= \mathbb{E}_{\pi_k} f^{(k)}. \end{aligned}$$

□

Now we are ready to establish the base case of the entropy's contraction.

LEMMA 10. *Let $f^{(2)} : \mathcal{M}(2) \to \mathbb{R}_{\geq 0}$ be a non-negative function defined on $\mathcal{M}(2)$. Then*

$$\mathrm{Ent}_{\pi_2}(f^{(2)}) \geq 2\mathrm{Ent}_{\pi_1}(f^{(1)}).$$

PROOF. Without loss of generality, we may assume that $\mathbb{E}_{\pi_2} f^{(2)} = 1$ and therefore $\mathbb{E}_{\pi_1} f^{(1)} = 1$ by (2) of Lemma 9. Note that for $v \in E$,

$$f^{(1)}(v) = \sum_{S \in \mathcal{M}(2): v \in S} \frac{w(S)}{w(v)} f^{(2)}(S).$$

We will use the following inequality, which is valid for any $a \geq 0$ and $b > 0$,

$$a \log \frac{a}{b} \geq a - b. \tag{24}$$

Noticing that $Z_1 = 2Z_2$, we have

$$\begin{aligned} &\mathrm{Ent}_{\pi_2}(f^{(2)}) - 2\mathrm{Ent}_{\pi_1}(f^{(1)}) \\ &\quad = \sum_{S \in \mathcal{M}(2)} \pi_2(S) f^{(2)}(S) \log f^{(2)}(S) \\ &\qquad - 2 \sum_{v \in E} \pi_1(v) \Bigg(\sum_{S \in \mathcal{M}(2): v \in S} \frac{w(S)}{w(v)} f^{(2)}(S) \Bigg) \log f^{(1)}(v) \\ &\quad = \sum_{S \in \mathcal{M}(2)} \Bigg(\pi_2(S) f^{(2)}(S) \log f^{(2)}(S) - 2 \sum_{v \in S} \pi_1(v) \frac{w(S)}{w(v)} f^{(2)}(S) \log f^{(1)}(v) \Bigg) \\ &\quad = \sum_{S \in \mathcal{M}(2)} \Bigg(\frac{w(S)}{Z_2} f^{(2)}(S) \log f^{(2)}(S) - 2 \sum_{v \in S} \frac{w(v)}{Z_1} \cdot \frac{w(S)}{w(v)} f^{(2)}(S) \log f^{(1)}(v) \Bigg) \\ &\quad = \sum_{S = \{u,v\} \in \mathcal{M}(2)} \frac{w(S)}{Z_2} f^{(2)}(S) \big(\log f^{(2)}(S) - \log f^{(1)}(v) - \log f^{(1)}(u) \big) \\ &\quad \geq \sum_{S = \{u,v\} \in \mathcal{M}(2)} \frac{w(S)}{Z_2} \big(f^{(2)}(S) - f^{(1)}(v) f^{(1)}(u) \big) \end{aligned}$$

$$= \sum_{S\in\mathcal{M}(2)} \pi_2(S) f^{(2)}(S) - \sum_{S=\{u,v\}\in\mathcal{M}(2)} \frac{w(S)}{Z_2} \cdot f^{(1)}(v) f^{(1)}(u)$$

$$= 1 - \frac{1}{2Z_2} \cdot \big(f^{(1)}\big)^{\mathrm{T}} W_\varnothing f^{(1)},$$

where the inequality is by (24) with $a = f^{(2)}(S)$ and $b = f^{(1)}(u) f^{(1)}(v)$ when $b > 0$, and when $b = 0$ we have $a = 0$ as well. Thus, the lemma follows from Lemma 6 with $I = \varnothing$ and $w(\varnothing) = Z_1 = 2Z_2$. $\square$

We generalize Lemma 10 as follows.

LEMMA 11. *Let $k \geq 2$ and $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\geq 0}$ be a nonnegative function defined on $\mathcal{M}(k)$. Then*

$$\mathrm{Ent}_{\pi_k}\big(f^{(k)}\big) \geq \frac{k}{k-1} \mathrm{Ent}_{\pi_{k-1}}\big(f^{(k-1)}\big).$$

PROOF. We do an induction on k. The base case of $k = 2$ follows from Lemma 10.

For the induction step, assume the lemma holds for all integers at most k for any matroid $\mathcal{M}$. Let $f^{(k+1)} : \mathcal{M}(k+1) \to \mathbb{R}_{\geq 0}$ be a nonnegative function such that $\mathbb{E}_{\pi_{k+1}} f^{(k+1)} = 1$.

Recall (11), where we define $\pi_{v,k}$ over $\mathcal{M}(k+1)$ instead of over $\mathcal{M}_v(k)$. For $I \in \mathcal{M}(k+1)$, $v \in \mathcal{M}(1)$ and $v \in I$,

$$\pi_{k+1}(I) = \frac{w(I)}{Z_{k+1}} = (k+1) \cdot \frac{w(v)}{(k+1)! Z_{k+1}} \cdot \frac{k! w(I)}{w(v)} = (k+1)\pi_1(v)\pi_{v,k}(I),$$

as $Z_1 = (k+1)! Z_{k+1}$. This means that

$$\pi_{k+1}(I) = \sum_{v\in\mathcal{M}(1), v\in I} \pi_1(v)\pi_{v,k}(I) = \sum_{v\in\mathcal{M}(1)} \pi_1(v)\pi_{v,k}(I). \tag{25}$$

Thus π_{k+1} is a mixture of $\pi_{v,k}$.

We use the "chain rule" of entropy to decompose $\mathrm{Ent}_{\pi_{k+1}}(f^{(k+1)})$ with respect to the entropy of $f^{(1)}$ ("projection") and the entropy conditioned on having each v ("restriction"). To be more precise, we have

$$\mathbb{E}_{\pi_{k+1}} f^{(k+1)} \log f^{(k+1)} = \sum_{v\in\mathcal{M}(1)} \pi_1(v) \mathbb{E}_{\pi_{v,k}} f^{(k+1)} \log f^{(k+1)}.$$

This implies that

$$\begin{aligned}\mathrm{Ent}_{\pi_{k+1}}\big(f^{(k+1)}\big) &= \sum_{v\in\mathcal{M}(1)} \pi_1(v) \mathrm{Ent}_{\pi_{v,k}}\big(f^{(k+1)}\big) \\ &\quad + \sum_{v\in\mathcal{M}(1)} \pi_1(v) \big(\mathbb{E}_{\pi_{v,k}} f^{(k+1)}\big) \log\big(\mathbb{E}_{\pi_{v,k}} f^{(k+1)}\big) \\ &= \sum_{v\in\mathcal{M}(1)} \pi_1(v) \mathrm{Ent}_{\pi_{v,k}}\big(f^{(k+1)}\big) + \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big),\end{aligned} \tag{26}$$

where we use (1) and (2) of Lemma 9. Similarly,

$$\mathrm{Ent}_{\pi_k}\big(f^{(k)}\big) = \sum_{v\in\mathcal{M}(1)} \pi_1(v) \mathrm{Ent}_{\pi_{v,k-1}}\big(f^{(k)}\big) + \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big). \tag{27}$$

For any $v \in \mathcal{M}(1)$, the contracted matroid $\mathcal{M}_v$ with weight function $w_v(I) = w(I \cup v)$ for $I \subseteq E \setminus \{v\}$ corresponds to an $(r-1)$-homogeneous strongly log-concave distribution. (Recall Definition 3.) Thus, we can apply the induction hypothesis on $\mathcal{M}_v$ at level k and get

$$\mathrm{Ent}_{\pi_{v,k}}\big(f^{(k+1)}\big) \geq \frac{k}{k-1} \cdot \mathrm{Ent}_{\pi_{v,k-1}}\big(f^{(k)}\big). \tag{28}$$

Strictly speaking, in (28) we should apply the induction hypothesis to $f_v^{(k)}$ which is the restriction of $f^{(k+1)}$ to $J \in \mathcal{M}(k+1)$ and $J \ni v$, and then "push it down" to $f_v^{(k-1)}$ defined over $I \in \mathcal{M}(k)$ and $I \ni v$ as

$$f_v^{(k-1)}(I) := \sum_{J \in \mathcal{M}(k+1): J \supset I} \frac{w(J)}{w(I)} \cdot f_v^{(k)}(J) = \sum_{J \in \mathcal{M}(k+1): J \supset I} \frac{w(J)}{w(I)} \cdot f^{(k+1)}(J).$$

However, $f_v^{(k)}$ agrees with $f^{(k+1)}$ on the support of $\pi_{v,k}$, and $f_v^{(k-1)}$ agrees with $f^{(k)}$ on the support of $\pi_{v,k-1}$. This validates (28).

Furthermore, using the induction hypothesis on $\mathcal{M}$ from level k to level 1, we have that

$$\mathrm{Ent}_{\pi_k}\big(f^{(k)}\big) \geq k \cdot \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big). \tag{29}$$

Thus, (27) and (29) together imply that

$$\sum_{v \in \mathcal{M}(1)} \pi_1(v) \mathrm{Ent}_{\pi_{v,k-1}}\big(f^{(k)}\big) \geq (k-1) \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big). \tag{30}$$

Putting everything together,

$$\begin{aligned}
& \mathrm{Ent}_{\pi_{k+1}}\big(f^{(k+1)}\big) \\
\text{(by (26))} \quad & = \sum_{v \in \mathcal{M}(1)} \pi_1(v) \mathrm{Ent}_{\pi_{v,k}}\big(f^{(k+1)}\big) + \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big) \\
\text{(by (28))} \quad & \geq \frac{k}{k-1} \sum_{v \in \mathcal{M}(1)} \pi_1(v) \mathrm{Ent}_{\pi_{v,k-1}}\big(f^{(k)}\big) + \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big) \\
& = \left(\frac{k+1}{k} + \frac{1}{k(k-1)}\right) \sum_{v \in \mathcal{M}(1)} \pi_1(v) \mathrm{Ent}_{\pi_{v,k-1}}\big(f^{(k)}\big) \\
& \quad + \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big) \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big) \\
\text{(by (30))} \quad & \geq \frac{k+1}{k} \sum_{v \in \mathcal{M}(1)} \pi_1(v) \mathrm{Ent}_{\pi_{v,k-1}}\big(f^{(k)}\big) + \frac{k+1}{k} \mathrm{Ent}_{\pi_1}\big(f^{(1)}\big) \\
\text{(by (27))} \quad & = \frac{k+1}{k} \mathrm{Ent}_{\pi_k}\big(f^{(k)}\big).
\end{aligned}$$

This concludes the inductive step and thus the proof. $\square$

REMARK. We remark that our decomposition of the relative entropy (26) is "horizontal" with respect to elements of $\mathcal{M}(1)$. This decomposition is different from the decomposition by Kaufman and Oppenheim (2018), Theorem 5.2 in a similar context, where they decompose "vertically" across all levels.

The decomposition (25) of π_k appears to be the key to Lemma 11. An alternative way to understand it is the following. Consider the process which first draws a basis $B \sim \pi$, and then repeatedly removes an element from the current set uniformly at random for at most

r repetitions. Let X_k be the outcome of this process after removing $r-k$ elements. Then $|X_k|=k$, and $\pi_k(I)=\Pr(X_k=I)$ for $I\in\mathcal{M}(k)$. Moreover,

$$\Pr(X_1=\{v\}\mid X_k=I)=\begin{cases}\frac{1}{k} & \text{if } v\in I;\\ 0 & \text{otherwise.}\end{cases}$$

By Bayes' rule,

$$\Pr(X_k=I\mid X_1=\{v\})\Pr(X_1=\{v\})=\Pr(X_1=\{v\}\mid X_k=I)\Pr(X_k=I).$$

Summing over v, since $\sum_{v\in\mathcal{M}(1)}\Pr(X_1=\{v\}\mid X_k=I)=1$, we have

$$\begin{aligned}&\sum_{v\in\mathcal{M}(1)}\Pr(X_k=I\mid X_1=\{v\})\Pr(X_1=\{v\})\\ &\quad=\Pr(X_k=I)\sum_{v\in\mathcal{M}(1)}\Pr(X_1=\{v\}\mid X_k=I)\qquad(31)\\ &\quad=\Pr(X_k=I).\end{aligned}$$

Noticing that $\Pr(X_k=I\mid X_1=\{v\})=\pi_{v,k-1}(I)$, equation (31) recovers (25).

By recalling (22) and (23), we observe that the analysis of the "going-down" half—and, similarly, the "going-up" half—of $P_k^{\vee}$ and $P_{k-1}^{\wedge}$ corresponds to premultiplying by $P_{k-1}^{\uparrow}$ – or, accordingly, $P_k^{\downarrow}$– to a function f. Hence, Lemma 11 implies that the relative entropy contracts by $1-\frac{1}{k}$ in the "going-down" half of the random walks $P_k^{\vee}$ and $P_{k-1}^{\wedge}$. What we show next is that the other half will not increase the relative entropy; a fact which is a special case of the so-called "data processing inequality."

LEMMA 12. *For any $k\geq 2$ and $f:\mathcal{M}(k-1)\to\mathbb{R}_{\geq 0}$,*

$$\operatorname{Ent}_{\pi_k}(P_k^{\downarrow}f)\leq\operatorname{Ent}_{\pi_{k-1}}(f).\qquad(32)$$

PROOF. Firstly, we verify that

$$\begin{aligned}\mathbb{E}_{\pi_k}P_k^{\downarrow}f&=\pi_kP_k^{\downarrow}f\\ &=\pi_{k-1}f=\mathbb{E}_{\pi_{k-1}}f.\qquad\text{(by Equation (20))}\end{aligned}$$

Thus, we can assume both are 1 without loss of generality. Then

$$\begin{aligned}\operatorname{Ent}_{\pi_k}(P_k^{\downarrow}f)&=\pi_k(P_k^{\downarrow}f\odot\log P_k^{\downarrow}f)\\ &\leq\pi_kP_k^{\downarrow}(f\odot\log f)\qquad\text{(by Jensen's inequality on } x\log x)\\ &=\pi_{k-1}(f\odot\log f)\qquad\text{(by Equation (20))}\\ &=\operatorname{Ent}_{\pi_{k-1}}(f),\end{aligned}$$

where $\odot$ stands for the Hadamard product. $\square$

With Lemmas 11 and 12 in hand, we can show the decay of relative entropy for $P_k^{\vee}$ and $P_k^{\wedge}$.

COROLLARY 13. *For any distribution τ on $\mathcal{M}(k)$,*

- *if $2\leq k\leq r$, then $D(\tau P_k^{\vee}\|\pi_k)\leq(1-\frac{1}{k})D(\tau\|\pi_k)$;*
- *if $1\leq k\leq r-1$, then $D(\tau P_k^{\wedge}\|\pi_k)\leq(1-\frac{1}{k+1})D(\tau\|\pi_k)$.*

PROOF. We will only prove this corollary for $P_k^\vee$ as the case of $P_k^\wedge$ is similar. We have that $D(\tau\|\pi_k) = \mathrm{Ent}_{\pi_k}(D_k^{-1}\tau^{\mathrm{T}})$ where $D_k := \mathrm{diag}(\pi_k)$. Since $P_k^\vee$ is reversible, $D_k^{-1}(P_k^\vee)^{\mathrm{T}} = P_k^\vee D_k^{-1}$. Therefore,

$$\begin{aligned}
& D(\tau P_k^\vee\|\pi_k) = \mathrm{Ent}_{\pi_k}(D_k^{-1}(P_k^\vee)^{\mathrm{T}}\tau^{\mathrm{T}}) = \mathrm{Ent}_{\pi_k}(P_k^\vee D_k^{-1}\tau^{\mathrm{T}}) \\
\text{(by Lemma 12)} \qquad & \le \mathrm{Ent}_{\pi_{k-1}}(P_{k-1}^\uparrow D_k^{-1}\tau^{\mathrm{T}}) \\
\text{(by Lemma 11)} \qquad & \le \left(1-\frac{1}{k}\right)\mathrm{Ent}_{\pi_k}(D_k^{-1}\tau^{\mathrm{T}}) \\
& = \left(1-\frac{1}{k}\right)D(\tau\|\pi_k). \qquad \square
\end{aligned}$$

It is well known that the decay of relative entropy implies a mLSI.

PROOF OF THEOREM 7. Given any $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\ge 0}$ such that $\mathbb{E}_{\pi_k} f^{(k)} = 1$, let $\tau = (D_k f^{(k)})^{\mathrm{T}}$ be the distribution corresponding to $f^{(k)}$. Then

$$\begin{aligned}
& D(\tau\|\pi_k) - D(\tau P_k^\vee\|\pi_k) \\
& = \sum_{S\in\mathcal{M}(k)} \tau(S)\log\left(\frac{\tau(S)}{\pi_k(S)}\right) - \sum_{S\in\mathcal{M}(k)} \tau P_k^\vee(S)\log\left(\frac{\tau P_k^\vee(S)}{\pi_k(S)}\right) \\
& = \sum_{S\in\mathcal{M}(k)} [\tau(\mathbf{I}-P_k^\vee)](S)\log\left(\frac{\tau(S)}{\pi_k(S)}\right) - \sum_{S\in\mathcal{M}(k)} \tau P_k^\vee(S)\log\left(\frac{\tau P_k^\vee(S)}{\tau(S)}\right) \\
& = \mathcal{E}_{P_k^\vee}(f^{(k)}, \log f^{(k)}) - D(\tau P_k^\vee\|\tau) \le \mathcal{E}_{P_k^\vee}(f^{(k)}, \log f^{(k)}).
\end{aligned}$$

Thus,

$$\begin{aligned}
& \mathcal{E}_{P_k^\vee}(f^{(k)}, \log f^{(k)}) \ge D(\tau\|\pi_k) - D(\tau P_k^\vee\|\pi_k) \\
\text{(by Corollary 13)} \qquad & \ge \frac{1}{k}D(\tau\|\pi_k) = \frac{1}{k}\mathrm{Ent}_{\pi_k}(D_k^{-1}\tau^{\mathrm{T}}) \\
& = \frac{1}{k}\mathrm{Ent}_{\pi_k}(f^{(k)}).
\end{aligned}$$

This proves the statement for $P_k^\vee$. The same proof can be used for $P_k^\wedge$ by replacing every occurrence of $P_k^\vee$ with $P_k^\wedge$, and the factor $\frac{1}{k}$ with $\frac{1}{k+1}$. $\square$

In fact, the contraction of relative entropy (Corollary 13) directly implies the mixing time bound of Corollary 8, as illustrated by the following.

A DIRECT PROOF OF COROLLARY 8. We will only prove this for $P_k^\vee$; the case of $P_k^\wedge$ is similar. Notice that Corollary 13 implies that

$$\begin{aligned}
D(\tau_0(P_k^\vee)^t\|\pi_k) & \le \left(1-\frac{1}{k}\right)^t D(\tau_0\|\pi_k) \\
& \le e^{-t/k}D(\tau_0\|\pi_k) = e^{-t/k}\log \pi_k(x_0)^{-1},
\end{aligned}$$

where τ_0 is the initial distribution with $\tau_0(x_0) = 1$ for some $x_0 \in \mathcal{M}(k)$. Then we use Pinsker's inequality ($2\|\tau-\sigma\|_{\mathrm{TV}}^2 \le D(\tau\|\sigma)$ for any two distributions τ, σ on the same state space), to show

$$2\|\tau_0(P_k^\vee)^t - \pi_k\|_{\mathrm{TV}}^2 \le D(\tau_0(P_k^\vee)^t\|\pi_k).$$

Setting $e^{-t/k}\log\pi_k(x_0)^{-1}\le 2\epsilon^2$, we conclude that

$$\|\tau_0(P_k^{\vee})^t-\pi_k\|_{\mathrm{TV}}\le\epsilon,$$

whenever

$$t\ge k\left(\log\log\pi_k(x_0)^{-1}+\log\frac{1}{2\epsilon^2}\right).$$

This gives us Corollary 8 for $P_k^{\vee}$. □

At the end of this section, let us comment that it is possible to prove the decay of variances similar to Lemma 11, with $\mathrm{Ent}(\cdot)$ replaced by $\mathrm{Var}(\cdot)$. This provides an alternative proof for the spectral gap of $P_{\mathrm{BX},\pi}$ to Anari et al. (2019), Kaufman and Oppenheim (2018). Indeed, the induction proof of Lemma 11 does not require any change when one replaces $\mathrm{Ent}(\cdot)$ by $\mathrm{Var}(\cdot)$, as both of them obey the same decomposition rules. However, the base case (namely Lemma 10) needs to be edited as follows.

LEMMA 14. *Let $f^{(2)}:\mathcal{M}(2)\to\mathbb{R}$. Then*

$$\mathrm{Var}_{\pi_2}(f^{(2)})\ge 2\mathrm{Var}_{\pi_1}(f^{(1)}).$$

PROOF. We begin by observing that

$$\mathrm{diag}(w_1)\big(2P_1^{\wedge}-\mathbf{I}\big)=W_{\varnothing}. \tag{33}$$

From this identity and Proposition 5, we deduce that the symmetric matrix $\mathrm{diag}(w_1)(2P_1^{\wedge}-\mathbf{I})$ has at most one positive eigenvalue. Premultiplying by the positive semidefinite matrix $\mathrm{diag}(w_1)^{-1}$, we get that $2P_1^{\wedge}-\mathbf{I}$ also has at most one positive eigenvalue (see, e.g., Anari et al. (2019), Lemma 2.6). Furthermore, the spectra of $2P_1^{\wedge}-\mathbf{I}$ and $2P_2^{\vee}-\mathbf{I}$ are the same up to some extra -1's. So, if $|\mathcal{M}(2)|\ge 2$ (otherwise the lemma holds trivially), $\lambda_2(P_2^{\vee})\le 1/2$ where λ_2 is the second largest eigenvalue. Then the spectral gap $\lambda(P_2^{\vee})=1-\lambda_2(P_2^{\vee})\ge 1/2$, which means that

$$\mathcal{E}_{P_2^{\vee}}(f^{(2)},f^{(2)})\ge\frac{1}{2}\mathrm{Var}_{\pi_2}(f^{(2)}).$$

However, this is equivalent to the statement of the lemma, as can be seen by the following equalities:

$$\begin{aligned}
\mathrm{Var}_{\pi_1}(f^{(1)})&=(f^{(1)})^{\mathrm{T}}D_1f^{(1)}-(\mathbb{E}_{\pi_1}f^{(1)})^2\\
\text{(by Lemma 9)}\qquad&=(f^{(2)})^{\mathrm{T}}(P_1^{\uparrow})^{\mathrm{T}}D_1P_1^{\uparrow}f^{(2)}-(\mathbb{E}_{\pi_2}f^{(2)})^2\\
\text{(by (19))}\qquad&=(f^{(2)})^{\mathrm{T}}D_2P_2^{\vee}f^{(2)}-(\mathbb{E}_{\pi_2}f^{(2)})^2\\
&=\mathrm{Var}_{\pi_2}(f^{(2)})-\mathcal{E}_{P_2^{\vee}}(f^{(2)},f^{(2)}).
\end{aligned}$$

□

5. Concentration. One application of the modified log-Sobolev inequalities is to show concentration inequalities, via the Herbst argument (see, e.g., Bobkov and Tetali (2006), Boucheron, Lugosi and Massart (2013)). In the discrete setting, concentration inequalities have been obtained by Goel (2004), Section 5 and can also be obtained by combining various results by Bobkov and Götze (1999), Bobkov, Houdré and Tetali (2006), Sammer (2005). The following lemma and its proof are a small modification of Hermon and Salez (2019), Lemma 5. For completeness, we include all details.

LEMMA 15. *Let P be the transition matrix of a reversible Markov chain with stationary distribution π on a finite set Ω, and $f : \Omega \to \mathbb{R}$ be some observable function. Then*

$$\Pr_{x\sim\pi}\big(f(x) - \mathbb{E}_\pi f \geq a\big) \leq \exp\left(-\frac{\rho(P)a^2}{2v(f)}\right),$$

where $a \geq 0$ and

$$v(f) := \max_{x\in\Omega}\left\{\sum_{y\in\Omega} P(x,y)\big(f(x) - f(y)\big)^2\right\}.$$

PROOF. For any $x \in \Omega$ and $t \in (0, +\infty)$, let

$$F_t(x) := \exp\big(tf(x) - ct^2\big),$$

where $c := \frac{v(f)}{2\rho(P)}$. We will use the inequality

$$z\big(e^z + 1\big) \geq 2\big(e^z - 1\big), \tag{34}$$

which holds for $z \geq 0$. To see this, notice that at $z = 0$ the equality holds, and for $z > 0$ the derivative of the left is larger than that of the right.

If $f(x) \geq f(y)$, we set $z = t(f(x) - f(y))$ in (34) and obtain

$$t\big(f(x) - f(y)\big)\big(F_t(x) + F_t(y)\big) \geq 2\big(F_t(x) - F_t(y)\big),$$

which in turn implies that

$$\big(F_t(x) - F_t(y)\big)\big(f(x) - f(y)\big) \leq \frac{t}{2}\big(F_t(x) + F_t(y)\big)\big(f(x) - f(y)\big)^2. \tag{35}$$

Notice that (35) also holds even if $f(x) < f(y)$ by swapping x and y. Thus, we have that

$$\begin{aligned}
&\text{(by (5))} & \mathcal{E}_P(F_t, \log F_t) &= \frac{t}{2}\sum_{x\in\Omega}\sum_{y\in\Omega}\pi(x)P(x,y)\big(F_t(x) - F_t(y)\big)\big(f(x) - f(y)\big) \\
&\text{(by (35))} & &\leq \frac{t^2}{4}\sum_{x\in\Omega}\sum_{y\in\Omega}\pi(x)P(x,y)\big(F_t(x) + F_t(y)\big)\big(f(x) - f(y)\big)^2 \\
&\text{(by the reversibility of } P) & &= \frac{t^2}{2}\sum_{x\in\Omega}\pi(x)F_t(x)\sum_{y\in\Omega}P(x,y)\big(f(x) - f(y)\big)^2 \\
& & &\leq \frac{t^2}{2}v(f)\mathbb{E}_\pi F_t.
\end{aligned}$$

This, together with $\mathcal{E}_P(F_t, \log F_t) \geq \rho(P)\mathrm{Ent}_\pi(F_t)$ (recall the definition of $\rho(P)$), yields

$$\mathrm{Ent}_\pi(F_t) \leq ct^2\mathbb{E}_\pi F_t.$$

By noticing that

$$\frac{d}{dt}\left(\frac{\log \mathbb{E}_\pi F_t}{t}\right) = \frac{\mathrm{Ent}_\pi(F_t) - ct^2\mathbb{E}_\pi F_t}{t^2\mathbb{E}_\pi F_t} \leq 0,$$

we deduce that for any $t > 0$,

$$\frac{\log \mathbb{E}_\pi F_t}{t} \leq \lim_{h\to 0^+}\frac{\log \mathbb{E}_\pi F_h}{h} = \mathbb{E}_\pi f,$$

or equivalently,

$$\mathbb{E}_\pi F_t \leq \exp(t\mathbb{E}_\pi f).$$

Finally, by Markov inequality, for any $a \geq 0$,

$$\begin{aligned}\Pr_{x\sim\pi}\big(f(x)-\mathbb{E}_\pi f \geq a\big) &= \Pr_{x\sim\pi}\big(F_t(x) \geq \exp\big(t\mathbb{E}_\pi f - ct^2 + at\big)\big)\\ &\leq \exp\big(ct^2 - at\big),\end{aligned}$$

where the right-hand side is minimized for $t = \frac{a}{2c} = \frac{a\rho(P)}{v(f)}$. $\square$

Corollary 2 follows from applying Lemma 15 to both f and $-f$ together with Theorem 1. We could also apply Lemma 15 together with Theorem 7 to get concentration inequalities for all π_k.

For a Lipschitz function $f : \Omega \to \mathbb{R}$ with Lipschitz constant c (under the graph distance in the bases-exchange graph), we have that $v(f) \leq c^2$. Thus, by Corollary 2, such a Lipschitz function satisfies the following concentration inequality:

$$\Pr_{x\sim\pi}\big(\big|f(x)-\mathbb{E}_\pi f\big| \geq a\big) \leq 2\exp\left(-\frac{a^2}{2rc^2}\right),$$

when π is an r-homogeneous strongly log-concave distribution.

For general matroids, an example is the function that counts the number of elements belonging to a specified subset of the ground set, which has Lipschitz constant $c = 1$. More examples were given by Pemantle and Peres (2014) for graphic matroids, such as functions that count the number of leaves in a spanning tree ($c = 2$), or the number of vertices with odd degrees ($c = 4$).

APPENDIX: STOCHASTIC COVERING PROPERTY AND STRONG LOG-CONCAVITY

The results obtained by Pemantle and Peres (2014) and Hermon and Salez (2019) only require a property which is weaker than the strong Rayleigh property (**SRP**), namely the *stochastic covering property* (**SCP**). Since strong log-concavity (**SLC**) is also a generalization of **SRP**, it is natural to wonder about the relationship between **SLC** and **SCP**. In this section, we show that **SLC** is incomparable to **SCP**. As a result, Theorem 1 and Corollary 2 do not subsume the results of Hermon and Salez (2019) and Pemantle and Peres (2014), respectively. Moreover, Corollary 2 is also incomparable to the concentration bound of Garbe and Vondrák (2018), whose result requires only *negative regression*, a property weaker than **SCP**.

First, let us define **SCP**. For $S \subseteq [n]$ and $x, y \in \{0,1\}^S$, we say x *covers* y, denoted by $x \rhd y$, if $x = y$ or $x = y + \mathbf{e}_i$ for some i, where $\mathbf{e}_i$ is the unit vector of coordinate i. In other words, x is obtained from y by increasing at most one coordinate. For two distributions μ and ν, we say μ *stochastically covers* ν, if there is a coupling such that $\Pr_{X\sim\mu, Y\sim\nu}(X \rhd Y) = 1$. With slight overload of notation, we also write $\mu \rhd \nu$. A distribution $\mu : \{0,1\}^{[n]} \to \mathbb{R}_{\geq 0}$ satisfies the **SCP** if for any $S \subseteq [n]$ and $x, y \in \{0,1\}^S$ such that $x \rhd y$, $\mu_y \rhd \mu_x$, where μ_x is the distribution of μ conditioned on agreeing with x over the index set S.

Furthermore, μ is said to satisfy the *negative cylinder dependence* (**NCD**), if for any $S \subseteq [n]$,

$$\mathbb{E}\prod_{i\in S} X_i \leq \prod_{i\in S}\mathbb{E}X_i,$$

$$\mathbb{E}\prod_{i\in S}(1-X_i) \leq \prod_{i\in S}\mathbb{E}(1-X_i),$$

where X_i is the indicator variable of coordinate i. It is known that **SCP** implies **NCD** (Pemantle and Peres (2014)). However, such negative dependence even when $|S| = 2$ is known *not* to hold for the uniform distribution over the bases of some matroids. See Huh, Schröter and Wang (2018) for the most comprehensive list of such examples that we are aware of. As the uniform distribution over a matroid's bases is **SLC**, **SLC** does not imply **SCP**.

On the other hand, **SCP** does not imply **SLC** either. We give a concrete example here. Let μ be supported on the bases of the uniform matroid of rank 2 over 4 elements. We choose μ such that

$$\mu(\{1,1,0,0\}) \propto \theta, \qquad \mu(\{1,0,1,0\}) \propto 2, \qquad \mu(\{1,0,0,1\}) \propto 1,$$
$$\mu(\{0,1,1,0\}) \propto 1, \qquad \mu(\{0,1,0,1\}) \propto 1, \qquad \mu(\{0,0,1,1\}) \propto 1.$$

It is straightforward to verify that if $0 \le \theta < 3-2\sqrt{2} \approx 0.17157$ or $\theta > 3+2\sqrt{2} \approx 5.82843$, then **SLC** fails. However, **SCP** holds as long as $0 \le \theta \le 6$. To see the latter claim, first verify that the distribution conditioned on choosing any $i \in [4]$ stochastically dominates the one conditioned on not choosing i. Then notice that in a homogeneous distribution, such stochastic dominance is the same as stochastic covering.

Here is some insight on how to find an example such as the above. When the generating polynomial g_μ is homogeneous and quadratic, it is **SLC** if and only if it has the **SRP** (Brändén and Huh (2019)), which in turn is equivalent to the following condition as $g_\mu \in \mathbb{R}[x_1, \dots, x_n]$ is multiaffine:

$$\frac{\partial}{x_i} g_\mu(\mathbf{x}) \cdot \frac{\partial}{x_j} g_\mu(\mathbf{x}) \ge g_\mu(\mathbf{x}) \cdot \frac{\partial^2}{\partial x_i \, \partial x_j} g_\mu(\mathbf{x}), \tag{36}$$

for any $i, j \in [n]$ and $\mathbf{x} \in \mathbb{R}^n$; see Brändén (2007). If we plug in $\mathbf{x} = \mathbf{1}$, then (36) becomes negative dependence for a pair of variables, which is a special case of **NCD** and thus a necessary condition for **SCP**. In our example, we choose μ so that (36) holds for $\mathbf{x} = \mathbf{1}$ but not for an arbitrary $\mathbf{x} \in \mathbb{R}^n$. It turns out that our choice is also sufficient for **SCP** in this particular setting.

Acknowledgments. We thank Mark Jerrum and Prasad Tetali for their helpful comments, and Arthur Sinulis for pointing out a saving of a factor 2 in Corollary 2. We also thank the anonymous referees whose comments have resulted in an improvement of the presentation.

Part of the work was done while HG was visiting the Simons institute for the theory of computing in the University of California, Berkeley.

REFERENCES

ALEV, V. and LAU, L. (2020). Improved analysis of higher order random walks and applications. In *STOC* 1198–1211. ACM, New York.

ANARI, N., OVEIS GHARAN, S. and REZAEI, A. (2016). Monte Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes. In *COLT* **49** 103–115.

ANARI, N., OVEIS GHARAN, S. and VINZANT, C. (2018). Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids. In 59*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2018 35–46. IEEE Computer Soc., Los Alamitos, CA. MR3899575 https://doi.org/10.1109/FOCS.2018.00013

ANARI, N., LIU, K., GHARAN, S. O. and VINZANT, C. (2019). Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. In *STOC'*19*—Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 1–12. ACM, New York. MR4003314 https://doi.org/10.1145/3313276.3316385

BOBKOV, S. G. and GÖTZE, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. MR1682772 https://doi.org/10.1006/jfan.1998.3326

BOBKOV, S. G., HOUDRÉ, C. and TETALI, P. (2006). The subgaussian constant and concentration inequalities. *Israel J. Math.* **156** 255–283. MR2282379 https://doi.org/10.1007/BF02773835

Bobkov, S. G. and Tetali, P. (2006). Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.* **19** 289–336. MR2283379 https://doi.org/10.1007/s10959-006-0016-3

Borcea, J., Brändén, P. and Liggett, T. M. (2009). Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.* **22** 521–567. MR2476782 https://doi.org/10.1090/S0894-0347-08-00618-8

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

Brändén, P. (2007). Polynomials with the half-plane property and matroid theory. *Adv. Math.* **216** 302–320. MR2353258 https://doi.org/10.1016/j.aim.2007.05.011

Brändén, P. and Huh, J. (2019). Lorentzian polynomials. Available at https://arxiv.org/abs/1902.03719.

Diaconis, P. and Saloff-Coste, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750. MR1410112 https://doi.org/10.1214/aoap/1034968224

Dubhashi, D. and Ranjan, D. (1998). Balls and bins: A study in negative dependence. *Random Structures Algorithms* **13** 99–124. MR1642566 https://doi.org/10.1002/(SICI)1098-2418(199809)13:2<99::AID-RSA1>3.0.CO;2-M

Feder, T. and Mihail, M. (1992). Balanced matroids. In *STOC* 26–38.

Garbe, K. and Vondrák, J. (2018). Concentration of Lipschitz functions of negatively dependent variables. Available at https://arxiv.org/abs/1804.10084.

Goel, S. (2004). Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Process. Appl.* **114** 51–79. MR2094147 https://doi.org/10.1016/j.spa.2004.06.001

Gurvits, L. (2009a). A polynomial-time algorithm to approximate the mixed volume within a simply exponential factor. *Discrete Comput. Geom.* **41** 533–555. MR2496316 https://doi.org/10.1007/s00454-009-9147-5

Gurvits, L. (2009b). On multivariate Newton-like inequalities. In *Advances in Combinatorial Mathematics* 61–78. Springer, Berlin. MR2683227 https://doi.org/10.1007/978-3-642-03562-3_4

Hermon, J. and Salez, J. (2019). Modified log-Sobolev inequalities for strong-Rayleigh measures. Available at https://arxiv.org/abs/1902.02775.

Huh, J., Schröter, B. and Wang, B. (2018). Correlation bounds for fields and matroids. Available at https://arxiv.org/abs/1806.02675.

Jerrum, M. (2003). *Counting, Sampling and Integrating: Algorithms and Complexity. Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR1960003 https://doi.org/10.1007/978-3-0348-8005-3

Jerrum, M., Sinclair, A. and Vigoda, E. (2004). A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM* **51** 671–697. MR2147852 https://doi.org/10.1145/1008731.1008738

Jerrum, M., Son, J.-B., Tetali, P. and Vigoda, E. (2004). Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.* **14** 1741–1765. MR2099650 https://doi.org/10.1214/105051604000000639

Kaufman, T. and Oppenheim, I. (2018). High order random walks: Beyond spectral gap. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. LIPIcs. Leibniz Int. Proc. Inform.* **116** Art. No. 47, 17. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3857285

Lee, T.-Y. and Yau, H.-T. (1998). Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.* **26** 1855–1873. MR1675008 https://doi.org/10.1214/aop/1022855885

Levin, D. A. and Peres, Y. (2017). *Markov Chains and Mixing Times*, 2nd ed. Amer. Math. Soc., Providence, RI. MR3726904

Morris, B. (2009). Improved mixing time bounds for the Thorp shuffle and L-reversal chain. *Ann. Probab.* **37** 453–477. MR2510013 https://doi.org/10.1214/08-AOP409

Morris, B. (2013). Improved mixing time bounds for the Thorp shuffle. *Combin. Probab. Comput.* **22** 118–132. MR3002577 https://doi.org/10.1017/S0963548312000478

Oxley, J. G. (1992). *Matroid Theory. Oxford Science Publications*. Oxford University Press, New York. MR1207587

Pemantle, R. and Peres, Y. (2014). Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. *Combin. Probab. Comput.* **23** 140–160. MR3197973 https://doi.org/10.1017/S0963548313000345

Sammer, M. D. (2005). Aspects of mass transportation in discrete concentration inequalities. Ph.D. thesis, Georgia Institute of Technology. MR2707115