Modified log-Sobolev inequalities for strongly log-concave distributions

Mary Cryan, Heng Guo, Giorgos Mousa
School of Informatics
University of Edinburgh
Edinburgh, United Kingdom
Email: mcryan@inf.ed.ac.uk, hguo@inf.ed.ac.uk, g.mousa@ed.ac.uk

Abstract—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 an asymptotically optimal mixing time bound for the bases-exchange walk for matroids, and a concentration bound for Lipschitz functions over these distributions.

Keywords-Matroids; bases-exchange walk; Markov chains; mixing time; concentration of measure;

I. 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 [1], [2]. This discovery leads to the breakthrough result that the exchange walk over the bases of a matroid is rapidly mixing [1], 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.[1] This chain is irreducible and it converges to the uniform distribution over the bases of a matroid. As shown in [2], 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 $P_{\text{BX},\pi}$ 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,

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

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 strongly log-concave distribution π,

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

where $\pi_{\min} = \min_{x \in \Omega} \pi(x)$. This will improve the previous bound $t_{\text{mix}}(P_{\text{BX},\pi}, \varepsilon) \leq r \left(\log \frac{1}{\pi_{\min}} + \log \frac{1}{\varepsilon} \right)$ due to [1]. 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 bound is asymptotically optimal without further assumptions, as the upper bound is achieved when π is the uniform distribution over the bases of some matroids [3].[2]

Our main improvement is a modified log-Sobolev inequality for π and $P_{\text{BX},\pi}$. 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 (normalised) relative entropy of $f : \Omega \to \mathbb{R}_{\geq 0}$ be

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

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

---

[1]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.

[2]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 $P_{\text{BX},\pi}$ in this case is just the 1/2-lazy random walk on the Boolean hypercube.

IEEE computer society

The modified log-Sobolev constant [4] is defined as

$$\rho(P)$$
$$:= \inf \left\{ \frac{\mathcal{E}_P(f, \log f)}{\text{Ent}_\pi(f)} \mid f : \Omega \to \mathbb{R}_{\geq 0}, \ \text{Ent}_\pi(f) \neq 0 \right\}.$$

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 $P_{\text{BX},\pi}$ is the corresponding bases-exchange walk. Then*

$$\rho(P_{\text{BX},\pi}) \geq \frac{1}{r}.$$

Since $t_{\text{mix}}(P, \varepsilon) \leq \frac{1}{\rho(P)} \left(\log \log \frac{1}{\pi_{\min}} + \log \frac{1}{2\varepsilon^2} \right)$ (cf. [4]), Theorem 1 directly implies the mixing time bound (1).

In fact, we show more than Theorem 1. Following [1] and [5], 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 $P_{\text{BX},\pi}$ 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 III and Theorem 7.

The previous work in [1], building upon [5], focuses on the spectral gap of $P_{\text{BX},\pi}$. 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 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 [6], [7], [8], [9], 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 in [10], are a proper subset of strongly log-concave distributions. The authors of [11] showed a lower bound on the modified log-Sobolev constant for strong Rayleigh distributions,[3] improving upon the spectral gap bound obtained in [12]. The work of [11] builds upon the previous work of [7] for balanced matroids [13]. All of these results follow an inductive framework inspired by [6], 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.

[3]The result of [11] in fact requires a weaker assumption, namely the *stochastic covering property* (**SCP**). We construct examples in Section A to show that **SCP** and **SLC** are in fact incomparable.

By the standard Herbst argument (see, e.g., [14], [15], [16]), 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 $P_{\text{BX},\pi}$ be the corresponding bases-exchange walk. For any observable function $f : \Omega \to \mathbb{R}$ and $a \geq 0$,*

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

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

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

There have been a number of results concerning concentration inequalities for Lipschitz functions of negatively correlated variables. The authors of [17] showed concentration for variables satisfying the *stochastic covering property* (**SCP**), which includes strong Rayleigh distributions as special cases. (See also [11].) Correcting an earlier proof in [18], the authors of [19] 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 generalises the concentration bound for Lipschitz functions in strong Rayleigh distributions. However, **SLC** is *not* a negative correlation property. We construct examples in Section A to show that **SCP** and **SLC** are in fact incomparable. Thus, Corollary 2 is incomparable to the results of [17], [11], [19]. It is not clear whether there is a larger class of distributions, generalising both **NR** and **SLC**, which retains this concentration bound.

In Section II we introduce necessary notions and briefly review relevant background. In Section III we formally state our main results. In Section IV we prove modified log-Sobolev constant lower bounds for the "down-up" walk. In Section V we deal with the "up-down" walk. In Section VI we show the concentration bound. In the Appendix we discuss stochastic covering property and strong log-concavity.

II. PRELIMINARIES

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

A. 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,

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

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

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

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

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

but we will not need this expression until Section VI. 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 characterisation. Let the variance of f be

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

Then

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

The usefulness of $\lambda(P)$ is due to the fact that

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

where $\pi_{\min} = \min_{x \in \Omega} \pi(x)$. See, for example, ([20], 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 non-negative 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

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

The constant $\alpha(P)$ gives a better control of the mixing time of P, as shown in [21],

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

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 in [4]:

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

Similar to (8), we have that

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

as shown in ([4], 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, ([4], 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 in [11], by taking the indicator function $\frac{1}{\pi(x)}\mathbf{1}_x$ 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 $\text{Ent}_\pi(f)$ 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 $\text{Ent}_\pi(f) = \mathbb{E}_\pi(f \log f)$. In this case, $\pi(\cdot)f(\cdot)$ is a distribution, and $\text{Ent}_\pi(f)$ is the relative entropy (or Kullback–Leibler divergence) between $\pi(\cdot)f(\cdot)$ and $\pi(\cdot)$.

B. Strongly log-concave distributions

We write ∂_i as shorthand for $\frac{\partial}{\partial x_i}$, and ∂_I for an index set $I = \{i_1, \ldots, i_k\}$ as shorthand for $\partial_{i_1} \ldots \partial_{i_k}$.

Definition 3. *A polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ with non-negative coefficients is* log-concave *at $\mathbf{x} \in \mathbb{R}_{\geq 0}$ if its Hessian $\nabla^2 \log p$ is negative semi-definite 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 in [22], [23]. There are also notions of *complete log-concavity* introduced in [24], and *Lorentzian* polynomials introduced in [2]. It turns out that for homogeneous polynomials the three notions are equivalent ([2], Theorem 5.3). (See also [1].)

The following property of strongly log-concave polynomials is particularly useful [24], [2].

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 semi-definite [24], but we will only need the proposition above.

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

C. 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:

- $\emptyset \in \mathcal{I}$;
- 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 non-empty, the second implies that $\mathcal{I}$ is downward closed, and the third is usually called the augmentation axiom. We direct the reader to [25] 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 $1 \leq k \leq r$. If we dropped the augmentation axiom, the resulting structure would be a non-empty collection of subsets of E that is downward closed, known as an (abstract) *simplicial complex*.

In ([2], Theorem 7.1) it was shown 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:[4]

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

for some normalisation 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

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

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 normalisation constant of π_k. In fact, for any $0 \leq k \leq r$, $k! Z_k = Z_0 = w(\emptyset)$.

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

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

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

---

[4]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 [1], [5], so we do not adopt it.

distributions $\pi_{I,k}$ for $k \leq r - |I|$ 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 $J \in \mathcal{M}(k+|I|)$,

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

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

If $|I| \leq r - 2$, 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

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

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*

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

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 $D = \text{diag}(\mathbf{w}_I)$ 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., [1], Lemma 2.4). We may decompose A as

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

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

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

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,

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

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. ∎

III. 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

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

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 $2 \leq k \leq r$,

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

Thus, the bases-exchange walk $P_{\text{BX}, \pi}$ 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 $2 \leq k \leq r$, $\rho(P_k^\vee) \geq \frac{1}{k}$;*
- *for any $1 \leq k \leq r - 1$, $\rho(P_k^\wedge) \geq \frac{1}{k+1}$.*

The first part of Theorem 7 is shown by Corollary 12, and the second part by Lemma 14. 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 $2 \leq k \leq r$,*
  $t_{\text{mix}}(P_k^\vee, \varepsilon) \leq k \left(\log \log \pi_{k, \min}^{-1} + \log \frac{1}{2\varepsilon^2} \right)$;
- *for any $1 \leq k \leq r - 1$,*
  $t_{\text{mix}}(P_k^\wedge, \varepsilon) \leq (k+1) \left(\log \log \pi_{k, \min}^{-1} + \log \frac{1}{2\varepsilon^2} \right)$.

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

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

For example, for the uniform distribution over bases of matroids, 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 [1]. The mixing time bound in Corollary 8 is asymptotically optimal, as it is achieved for the bases of some matroids ([3], Ex. 9.14). As mentioned in the introduction, 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 rank of this matroid is n, and $\pi_{\min} = \frac{1}{2^n}$. The Markov chain $P_{\mathrm{BX},\pi}$ in this case is just the 1/2-lazy random walk on the n-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 VI.

IV. The down-up walk

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

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

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

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

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

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 normalise $\mathbb{E}_{\pi_k}f^{(k)} = 1$, then $(f^{(k)})^{\mathsf{T}}D_k$ is a distribution (viewed as a row vector). Then, it is easy to verify that

$$\left(f^{(k-1)}\right)^{\mathsf{T}}D_{k-1} = \left(f^{(k)}\right)^{\mathsf{T}}D_k P_k^\downarrow.$$

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 non-negative function on $\mathcal{M}(k)$ such that $\mathbb{E}_{\pi_k}f^{(k)} = 1$. 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 $1 \leq i \leq k$, $\mathbb{E}_{\pi_i}f^{(i)} = 1$.*

Proof: For (1), first notice that

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

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

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

For (2), we have that

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

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*

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

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 $b > 0$,

$$a \log \frac{a}{b} \geq a - b. \tag{23}$$

Noticing that $Z_1 = 2Z_2$, we have

$$\text{Ent}_{\pi_2}\left(f^{(2)}\right) - 2\text{Ent}_{\pi_1}\left(f^{(1)}\right)$$

$$= \sum_{S \in \mathcal{M}(2)} \pi_2(S) f^{(2)}(S) \log f^{(2)}(S)$$

$$-2 \sum_{v \in E} \pi_1(v) \left(\sum_{S \in \mathcal{M}(2): v \in S} \frac{w(S)}{w(v)} f^{(2)}(S) \right) \log f^{(1)}(v)$$

$$= \sum_{S \in \mathcal{M}(2)} \Big(\pi_2(S) f^{(2)}(S) \log f^{(2)}(S)$$

$$-2 \sum_{v \in S} \pi_1(v) \frac{w(S)}{w(v)} f^{(2)}(S) \log f^{(1)}(v) \Big)$$

$$= \sum_{S \in \mathcal{M}(2)} \Big(\frac{w(S)}{Z_2} f^{(2)}(S) \log f^{(2)}(S)$$

$$-2 \sum_{v \in S} \frac{w(v)}{Z_1} \cdot \frac{w(S)}{w(v)} f^{(2)}(S) \log f^{(1)}(v) \Big)$$

$$= \sum_{S=\{u,v\} \in \mathcal{M}(2)} \frac{w(S)}{Z_2} f^{(2)}(S) \Big(\log f^{(2)}(S)$$

$$- \log f^{(1)}(v) - \log f^{(1)}(u) \Big)$$

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

$$= \sum_{S \in \mathcal{M}(2)} \pi_2(S) f^{(2)}(S)$$

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

$$= 1 - \frac{1}{2Z_2} \cdot \left(f^{(1)} \right)^{\mathsf{T}} W_\emptyset f^{(1)},$$

where the inequality is by (23) 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 $I = \emptyset$ and $w(\emptyset) = Z_1 = 2Z_2$. ∎

We generalise Lemma 10 as follows.

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

$$\text{Ent}_{\pi_k}\left(f^{(k)}\right) \geq \frac{k}{k-1} \text{Ent}_{\pi_{k-1}}\left(f^{(k-1)}\right).$$

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 non-negative 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 implies that

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

$$= \sum_{v \in \mathcal{M}(1)} \pi_1(v)\pi_{v,k}(I). \tag{24}$$

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

We use the "chain rule" of entropy to decompose $\text{Ent}_{\pi_{k+1}}\left(f^{(k+1)}\right)$ 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

$$\text{Ent}_{\pi_{k+1}}\left(f^{(k+1)}\right)$$

$$= \sum_{v \in \mathcal{M}(1)} \pi_1(v)\text{Ent}_{\pi_{v,k}}\left(f^{(k+1)}\right)$$

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

$$= \sum_{v \in \mathcal{M}(1)} \pi_1(v)\text{Ent}_{\pi_{v,k}}\left(f^{(k+1)}\right) + \text{Ent}_{\pi_1}\left(f^{(1)}\right), \tag{25}$$

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

$$\text{Ent}_{\pi_k}\left(f^{(k)}\right) = \sum_{v \in \mathcal{M}(1)} \pi_1(v)\text{Ent}_{\pi_{v,k-1}}\left(f^{(k)}\right)$$

$$+ \text{Ent}_{\pi_1}\left(f^{(1)}\right). \tag{26}$$

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

$$\text{Ent}_{\pi_{v,k}}\left(f^{(k+1)}\right) \geq \frac{k}{k-1} \cdot \text{Ent}_{\pi_{v,k-1}}\left(f^{(k)}\right). \quad (27)$$

Strictly speaking, in (27) 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 (27).

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

$$\text{Ent}_{\pi_k}\left(f^{(k)}\right) \geq k \cdot \text{Ent}_{\pi_1}\left(f^{(1)}\right). \quad (28)$$

Thus, (26) and (28) together imply that

$$\sum_{v \in \mathcal{M}(1)} \pi_1(v) \text{Ent}_{\pi_{v,k-1}}\left(f^{(k)}\right) \geq (k-1) \text{Ent}_{\pi_1}\left(f^{(1)}\right). \quad (29)$$

Putting everything together,

$$\text{Ent}_{\pi_{k+1}}\left(f^{(k+1)}\right)$$
$$= \sum_{v \in \mathcal{M}(1)} \pi_1(v) \text{Ent}_{\pi_{v,k}}\left(f^{(k+1)}\right) + \text{Ent}_{\pi_1}\left(f^{(1)}\right),$$
$$\text{(by (25))}$$
$$\geq \frac{k}{k-1} \sum_{v \in \mathcal{M}(1)} \pi_1(v) \text{Ent}_{\pi_{v,k-1}}\left(f^{(k)}\right) + \text{Ent}_{\pi_1}\left(f^{(1)}\right),$$
$$\text{(by (27))}$$
$$= \left(\frac{k+1}{k} + \frac{1}{k(k-1)}\right) \sum_{v \in \mathcal{M}(1)} \pi_1(v) \text{Ent}_{\pi_{v,k-1}}\left(f^{(k)}\right)$$
$$+ \text{Ent}_{\pi_1}\left(f^{(1)}\right)$$
$$\geq \frac{k+1}{k} \sum_{v \in \mathcal{M}(1)} \pi_1(v) \text{Ent}_{\pi_{v,k-1}}\left(f^{(k)}\right)$$
$$+ \frac{k+1}{k} \text{Ent}_{\pi_1}\left(f^{(1)}\right) \quad \text{(by (29))}$$
$$= \frac{k+1}{k} \text{Ent}_{\pi_k}\left(f^{(k)}\right). \quad \text{(by (26))}$$

This concludes the inductive step and thus the proof. ∎

Remark. *We remark that our decomposition of the relative entropy* (25) *is "vertical" with respect to elements of $\mathcal{M}(1)$. This decomposition is different from the decomposition of [5] in a similar context, where they decompose "horizontally" across all levels.*

The decomposition (24) 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

$$\sum_{v \in \mathcal{M}(1)} \Pr(X_k = I \mid X_1 = \{v\}) \Pr(X_1 = \{v\})$$
$$= \Pr(X_k = I) \sum_{v \in \mathcal{M}(1)} \Pr(X_1 = \{v\} \mid X_k = I)$$
$$= \Pr(X_k = I). \quad (30)$$

Noticing that $\Pr(X_k = I \mid X_1 = \{v\}) = \pi_{v,k-1}(I)$, equation (30) recovers (24).

Lemma 11 implies that the relative entropy contracts by $1 - \frac{1}{k}$ in the first half of the random walk $P_k^\vee$. Since the second half of the random walk will not increase the relative entropy,[5] we have the following corollary.

Corollary 12. *For any $2 \leq k \leq r$,*

$$\rho(P_k^\vee) \geq \frac{1}{k}.$$

Proof: Given any $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\geq 0}$ such that $\mathbb{E}_{\pi_k} f^{(k)} = 1$, let $D_k = \text{diag}(\pi_k)$. Then we have

$$\text{Ent}_{\pi_{k-1}}\left(f^{(k-1)}\right) = \left(f^{(k-1)}\right)^\mathsf{T} D_{k-1} \log f^{(k-1)}$$
$$= \left(f^{(k)}\right)^\mathsf{T} \left(P_{k-1}^\uparrow\right)^\mathsf{T} D_{k-1} \log f^{(k-1)}$$
$$= \left(f^{(k)}\right)^\mathsf{T} D_k P_k^\downarrow \log P_{k-1}^\uparrow f^{(k)}$$
$$\text{(by (19))}$$
$$\geq \left(f^{(k)}\right)^\mathsf{T} D_k P_k^\downarrow P_{k-1}^\uparrow \log f^{(k)}$$
$$\text{(by Jensen's inequality)}$$
$$= \text{Ent}_{\pi_k}\left(f^{(k)}\right) - \mathcal{E}_{P_k^\vee}\left(f^{(k)}, \log f^{(k)}\right).$$

[5]This fact is shown as (∗) in our direct proof of Corollary 8 below.

Together with Lemma 11 we have that $\rho(P_k^\vee) \geq \frac{1}{k}$. ∎

In fact, the contraction of relative entropy (Lemma 11) directly implies the mixing time bound of Corollary 8, as illustrated by the following.

A direct proof for Corollary 8: We will only prove this for $P_k^\vee$; the case of $P_k^\wedge$ is similar. For any intermediate distribution τ on $\mathcal{M}(k)$, the relative entropy $D(\tau \| \pi_k) = \text{Ent}_{\pi_k}(D_k^{-1}\tau^\mathsf{T})$ where $D_k = \text{diag}(\pi_k)$. Moreover, after one step of the random walk, the distribution is $\tau P_k^\vee$. Since $P_k^\vee$ is reversible, $D_k^{-1}(P_k^\vee)^\mathsf{T} = P_k^\vee D_k^{-1}$. We claim that

$$
\begin{aligned}
D(\tau P_k^\vee \| \pi_k) &= \text{Ent}_{\pi_k}\left(D_k^{-1}(P_k^\vee)^\mathsf{T}\tau^\mathsf{T}\right) \\
&= \text{Ent}_{\pi_k}\left(P_k^\vee D_k^{-1}\tau^\mathsf{T}\right) \\
&= \text{Ent}_{\pi_k}\left(P_k^\downarrow P_{k-1}^\uparrow D_k^{-1}\tau^\mathsf{T}\right) \\
&\leq \text{Ent}_{\pi_{k-1}}\left(P_{k-1}^\uparrow D_k^{-1}\tau^\mathsf{T}\right) \quad\quad (*) \\
&\leq \left(1 - \frac{1}{k}\right)\text{Ent}_{\pi_k}\left(D_k^{-1}\tau^\mathsf{T}\right) \\
&\quad\quad\quad\quad\quad\quad\quad \text{(by Lemma 11)} \\
&= \left(1 - \frac{1}{k}\right)D(\tau \| \pi_k), \quad\quad (31)
\end{aligned}
$$

where $(*)$ is saying that the entropy will not increase after applying $P_k^\downarrow$. We defer our proof of $(*)$ to the end of this proof. To continue, notice that (31) implies that

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

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

$$
2\left\|\tau_0(P_k^\vee)^t - \pi_k\right\|_{\text{TV}}^2 \leq D\left(\tau_0(P_k^\vee)^t \| \pi_k\right).
$$

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

$$
\left\|\tau_0(P_k^\vee)^t - \pi_k\right\|_{\text{TV}} \leq \epsilon,
$$

whenever

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

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

We still need to prove $(*)$. We claim that

$$
\text{Ent}_{\pi_k}\left(P_k^\downarrow f\right) \leq \text{Ent}_{\pi_{k-1}}(f) \quad\quad (32)
$$

for any $k \geq 2$ and $f : \mathcal{M}(k-1) \to \mathbb{R}_{\geq 0}$. Firstly, we verify that

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

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

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

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

V. The Up-Down Walk

In this section we establish an analogous result of Corollary 12, namely that for any $1 \leq k \leq r-1$, $\rho(P_k^\wedge) \geq \frac{1}{k+1}$. Although $\rho(P_k^\wedge)$ with $\rho(P_{k+1}^\vee)$ share the same spectral gap (recall (17) and (18)), it is not clear how to directly relate $\rho(P_k^\wedge)$ with $\rho(P_{k+1}^\vee)$. In fact, even adapting the proof of Corollary 12 seems difficult. We will use a different decomposition approach.

Note that the mixing time bound can be shown directly from Lemma 11 without going through modified log-Sobolev inequalities (mLSI), as is done in the last section. However, concentration bounds like Corollary 2 apparently still need mLSI.

Once again, we start with the base case.

Lemma 13. *Let I be an independent set of $\mathcal{M}$ such that $|I| \leq r - 2$. Then $\rho(P_{I,1}^\wedge) \geq 1/2$.*

Proof: Recall that we may assume $\mathbb{E}_{\pi_{I,1}}f = 1$ and thus $\text{Ent}_{\pi_1}(f) = \mathbb{E}_{\pi_{I,1}}(f\log f)$. Also, recall (13), for any $u, v \in E \setminus I$,

$$
P_{I,1}^\wedge(u,v) = \begin{cases}
\frac{1}{2} & \text{if } u = v; \\
\frac{w(\{u,v\})}{2w(u)} & \text{if } \{u,v\} \in \mathcal{M}(2); \\
0 & \text{otherwise.}
\end{cases}
$$

Rewriting the above,

$$
P_{I,1}^\wedge = \frac{\mathbf{I}}{2} + \frac{1}{2}\text{diag}(\mathbf{w}_I^{-1})W_I,
$$

where $\mathbf{w}_I = \{w_I(v)\}_{v \in E \setminus I}$. Thus, by (4),

$$
\begin{aligned}
\mathcal{E}_{P_{I,1}^\wedge}(f, \log f) &= f^\mathsf{T}\text{diag}(\pi_{I,1})\left(\mathbf{I} - P_{I,1}^\wedge\right)\log f \\
&= \frac{1}{2}\Big(\mathbb{E}_{\pi_{I,1}}(f\log f) \\
&\quad\quad -f^\mathsf{T}\text{diag}(\pi_{I,1})\text{diag}(\mathbf{w}_I^{-1})W_I\log f\Big) \\
&= \frac{1}{2}\left(\text{Ent}_{\pi_{I,1}}(f) - \frac{1}{w(I)}\cdot f^\mathsf{T}W_I\log f\right).
\end{aligned}
$$

As $\log x \leq x - 1$,

$$
\begin{aligned}
f^\mathsf{T}W_I\log f &\leq f^\mathsf{T}W_If - f^\mathsf{T}W_I\mathbf{1} \\
&\leq w(I) - w(I) = 0,
\end{aligned}
$$

where in the last line we used Lemma 6 and $\mathbb{E}_{\pi_{I,1}} f = 1$. This finishes the proof. $\blacksquare$

Lemma 13 is a strengthening of the fact that the lazy random walk on 1-skeletons of links of a matroid $\mathcal{M}$ (namely $P_{I,1}^{\wedge}$) has spectral gap at least $1/2$, (cf. [1]).

Lemma 14. *For any $1 \le k \le r - 1$,*

$$\rho(P_k^{\wedge}) \ge \frac{1}{k+1}.$$

Proof: Recall (13) that

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

For $K \in \mathcal{M}(k-1)$, we extend $P_{K,1}^{\wedge}$ to a square matrix indexed by $\mathcal{M}(k)$ as follows,

$$P_{K,1}^{\wedge}(I, J) = \begin{cases} 0 & \text{if } K \not\subset I; \\ \frac{1}{2} & \text{if } K \subset I \text{ and } I = J; \\ \frac{w(I \cup J)}{2w(I)} & \text{if } K = I \cap J. \end{cases}$$

Let $S_K = \{K \cup \{v\} \mid v \in \mathcal{M}_K(1)\}$ be the support of $\pi_{K,1}$. Notice that for any $I \in \mathcal{M}(k)$,

$$|\{K \mid K \in \mathcal{M}(k-1), K \subset I\}| = k,$$

and if $I \cup J \in \mathcal{M}(k+1)$, then $I \cap J \in \mathcal{M}(k-1)$. We have

$$P_k^{\wedge} - \frac{1}{k+1} \cdot \mathbf{I} = \frac{2}{k+1} \sum_{K \in \mathcal{M}(k-1)} \left(P_{K,1}^{\wedge} - \frac{1}{2} \cdot \mathbf{I}_{S_K} \right),$$

where $\mathbf{I}_{S_K}$ is the diagonal matrix with 1 on S_k and 0 otherwise. Equivalently,

$$\mathbf{I} - P_k^{\wedge} = \frac{k}{k+1} \cdot \mathbf{I} + \frac{2}{k+1} \sum_{K \in \mathcal{M}(k-1)} \left(\frac{1}{2} \cdot \mathbf{I}_{S_K} - P_{K,1}^{\wedge} \right)$$

$$= \frac{2}{k+1} \sum_{K \in \mathcal{M}(k-1)} \left(\mathbf{I}_{S_K} - P_{K,1}^{\wedge} \right). \quad (33)$$

Furthermore, we have entropy decompositions similar to (26). For any $I \in \mathcal{M}(k)$, $K \in \mathcal{M}(k-1)$ and $K \subset I$,

$$\pi_k(I) = \frac{w(I)}{Z_k} = \frac{w(K)}{Z_k} \cdot \frac{w(I)}{w(K)} = k \pi_{k-1}(K) \pi_{K,1}(I),$$

as $Z_{k-1} = k Z_k$. This implies that

$$\pi_k(I) = \sum_{K \in \mathcal{M}(k-1), K \subset I} \pi_{k-1}(K) \pi_{K,1}(I)$$

$$= \sum_{K \in \mathcal{M}(k-1)} \pi_{k-1}(K) \pi_{K,1}(I).$$

Then, for any $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\ge 0}$ such that $\mathbb{E}_{\pi_k} f^{(k)} = 1$, we have

$$\text{Ent}_{\pi_k}\left(f^{(k)}\right) = \sum_{K \in \mathcal{M}(k-1)} \pi_{k-1}(K) \text{Ent}_{\pi_{K,1}}\left(f^{(k)}\right)$$

$$+ \sum_{K \in \mathcal{M}(k-1)} \pi_{k-1}(K) \, \mathbb{E}_{\pi_{K,1}} f^{(k)} \log \mathbb{E}_{\pi_{K,1}} f^{(k)}$$

$$= \sum_{K \in \mathcal{M}(k-1)} \pi_{k-1}(K) \text{Ent}_{\pi_{K,1}}\left(f^{(k)}\right)$$

$$+ \text{Ent}_{\pi_{k-1}}\left(f^{(k-1)}\right), \quad (34)$$

where $f^{(k-1)}$ is defined in (22). Then Lemma 11 implies that

$$\sum_{K \in \mathcal{M}(k-1)} \pi_{k-1}(K) \text{Ent}_{\pi_{K,1}}\left(f^{(k)}\right) \ge \frac{1}{k} \cdot \text{Ent}_{\pi_k}\left(f^{(k)}\right). \quad (35)$$

On the other hand, it is straightforward from (33) that

$$\mathcal{E}_{P_k^{\wedge}}\left(f^{(k)}, \log f^{(k)}\right)$$

$$= \sum_{K \in \mathcal{M}(k-1)} \frac{2}{k+1} \mathbb{E}_{\pi_k} f \left(\mathbf{I}_{S_K} - P_{K,1}^{\wedge}\right) \log f$$

$$= \frac{2k}{k+1} \sum_{K \in \mathcal{M}(k-1)} \pi_{k-1}(K) \mathcal{E}_{P_{K,1}^{\wedge}}(f, \log f)$$

$$\ge \frac{k}{k+1} \sum_{K \in \mathcal{M}(k-1)} \pi_{k-1}(K) \text{Ent}_{\pi_{K,1}}(f) \quad \text{(by Lemma 13)}$$

$$\ge \frac{1}{k+1} \cdot \text{Ent}_{\pi_k}\left(f^{(k)}\right). \quad \text{(by (35))}$$

This finishes the proof. $\blacksquare$

VI. CONCENTRATION

One application of the modified log-Sobolev inequalities is to show concentration inequalities, via the Herbst argument (see, e.g., [4], [16]). In the discrete setting, concentration inequalities have been obtained in ([14], Section 5) and can also be obtained by combining various results in [26], [15], [27]. The following lemma and its proof are a small modification of ([11], 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,*

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

where $a \ge 0$ and

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

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

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

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

$$z(e^z + 1) \geq 2(e^z - 1), \tag{36}$$

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 (36) and obtain

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

which in turn implies that

$$(F_t(x) - F_t(y))(f(x) - f(y)) \leq$$
$$\frac{t}{2}(F_t(x) + F_t(y))(f(x) - f(y))^2. \tag{37}$$

Notice that (37) also holds even if $f(x) < f(y)$ by swapping x and y. Thus, we have that

$$\mathcal{E}_P\left(F_t, \log F_t\right)$$
$$= \frac{t}{2} \sum_{x \in \Omega} \sum_{y \in \Omega} \pi(x) P(x, y)(F_t(x) - F_t(y))(f(x) - f(y))$$

$$\text{(by (5))}$$

$$\leq \frac{t^2}{4} \sum_{x \in \Omega} \sum_{y \in \Omega} \pi(x) P(x, y)(F_t(x) + F_t(y))(f(x) - f(y))^2$$

$$\text{(by (37))}$$

$$= \frac{t^2}{2} \sum_{x \in \Omega} \pi(x) F_t(x) \sum_{y \in \Omega} P(x, y)(f(x) - f(y))^2$$

$$\text{(by the reversibility of } P)$$

$$\leq \frac{t^2}{2} v(f) \, \mathbb{E}_\pi F_t.$$

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

$$\text{Ent}_\pi\left(F_t\right) \leq ct^2 \, \mathbb{E}_\pi F_t.$$

By noticing that

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

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,

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

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

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

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

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:

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

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 in [17] 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

The results obtained in [17] and [11] 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 generalisation 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 [11] and [17], respectively. Moreover, Corollary 2 is also incomparable to the concentration bound of [19], 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** [17]. However, such negative dependence even when $|S| = 2$ is known *not* to hold for the uniform distribution over the bases of some matroids. See [28] 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

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

It is straightforward to verify that if $0 \leq \theta < 3 - 2\sqrt{2} \approx 0.17157$ or $\theta > 3 + 2\sqrt{2} \approx 5.82843$, then **SLC** fails. However, **SCP** holds as long as $0 \leq \theta \leq 6$. 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** [2], which in turn is equivalent to the following condition as $g_\mu \in \mathbb{R}[x_1, \ldots, x_n]$ is multiaffine:

$$\frac{\partial}{x_i} g_\mu(\mathbf{x}) \cdot \frac{\partial}{x_j} g_\mu(\mathbf{x}) \geq g_\mu(\mathbf{x}) \cdot \frac{\partial^2}{\partial x_i \partial x_j} g_\mu(\mathbf{x}), \qquad (38)$$

for any $i, j \in [n]$ and $\mathbf{x} \in \mathbb{R}^n$. See [29]. If we plug in $\mathbf{x} = \mathbf{1}$, then (38) 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 (38) 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.

REFERENCES

[1] N. Anari, K. Liu, S. Oveis Gharan, and C. Vinzant, "Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid," in *STOC*, 2019, pp. 1–12.

[2] P. Brändén and J. Huh, "Lorentzian polynomials," *arXiv*, vol. abs/1902.03719, 2019.

[3] M. Jerrum, *Counting, sampling and integrating: algorithms and complexity*, ser. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003, Chapter 9 (unpublished) is available at http://www.maths.qmul.ac.uk/~mj/ETHbook/chapter9.pdf.

[4] S. G. Bobkov and P. Tetali, "Modified logarithmic Sobolev inequalities in discrete settings," *J. Theoret. Probab.*, vol. 19, no. 2, pp. 289–336, 2006.

[5] T. Kaufman and I. Oppenheim, "High order random walks: Beyond spectral gap," in *APPROX-RANDOM*, ser. LIPIcs, vol. 116. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018, pp. 47:1–47:17.

[6] T.-Y. Lee and H.-T. Yau, "Logarithmic Sobolev inequality for some models of random walks," *Ann. Probab.*, vol. 26, no. 4, pp. 1855–1873, 1998.

[7] M. Jerrum, J.-B. Son, P. Tetali, and E. Vigoda, "Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains," *Ann. Appl. Probab.*, vol. 14, no. 4, pp. 1741–1765, 2004.

[8] B. Morris, "Improved mixing time bounds for the Thorp shuffle and L-reversal chain," *Ann. Probab.*, vol. 37, no. 2, pp. 453–477, 2009.

[9] ——, "Improved mixing time bounds for the Thorp shuffle," *Combin. Probab. Comput.*, vol. 22, no. 1, pp. 118–132, 2013.

[10] J. Borcea, P. Brändén, and T. M. Liggett, "Negative dependence and the geometry of polynomials," *J. Amer. Math. Soc.*, vol. 22, no. 2, pp. 521–567, 2009.

[11] J. Hermon and J. Salez, "Modified log-Sobolev inequalities for strong-Rayleigh measures," *arXiv*, vol. abs/1902.02775, 2019.

[12] N. Anari, S. Oveis Gharan, and A. Rezaei, "Monte Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes," in *COLT*, vol. 49, 2016, pp. 103–115.

[13] T. Feder and M. Mihail, "Balanced matroids," in *STOC*, 1992, pp. 26–38.

[14] S. Goel, "Modified logarithmic Sobolev inequalities for some models of random walk," *Stochastic Process. Appl.*, vol. 114, no. 1, pp. 51–79, 2004.

[15] M. D. Sammer, *Aspects of mass transportation in discrete concentration inequalities*, 2005, thesis (Ph.D.)–Georgia Institute of Technology.

[16] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*. Oxford University Press, Oxford, 2013.

[17] R. Pemantle and Y. Peres, "Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures," *Combin. Probab. Comput.*, vol. 23, no. 1, pp. 140–160, 2014.

[18] D. Dubhashi and D. Ranjan, "Balls and bins: a study in negative dependence," *Random Structures Algorithms*, vol. 13, no. 2, pp. 99–124, 1998.

[19] K. Garbe and J. Vondrák, "Concentration of Lipschitz functions of negatively dependent variables," *arXiv*, vol. abs/1804.10084, 2018.

[20] D. A. Levin and Y. Peres, *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017, second edition.

[21] P. Diaconis and L. Saloff-Coste, "Logarithmic Sobolev inequalities for finite Markov chains," *Ann. Appl. Probab.*, vol. 6, no. 3, pp. 695–750, 1996.

[22] L. Gurvits, "A polynomial-time algorithm to approximate the mixed volume within a simply exponential factor," *Discrete Comput. Geom.*, vol. 41, no. 4, pp. 533–555, 2009.

[23] ——, "On multivariate Newton-like inequalities," in *Advances in combinatorial mathematics*. Springer, Berlin, 2009, pp. 61–78.

[24] N. Anari, S. Oveis Gharan, and C. Vinzant, "Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids," in *FOCS*, 2018, pp. 35–46, available at abs/1807.00929.

[25] J. G. Oxley, *Matroid theory*. Oxford University Press, 1992.

[26] S. G. Bobkov and F. Götze, "Exponential integrability and transportation cost related to logarithmic Sobolev inequalities," *J. Funct. Anal.*, vol. 163, no. 1, pp. 1–28, 1999.

[27] S. G. Bobkov, C. Houdré, and P. Tetali, "The subgaussian constant and concentration inequalities," *Israel J. Math.*, vol. 156, pp. 255–283, 2006.

[28] J. Huh, B. Schröter, and B. Wang, "Correlation bounds for fields and matroids," *arXiv*, vol. abs/1806.02675, 2018.

[29] P. Brändén, "Polynomials with the half-plane property and matroid theory," *Adv. Math.*, vol. 216, no. 1, pp. 302–320, 2007.