Modified log-Sobolev inequalities for strongly log-concave distributions

MODIFIED LOG-SOBOLEV INEQUALITIES FOR STRONGLY LOG-CONCAVE DISTRIBUTIONS

MARY CRYAN, HENG GUO, AND GIORGOS MOUSA

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 $1/r$. As a consequence, we obtain an asymptotically optimal mixing time bound for this chain. Applications include the bases-exchange random walk in a matroid.

1. Introduction

Let $\pi : \{0,1\}^n \to \mathbb{R}_{\geqslant 0}$ be a discrete distribution. Consider the generating polynomial of π:

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

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

An important example of strongly log-concave distributions is the uniform distribution over the bases of a matroid (Anari et al., 2018a; Brändén and Huh, 2019). This discovery leads to the breakthrough result that the exchange walk over the bases of a matroid is rapidly mixing Anari et al. (2018a), 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. 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 $P_{BX,\pi}$ 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,

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

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 strongly log-concave distribution π,

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

where $\pi_{min} = \min_x \pi(x)$. This will improve the previous bound $t_{mix}(P_{BX,\pi}, \varepsilon) \leqslant r\left(\log\frac{1}{\pi_{min}} + \log\frac{1}{\varepsilon}\right)$ due to Anari et al. (2018a). Since π_{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 (Jerrum, 2003).[1]

[1]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

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

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

where f, g are two functions over Ω, and $\mathbf{I}$ denotes the identity matrix. Moreover, let the entropy of $f : \Omega \to \mathbb{R}_{\geqslant 0}$ 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 normalise $\mathbb{E}_\pi f = 1$, 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

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

Our main theorem is the following, which is a special case of Theorem 6.

Theorem 1. *Let π be an r-homogeneous strongly log-concave distribution. Then*

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

Since $t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\rho(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right)$ (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. (2018a) 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 $P_{BX,\pi}$ 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 6.

The previous work of Anari et al. (2018a), building upon (Kaufman and Oppenheim, 2018), focuses on the spectral gap of $P_{BX,\pi}$. 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 10, shows that the relative entropy contracts 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, which appears to be novel.

Prior to our work, similar bounds have been obtained only for strongly Rayleigh distributions, which, introduced by Borcea et al. (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 strongly Rayleigh distributions, improving upon the spectral gap bound of Anari et al. (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. The approach we took is entirely different.

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 prove modified log-Sobolev constant lower bounds for the "down-up" walk. In Section 5 we finish by dealing with the "up-down" walk.

2. Preliminaries

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

block has two elements and each basis is formed by choosing exactly one element from every block. The Markov chain $P_{BX,\pi}$ in this case is just a lazy random walk on the Boolean hypercube.

2.1. **Markov chains.** Let Ω be a discrete state space and π be a distribution over Ω. Let $P : \Omega \times \Omega \to \mathbb{R}_{\geqslant 0}$ 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 $\|f\|_\pi := \langle f, f\rangle_\pi$.

The adjoint operator P^* of P is defined as $P^*(x,y) = \frac{\pi(y)P(y,x)}{\pi(x)}$. Indeed, P^* is the (unique) operator that 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,

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

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

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

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

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

but we will not need this expression in this paper. 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 following

$$t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{\lambda(P)}\left(\frac{1}{2}\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon}\right), \tag{5}$$

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 $\mathcal{E}_P\left(\sqrt{f},\sqrt{f}\right)$ with the following entropy-like quantity:

$$\mathrm{Ent}_\pi(f) := \mathbb{E}_\pi(f\log f) - \mathbb{E}_\pi f \log \mathbb{E}_\pi f \tag{6}$$

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

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

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

$$t_{\mathrm{mix}}(P,\varepsilon) \leqslant \frac{1}{4\alpha(P)}\left(\log\log\frac{1}{\pi_{\min}} + \log\frac{1}{2\varepsilon^2}\right). \tag{7}$$

The saving seems modest comparing to (5), 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):

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

Similar to (7), we have that

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

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

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

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

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 (7) and (8). Indeed, as observed by Hermon and Salez (2019), by taking the indicator function $\frac{1}{\pi(x)}\mathbb{1}_x$ for all $x\in\Omega$,

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

In our setting of r-homogeneous strongly log-concave distributions, we cannot hope for an 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 (6), 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 a simplification $\mathrm{Ent}_\pi(f) = \mathbb{E}_\pi(f\log f)$. Indeed, 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 $I = \{i_1, \ldots, i_k\}$ as shorthand for $\partial_{i_1} \ldots \partial_{i_k}$.

Definition 2. *A polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ with non-negative coefficients is* log-concave *at $\mathbf{x} \in \mathbb{R}_{\geqslant 0}$ 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 by Gurvits (2009a,b). There are also notions of *complete log-concavity* introduced by Anari et al. (2018b), and *Lorentzian* polynomials introduced by Brändén and Huh (2019). It turns out that all three notions are equivalent. See Brändén and Huh (2019, Theorem 5.3).

The following property of strongly log-concave polynomials is particularly useful (Anari et al., 2018b; Brändén and Huh, 2019).

Proposition 3. *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, $\nabla^2 \partial_I p(\mathbf{1})$ having at most one positive eigenvalue is equivalent to $\nabla^2 \log \partial_I p(\mathbf{1})$ being negative semi-definite, but we will only need the direction 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 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 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 $1 \leqslant k \leqslant r$. 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 a (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:[2]

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

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 $\pi = \pi_r$. Let $Z_k = \sum_{I \in \mathcal{M}(k)} w(I)$ be the normalisation constant of π_k. In fact, for any $0 \leqslant k \leqslant r$, $k! Z_k = Z_0 = w(\emptyset)$.

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

$$\partial_I g_\pi(\mathbf{1}) = \sum_{B \in \mathcal{B}, I \subset B} \pi(B) = \frac{1}{Z_r} \sum_{B \in \mathcal{B}, I \subset B} w(B). \tag{9}$$

[2] 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., 2018a; Kaufman and Oppenheim, 2018), so we do not adopt it.

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 $k \leqslant r - |I|$ 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 $J \in \mathcal{M}(k + |I|)$,

$$\pi_{I,k}(J) := \begin{cases} \frac{k! w(J)}{w(I)} & \text{if } I \subset J; \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

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

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

$$\begin{aligned} w_I(\{u, v\}) &= w(I \cup \{u, v\}) \\ &= (r - |I| - 2)! \sum_{B \in \mathcal{B},\ I \cup \{u,v\} \subseteq B} w(B) \\ &= (r - |I| - 2)! Z_r \cdot \partial_u \partial_v \partial_I g_\pi(\mathbf{1}). \end{aligned} \tag{by (9)}$$

In other words, W_I is $\nabla^2 \partial_I g_\pi$ multiplied by the scalar $(r-|I|-2)!Z_r$. Thus, Proposition 3 implies the following.

Proposition 4. *Let π be an r-homogeneous strongly log-concave distribution over $\mathcal{M} = (E, \mathcal{I})$. If $I \in \mathcal{I}$ and $|I| \leqslant r - 2$, then the matrix W_I has at most one positive eigenvalue.*

Proposition 4 implies the following bound for a quadratic form, which will be useful later.

Lemma 5. *Let $f : \mathcal{M}_I(1) \to \mathbb{R}$ be a function such that $\mathbb{E}_{\pi_{I,1}} f = 1$. Then*

$$f^T W_I f \leqslant w(I).$$

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 4, W_I has at most one positive eigenvalue, and thus so does A. We may decompose A as

$$A = \sum_{i=1}^{|E \setminus I|} \lambda_i g_i g_i^T, \tag{11}$$

where $\{g_i\}$ is an orthonormal basis and $\lambda_i \leqslant 0$ for all $i \geqslant 2$. 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 (11) directly implies that

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

where $h_i = g_i D^{1/2}$. 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,

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

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

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 $1 \leqslant k \leqslant r-1$ and $I, J \in \mathcal{M}(k)$, we have that

$$P_k^{\wedge}(I,J) = \begin{cases} \frac{1}{k+1} & \text{if } I = J; \\ \frac{w(I \cup J)}{(k+1)w(I)} & \text{if } I \cup J \in \mathcal{M}(k+1); \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

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 $2 \leqslant k \leqslant r$,

$$P_k^{\vee}(I,J) = \begin{cases} \sum_{I' \in \mathcal{M}(k-1), I' \subset I} \frac{w(I)}{k w(I')} & \text{if } I = J; \\ \frac{w(J)}{k w(I \cap J)} & \text{if } |I \cap J| = k-1; \\ 0 & \text{if } |I \cap J| < k-1. \end{cases} \tag{13}$$

Thus, the bases-exchange walk $P_{BX,\pi}$ 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 6. *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 $2 \leqslant k \leqslant r$, $\rho(P_k^{\vee}) \geqslant \frac{1}{k}$;*
- *for any $1 \leqslant k \leqslant r-1$, $\rho(P_k^{\wedge}) \geqslant \frac{1}{k+1}$.*

The first part of Theorem 6 is shown by Corollary 11, and the second part by Lemma 13. 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 (16) and (17) below).

By (8), we have the following corollary.

Corollary 7. *In the same setting as Theorem 6, we have that*

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

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

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

For example, for the uniform distribution over bases of matroids, Corollary 7 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. (2018a). The mixing time bound in Corollary 7 is asymptotically optimal, as it is achieved for the bases of some matroids (Jerrum, 2003, 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 $P_{BX,\pi}$ in this case is just a lazy random walk on the n-dimensional Boolean hypercube, which has mixing time $\Theta(n \log n)$, matching the upper bound in Corollary 7.

4. 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

$$P_k^{\uparrow} := \operatorname{diag}(\mathbf{w}_k)^{-1} A_k \operatorname{diag}(\mathbf{w}_{k+1}), \tag{14}$$

$$P_{k+1}^{\downarrow} := \frac{1}{k+1} A_k^{\mathsf{T}}. \tag{15}$$

Then

$$P_k^{\wedge} = P_k^{\uparrow} P_{k+1}^{\downarrow}, \tag{16}$$

$$P_{k+1}^{\vee} = P_{k+1}^{\downarrow} P_k^{\uparrow}. \tag{17}$$

Let $D_k = \operatorname{diag}(\pi_k)$. Using (14) and (15), we get that

$$D_{k+1} P_{k+1}^{\downarrow} = (P_k^{\uparrow})^{\mathsf{T}} D_k. \tag{18}$$

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

$$f^{(i)} := \prod_{j=i}^{k-1} P_j^{\uparrow} f^{(k)}. \tag{19}$$

Intuitively, $f^{(i)}$ is the function $f^{(k)}$ "pushed down" to level i. The key lemma, namely Lemma 10, is that this operation contracts the relative entropy by a factor of $1 - \frac{1}{i}$ from level i to level $i-1$.

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

Lemma 8. *Let $k \geqslant 2$ and $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\geqslant 0}$ 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 $1 \leqslant i < k$, $J \in \mathcal{M}(i)$, $f^{(i)}(J) = \mathbb{E}_{\pi_{J,k-i}} f^{(k)}$;

(2) for any $1 \leqslant i \leqslant k$, $\mathbb{E}_{\pi_i} f^{(i)} = 1$.

Proof. For (1), we do an induction on i from $k-1$ to 1. The base case of $k-1$ is straightforward to verify. For the induction step, suppose the claim holds for all integers larger than i ($i < k-1$). Then we have that

$$\begin{aligned}
f^{(i)}(J) = P_i^{\uparrow} f^{(i+1)}(J) &= \sum_{I\in\mathcal{M}(i+1):I\supset J} \frac{w(I)}{w(J)} \cdot f^{(i+1)}(I) \\
\text{(by IH)} \qquad &= \sum_{I\in\mathcal{M}(i+1):I\supset J} \frac{w(I)}{w(J)} \cdot \mathbb{E}_{\pi_{I,k-i-1}} f^{(k)} \\
&= \sum_{I\in\mathcal{M}(i+1):I\supset J} \frac{w(I)}{w(J)} \sum_{K\in\mathcal{M}(k):K\supset I} \frac{(k-i-1)!\,w(K)}{w(I)} \cdot f^{(k)}(K) \\
&= \sum_{K\in\mathcal{M}(k):K\supset J} |\{I \in \mathcal{M}(i+1) : J \subset I \subset K\}| \cdot \frac{(k-i-1)!\,w(K)}{w(J)} \cdot f^{(k)}(K) \\
&= \sum_{K\in\mathcal{M}(k):K\supset J} \frac{(k-i)!\,w(K)}{w(J)} \cdot f^{(k)}(K) \\
&= \mathbb{E}_{\pi_{J,k-i}} f^{(k)}.
\end{aligned}$$

For (2), we have that

$$\begin{aligned}
\mathbb{E}_{\pi_i} f^{(i)} &= \sum_{J\in\mathcal{M}(i)} \pi_i(J)\, \mathbb{E}_{\pi_{J,k-i}} f^{(k)} \\
\text{(by (1))} \qquad &= \sum_{J\in\mathcal{M}(i)} \frac{w(J)}{Z_i} \sum_{K\in\mathcal{M}(k):K\supset J} \frac{(k-i)!w(K)}{w(J)} \cdot f^{(k)}(K) \\
&= \sum_{K\in\mathcal{M}(k)} \sum_{J\in\mathcal{M}(i):J\subset K} \frac{(k-i)!w(K)}{Z_i} \cdot f^{(k)}(K) \\
\text{(as } k!Z_k = i!Z_i\text{)} \qquad &= \sum_{K\in\mathcal{M}(k)} \frac{k!w(K)}{i!Z_i} \cdot f^{(k)}(K) = \sum_{K\in\mathcal{M}(k)} \frac{w(K)}{Z_k} \cdot f^{(k)}(K) \\
&= \mathbb{E}_{\pi_k} f^{(k)} = 1. \qquad \square
\end{aligned}$$

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

Lemma 9. *Let $f^{(2)} : \mathcal{M}(2) \to \mathbb{R}_{\geqslant 0}$ be a non-negative function defined on $\mathcal{M}(2)$. Then*

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

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 8. 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 $a \geqslant 0$ and $b > 0$,

$$a \log \frac{a}{b} \geqslant a - b. \tag{20}$$

Noticing that $Z_1 = 2Z_2$, we have

$$\begin{aligned}
&\mathrm{Ent}_{\pi_2}\left(f^{(2)}\right) - 2\mathrm{Ent}_{\pi_1}\left(f^{(1)}\right) \\
&= \sum_{S\in\mathcal{M}(2)} \pi_2(S) f^{(2)}(S) \log f^{(2)}(S) - 2\sum_{v\in E} \pi_1(v) \left(\sum_{S\in\mathcal{M}(2):v\in S} \frac{w(S)}{w(v)} f^{(2)}(S)\right) \log f^{(1)}(v) \\
&= \sum_{S\in\mathcal{M}(2)} \left(\pi_2(S) f^{(2)}(S) \log f^{(2)}(S) - 2\sum_{v\in S} \pi_1(v) \frac{w(S)}{w(v)} f^{(2)}(S) \log f^{(1)}(v)\right) \\
&= \sum_{S\in\mathcal{M}(2)} \left(\frac{w(S)}{Z_2} f^{(2)}(S) \log f^{(2)}(S) - 2\sum_{v\in S} \frac{w(v)}{Z_1} \cdot \frac{w(S)}{w(v)} f^{(2)}(S) \log f^{(1)}(v)\right) \\
&= \sum_{S=\{u,v\}\in\mathcal{M}(2)} \frac{w(S)}{Z_2} f^{(2)}(S) \left(\log f^{(2)}(S) - \log f^{(1)}(v) - \log f^{(1)}(u)\right) \\
&\geqslant \sum_{S=\{u,v\}\in\mathcal{M}(2)} \frac{w(S)}{Z_2} \left(f^{(2)}(S) - f^{(1)}(v) f^{(1)}(u)\right) \\
&= \sum_{S\in\mathcal{M}(2)} \pi_2(S) f^{(2)}(S) - \sum_{S=\{u,v\}\in\mathcal{M}(2)} \frac{w(S)}{Z_2} \cdot f^{(1)}(v) f^{(1)}(u) \\
&= 1 - \frac{1}{2Z_2} \cdot \left(f^{(1)}\right)^{\mathsf{T}} W_\emptyset f^{(1)},
\end{aligned}$$

where the inequality is by (20) 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 5 with $I = \emptyset$ and $w(\emptyset) = Z_1 = 2Z_2$. $\square$

We generalise Lemma 9 as follows.

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

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

Proof. We do an induction on k. The base case of $k = 2$ follows from Lemma 9.

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

Recall (10), 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 $Z_1 = (k+1)!Z_{k+1}$. It implies that

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

Then we have

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

Thus, we have the decomposition

$$\begin{aligned}
&\mathrm{Ent}_{\pi_{k+1}}\left(f^{(k+1)}\right) \\
&= \sum_{v \in \mathcal{M}(1)} \pi_1(v)\mathrm{Ent}_{\pi_{v,k}}\left(f^{(k+1)}\right) + \sum_{v \in \mathcal{M}(1)} \pi_1(v)\left(\mathbb{E}_{\pi_{v,k}} f^{(k+1)}\right)\log\left(\mathbb{E}_{\pi_{v,k}} f^{(k+1)}\right) \\
&= \sum_{v \in \mathcal{M}(1)} \pi_1(v)\mathrm{Ent}_{\pi_{v,k}}\left(f^{(k+1)}\right) + \mathrm{Ent}_{\pi_1}\left(f^{(1)}\right), \qquad (21)
\end{aligned}$$

where we use (1) and (2) of Lemma 8. Similarly,

$$\mathrm{Ent}_{\pi_k}\left(f^{(k)}\right) = \sum_{v \in \mathcal{M}(1)} \pi_1(v)\mathrm{Ent}_{\pi_{v,k-1}}\left(f^{(k)}\right) + \mathrm{Ent}_{\pi_1}\left(f^{(1)}\right) \qquad (22)$$

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 2.) Thus, we can apply the induction hypothesis on $\mathcal{M}_v$ at level k and get

$$\mathrm{Ent}_{\pi_{v,k}}\left(f^{(k+1)}\right) \geqslant \frac{k}{k-1} \cdot \mathrm{Ent}_{\pi_{v,k-1}}\left(f^{(k)}\right). \qquad (23)$$

Strictly speaking, in (23) 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 (23).

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

$$\mathrm{Ent}_{\pi_k}\left(f^{(k)}\right) \geqslant k \cdot \mathrm{Ent}_{\pi_1}\left(f^{(1)}\right). \qquad (24)$$

Thus, (22) and (24) together imply that

$$\sum_{\nu\in\mathcal{M}(1)} \pi_1(\nu)\mathrm{Ent}_{\pi_{\nu,k-1}}\left(f^{(k)}\right) \geqslant (k-1)\mathrm{Ent}_{\pi_1}\left(f^{(1)}\right). \tag{25}$$

Putting everything together,

$$\begin{aligned}
\text{(by (21))}\quad \mathrm{Ent}_{\pi_{k+1}}\left(f^{(k+1)}\right) &= \sum_{\nu\in\mathcal{M}(1)} \pi_1(\nu)\mathrm{Ent}_{\pi_{\nu,k}}\left(f^{(k+1)}\right) + \mathrm{Ent}_{\pi_1}\left(f^{(1)}\right),\\
\text{(by (23))}\quad &\geqslant \frac{k}{k-1}\sum_{\nu\in\mathcal{M}(1)} \pi_1(\nu)\mathrm{Ent}_{\pi_{\nu,k-1}}\left(f^{(k)}\right) + \mathrm{Ent}_{\pi_1}\left(f^{(1)}\right),\\
&= \left(\frac{k+1}{k} + \frac{1}{k(k-1)}\right)\sum_{\nu\in\mathcal{M}(1)} \pi_1(\nu)\mathrm{Ent}_{\pi_{\nu,k-1}}\left(f^{(k)}\right) + \mathrm{Ent}_{\pi_1}\left(f^{(1)}\right)\\
\text{(by (25))}\quad &\geqslant \frac{k+1}{k}\sum_{\nu\in\mathcal{M}(1)} \pi_1(\nu)\mathrm{Ent}_{\pi_{\nu,k-1}}\left(f^{(k)}\right) + \frac{k+1}{k}\mathrm{Ent}_{\pi_1}\left(f^{(1)}\right)\\
\text{(by (22))}\quad &= \frac{k+1}{k}\mathrm{Ent}_{\pi_k}\left(f^{(k)}\right).
\end{aligned}$$

This concludes the inductive step and thus the proof. □

Lemma 10 implies that the 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 entropy, we have the following corollary.

Corollary 11. *For any* $2 \leqslant k \leqslant r$,

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

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

$$\begin{aligned}
\mathrm{Ent}_{\pi_{k-1}}\left(f^{(k-1)}\right) &= \left(f^{(k-1)}\right)^{T} D_{k-1}\log f^{(k-1)}\\
&= \left(f^{(k)}\right)^{T}\left(P_{k-1}^{\uparrow}\right)^{T} D_{k-1}\log f^{(k-1)}\\
\text{(by (18))}\quad &= \left(f^{(k)}\right)^{T} D_k P_k^{\downarrow}\log P_{k-1}^{\uparrow} f^{(k)}\\
\text{(by Jensen's inequality)}\quad &\geqslant \left(f^{(k)}\right)^{T} D_k P_k^{\downarrow} P_{k-1}^{\uparrow}\log f^{(k)}\\
&= \mathrm{Ent}_{\pi_k}\left(f^{(k)}\right) - \mathcal{E}_{P_k^{\vee}}\left(f^{(k)}, \log f^{(k)}\right).
\end{aligned}$$

Together with Lemma 10 we have that $\rho(P_k^{\vee}) \geqslant \frac{1}{k}$. □

5. The up-down walk

In this section we establish an analogous result of Corollary 11, namely for any $1 \leqslant k \leqslant r-1$, $\rho(P_k^{\wedge}) \geqslant \frac{1}{k+1}$. Although $\rho(P_k^{\wedge})$ with $\rho(P_{k+1}^{\vee})$ share the same spectral gap (recall (16) and (17)), we do not how to directly relate $\rho(P_k^{\wedge})$ with $\rho(P_{k+1}^{\vee})$. In fact, even adapting the proof of Corollary 7 seems difficult. We will use a different decompositional approach.

Once again, we start with the base case.

Lemma 12. *Let* I *be an independent set of* $\mathcal{M}$ *such that* $|I| \leqslant r-2$. *Then* $\rho(P_{I,1}^{\wedge}) \geqslant 1/2$.

Proof. Recall that we may assume $\mathbb{E}_{\pi_{I,1}} f = 1$ and thus $\mathrm{Ent}_{\pi_1}(f) = \mathbb{E}_{\pi_{I,1}}(f \log f)$. Also, recall (12), for any $u, v \in E \setminus I$,

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

Rewriting the above,

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

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

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

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

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

where in the last line we used Lemma 5 and $\mathbb{E}_{\pi_{I,1}} f = 1$. This finishes the proof. □

Lemma 12 is a strengthening of the fact that the lazy random walk on 1-skeletons of links of a matroid $\mathcal{M}$ (namely $P^{\wedge}_{I,1}$) has spectral gap at least $1/2$, (cf. Anari et al., 2018a).

Lemma 13. *For any* $1 \leqslant k \leqslant r - 1$,

$$\rho(P^{\wedge}_k) \geqslant \frac{1}{k+1}.$$

Proof. Recall (12) that

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

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

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

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

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

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

$$\begin{aligned}\mathbf{I} - P_k^{\wedge} &= \frac{k}{k+1}\cdot\mathbf{I} + \frac{2}{k+1}\sum_{K\in\mathcal{M}(k-1)}\left(\frac{1}{2}\cdot\mathbf{I}_{S_K} - P_{K,1}^{\wedge}\right)\\ &= \frac{2}{k+1}\sum_{K\in\mathcal{M}(k-1)}\left(\mathbf{I}_{S_K} - P_{K,1}^{\wedge}\right). \qquad (26)\end{aligned}$$

Furthermore, we have entropy decompositions similar to (22). 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 $Z_{k-1} = kZ_k$. 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 $f^{(k)}:\mathcal{M}(k)\to\mathbb{R}_{\geqslant 0}$ such that $\mathbb{E}_{\pi_k} f^{(k)} = 1$, we have

$$\begin{aligned}\mathrm{Ent}_{\pi_k}\left(f^{(k)}\right) &= \sum_{K\in\mathcal{M}(k-1)}\pi_{k-1}(K)\mathrm{Ent}_{\pi_{K,1}}\left(f^{(k)}\right) + \sum_{K\in\mathcal{M}(k-1)}\pi_{k-1}(K)\,\mathbb{E}_{\pi_{K,1}} f^{(k)}\log\mathbb{E}_{\pi_{K,1}} f^{(k)}.\\ &= \sum_{K\in\mathcal{M}(k-1)}\pi_{k-1}(K)\mathrm{Ent}_{\pi_{K,1}}\left(f^{(k)}\right) + \mathrm{Ent}_{\pi_{k-1}}\left(f^{(k-1)}\right), \qquad (27)\end{aligned}$$

where $f^{(k-1)}$ is defined in (19). Then Lemma 10 implies that

$$\sum_{K\in\mathcal{M}(k-1)}\pi_{k-1}(K)\mathrm{Ent}_{\pi_{K,1}}\left(f^{(k)}\right) \geqslant \frac{1}{k}\cdot\mathrm{Ent}_{\pi_k}\left(f^{(k)}\right). \qquad (28)$$

On the other hand, it is straightforward from (26) that

$$\begin{aligned}\mathcal{E}_{P_k^{\wedge}}\left(f^{(k)},\log f^{(k)}\right) &= \sum_{K\in\mathcal{M}(k-1)}\frac{2}{k+1}\,\mathbb{E}_{\pi_k} f\left(\mathbf{I}_{S_K} - P_{K,1}^{\wedge}\right)\log f\\ &= \frac{2k}{k+1}\sum_{K\in\mathcal{M}(k-1)}\pi_{k-1}(K)\mathcal{E}_{P_{K,1}^{\wedge}}(f,\log f)\\ &\geqslant \frac{k}{k+1}\sum_{K\in\mathcal{M}(k-1)}\pi_{k-1}(K)\mathrm{Ent}_{\pi_{K,1}}(f) \qquad \text{(by Lemma 12)}\\ &\geqslant \frac{1}{k+1}\cdot\mathrm{Ent}_{\pi_k}\left(f^{(k)}\right). \qquad \text{(by (28))}\end{aligned}$$

This finishes the proof. □

Acknowledgement

We thank Mark Jerrum for his helpful comments.

References

Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. Monte Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes. In *COLT*, volume 49, pages 103–115, 2016. 2

Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. *arXiv*, abs/1811.01816, 2018a. 1, 2, 5, 7, 12

Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids. In *FOCS*, pages 35–46, 2018b. Available at abs/1807.00929. 5

Sergey G. Bobkov and Prasad Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19(2):289–336, 2006. 2, 4

Julius Borcea, Petter Brändén, and Thomas M. Liggett. Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.*, 22(2):521–567, 2009. 2

Petter Brändén and June Huh. Lorentzian polynomials. *arXiv*, abs/1902.03719, 2019. 1, 5

Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996. 4

Tomás Feder and Milena Mihail. Balanced matroids. In *STOC*, pages 26–38, 1992. 2

Leonid Gurvits. A polynomial-time algorithm to approximate the mixed volume within a simply exponential factor. *Discrete Comput. Geom.*, 41(4):533–555, 2009a. 5

Leonid Gurvits. On multivariate Newton-like inequalities. In *Advances in combinatorial mathematics*, pages 61–78. Springer, Berlin, 2009b. 5

Jonathan Hermon and Justin Salez. Modified log-Sobolev inequalities for strong-Rayleigh measures. *arXiv*, abs/1902.02775, 2019. 2, 4

Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. 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`. 1, 7

Mark Jerrum, Jung-Bae Son, Prasad Tetali, and Eric Vigoda. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.*, 14(4):1741–1765, 2004. 2

Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. In *APPROX-RANDOM*, volume 116 of *LIPIcs*, pages 47:1–47:17. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018. 2, 5

Tzong-Yow Lee and Horng-Tzer Yau. Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.*, 26(4):1855–1873, 1998. 2

David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition. 4

James G. Oxley. *Matroid theory*. Oxford University Press, 1992. 5

School of Informatics, University of Edinburgh, Informatics Forum, Edinburgh, EH8 9AB, United Kingdom.
E-mail address: `mcryan@inf.ed.ac.uk, hguo@inf.ed.ac.uk, s1687200@sms.ed.ac.uk`