Chapter 24

Entropic Independence

We end with a very recent paper by Anari, Jain, Koehler, Pham, Vuong [AJK+21a], who introduced the notion of entropic independence and used it to lower bound the modified log-Sobolev constants of fractionally log-concave distributions. The presentation is taken directly from [AJK+21a].

24.1 Larger-Step Down-Up Walks

We will consider the following natural generalization of Definition 20.5 for down-up walks.

Definition 24.1 (Larger-Step Down-Up Walks). *For a probability distribution $\mu : \binom{[n]}{k} \to \mathbb{R}$ and an integer $l \leq k$, the $k \leftrightarrow l$ down-up random walk $P^{\triangledown}_{k\leftrightarrow l}$ is the sequence of random sets $S_0, S_1, \ldots$ generated by the following algorithm:*

for** $t = 0, 1, \ldots$ **do
Select T_t uniformly at random from subsets of size l of S_t.
Select S_{t+1} with probability $\propto \mu(S_{t+1})$ from supersets of size k of T_t.
end for

We also take this opportunity to redefine the down and up operators, but use different notations, that are more natural for the analysis of random walks. Compare with Definition 20.2 and see Remark 20.3. We will think of these as matrices that act on probability distributions on the left.

Definition 24.2 (Down Operator). *For a ground set $[n]$ and $n \geq k \geq l$, define the row-stochastic down operator $D_{k\to l} \in \mathbb{R}^{\binom{[n]}{k}\times\binom{[n]}{l}}$ as*

$$D_{k\to l}(S,T) = \begin{cases} \frac{1}{\binom{k}{l}} & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 24.3 (Up Operator). *For a ground set $[n]$ and $n \geq k \geq l$, and probability distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, define the up operator $U_{l\to k} \in \mathbb{R}^{\binom{[n]}{l}\times\binom{[n]}{k}}$ as*

$$U_{l\to k}(T,S) = \begin{cases} \frac{\mu(S)}{\sum_{S'\supseteq T}\mu(S')} & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

The following properties should be familiar.

Exercise 24.4 (Stationary Distributions of $P^{\bigtriangledown}_{k\leftrightarrow l}$ and $P^{\bigtriangleup}_{k\leftrightarrow l}$)**.** *Define $\mu_k := \mu$ and more generally define $\mu_l := \mu_k D_{k\to l}$. The operators $P^{\bigtriangledown}_{k\leftrightarrow l} := D_{k\to l}U_{l\to k}$ and $P^{\bigtriangleup}_{l\leftrightarrow k} := U_{l\to k}D_{k\to l}$ both define Markov chains that are time-reversible and have nonnegative eigenvalues. Moreover μ_k and μ_l are respectively their stationary distributions.*

24.2 Entropic Independence

The notion of entropic independence is about the entropy contraction of the down operator.

Definition 24.5 (Entropic Independence)**.** *A probability distribution μ on $\binom{[n]}{k}$ is said to be $(1/\alpha)$-entropically independent, for $\alpha \in (0,1]$, if for all probability distributions ν on $\binom{[n]}{k}$,*

$$\mathcal{D}_{\mathrm{KL}}\big(\nu D_{k\to 1} \parallel \mu D_{k\to 1}\big) \le \frac{1}{\alpha k} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

We remark that spectral independence can be used to prove the variance contraction of the down operator (see [Ove22]), which can be used to prove a lower bound on the spectral gap as in Chapter 23. So Definition 24.5 can be seen as a natural analog of spectral independence for proving a lower bound on the modified log-Sobolev constant as in Chapter 23.

We study two main results from [AJK+21a]. The first one characterizes entropic independence and fractionally log-concavity. In particular, it proves that an α-fractionally log-concave distribution is $(1/\alpha)$-entropically independent for all conditional distributions.

Theorem 24.6 (Entropic Independence and Fractionally Log-Concavity [AJK+21a])**.** *Let μ be a probability distribution on $\binom{[n]}{k}$ and let $\mu_1 = \mu D_{k\to 1} \in \mathbb{R}^n$. Then, for any $\alpha \in (0,1]$,*

$$\mu \text{ is } \frac{1}{\alpha}\text{-entropically independent} \iff \forall (z_1,\ldots,z_n) \in \mathbb{R}^n_{\ge 0}, \quad g_\mu(z_1^\alpha,\ldots,z_n^\alpha)^{\frac{1}{k\alpha}} \le \sum_{i=1}^n \mu_1(i)\cdot z_i.$$

Consequently, if μ is α-fractionally log-concave, then μ is $(1/\alpha)$-entropically independent. Moreover,

$$\mu \text{ is } \alpha\text{-fractionally log-concave} \iff \lambda{*}\mu \text{ is } (1/\alpha)\text{-entropically independent } \forall \lambda = (\lambda_1,\ldots,\lambda_n) \in \mathbb{R}^n_{>0},$$

where $\lambda{*}\mu$ is the distribution scaled by external field λ such that $g_{\lambda*\mu}(z_1,\ldots,z_n) \propto g_\mu(\lambda_1 z_1,\ldots,\lambda_n z_n)$.

The second one proves that if μ is $(1/\alpha)$-entropically independent for all conditional distributions, then the $k \leftrightarrow k - \frac{1}{\alpha}$ down-up walk is fast mixing.

Theorem 24.7 (Local to Global Entropy Contraction [AJK+21a])**.** *Suppose $\mu : \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ is α-fractionally log-concave, or more generally $(1/\alpha)$-entropically independent for all conditional distributions. Let $l \le k - \lceil 1/\alpha \rceil$. Then, for all probability distributions ν on $\binom{[n]}{k}$,*

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l}) \le (1-\kappa)\cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

Consequently, the $k \leftrightarrow l$ down-up walk with respect to μ has modified log-Sobolev constant $\Omega(\kappa)$ where Ω hides an absolute constant and when $1/\alpha$ is an integer then

$$\kappa = \binom{k-l}{1/\alpha} \Big/ \binom{k}{1/\alpha}.$$

Combining Theorem 24.6 and Theorem 24.7, the $k \leftrightarrow k - \frac{1}{\alpha}$ down-up is fast mixing for α-fractionally log-concave distributions. Note that this is a generalization of the result of Cryan, Guo, and Mousa in Theorem 23.19 for strongly log-concave distributions, which are 1-fractionally log-concave. This also gives the optimal mixing time analysis for the monomer-dimer systems and the non-symmetric determinantal point process that we discussed in Chapter 22 (see [AJK+21a]).

24.3 Fractional Log-Concavity Implies Local Entropic Contraction

In this section, we prove one direction of Theorem 24.6, that fractionally log-concavity implies entropic independence. We refer the reader to [AJK+21a] for the other direction, which is not needed for the conclusion of fast mixing.

The following statement will be used to replace log-concavity by $\frac{1}{d}$-th root concavity.

Problem 24.8. *Let $\mathcal{C} \subseteq \mathbb{R}^n_{\geq 0}$ denote a convex cone. For a d-homogeneous function $f : C \to \mathbb{R}_{\geq 0}$, f is log-concave if and only if $f^{1/d}$ is concave.*

The first step is to show that fractional log-concavity implies that the transformed generating polynomial is upper bounded by its linear tangent.

Lemma 24.9 (Linear Tangent Upper Bound)**.** *If μ is a α-fractionally log-concave distribution on $\binom{[n]}{k}$, then*

$$g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \leq \Big(\sum_{i=1}^{n} \mu_1(i) \cdot z_i \Big)^{\alpha k}.$$

Proof. Let $f(z_1, \ldots, z_n) := g_\mu(z_1^\alpha, \ldots, z_n^\alpha)^{\frac{1}{\alpha k}}$ be the transformed generating polynomial. As the polynomial $g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is αk-homogeneous and log-concave, it follows from Problem 24.8 that f is concave. Therefore, by concavity, for all $z_1, \ldots, z_n > 0$,

$$f(z_1, \ldots, z_n) \ \leq \ f(\vec{1}) + \langle \nabla f(\vec{1}), \vec{z} - \vec{1} \rangle \ = \ f(\vec{1}) + \sum_{i=1}^{n} \partial_i f(\vec{1}) \cdot (z_i - 1) \ = \ \sum_{i=1}^{n} \partial_i f(\vec{1}) \cdot z_i,$$

where the last equality is because f is 1-homogeneous and so $\sum_{i=1}^{n} \partial_i f(\vec{1}) = f(\vec{1})$. By the chain rule,

$$\partial_i f(\vec{1}) \ = \ \big(\alpha \cdot \partial_i g_\mu(\vec{1})\big) \Big(\frac{1}{\alpha k} \cdot g_\mu(\vec{1})^{\frac{1}{\alpha k} - 1} \Big) \ = \ \frac{1}{k} \Pr_{S \sim \mu}[i \in S] \ = \ \mu_1(i),$$

where the last equality is by the definition that $\mu_1 = \mu D_{k \to 1}$. Therefore,

$$f(z_1, \ldots, z_n) \leq \sum_{i=1}^{n} \mu_1(i) \cdot z_i \quad \Longrightarrow \quad g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \leq \Big(\sum_{i=1}^{n} \mu_1(i) \cdot z_i \Big)^{\alpha k}.$$

□

The second step is to show that the linear tangent upper bound implies entropic independence. A key idea in the proof is to fix the marginal probability and to use the following Gurvits' capacity-type bound proved by Singh and Vishnoi, which is obtained by convex duality.

Lemma 24.10 (Capacity Bound of Relative Entropy)**.** *Consider a homogeneous distribution* $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ *and let* $g_\mu(z_1, \ldots, z_n)$ *be its multivariate generating polynomial. Then, for any* $q \in \mathbb{R}^n_{\geq 0}$ *with* $\sum_{i=1}^n q_i = 1$,

$$\inf \left\{ \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \mid \nu D_{k\to 1} = q \right\} = -\log \left(\inf_{z_1,\ldots,z_n>0} \frac{g_\mu(z_1,\ldots,z_n)}{z_1^{kq_1} \cdots z_n^{kq_n}} \right).$$

Lemma 24.11 (Linear Tangent Upper Bound Implies Entropic Independence)**.** *If* μ *is a homogeneous distribution whose generating polynomial* g_μ *satisfies*

$$g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \leq \left(\sum_{i=1}^n \mu_1(i) \cdot z_i \right)^{\alpha k}.$$

for all $z_1, \ldots, z_n \in \mathbb{R}^n_{\geq 0}$*, then* μ *is* $(1/\alpha)$*-entropically independent.*

Proof. Let ν be an arbitrary probability distribution on $\binom{[n]}{k}$ and let $\nu_1 := \nu D_{k\to 1}$, so that $\nu_1 \in \mathbb{R}^n_{\geq 0}$ and $\sum_{i=1}^n \nu_1(i) = 1$. By Lemma 24.10,

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \geq \inf \left\{ \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \mid \nu D_{k\to 1} = \nu_1 \right\} = -\log \left(\inf_{z_1,\ldots,z_n>0} \frac{g_\mu(z_1,\ldots,z_n)}{z_1^{k\cdot\nu_1(1)} \cdots z_n^{k\cdot\nu_1(n)}} \right).$$

By the linear tangent upper bound,

$$\inf_{z_1,\ldots,z_n>0} \frac{g_\mu(z_1,\ldots,z_n)}{z_1^{k\cdot\nu_1(1)} \cdots z_n^{k\cdot\nu_1(n)}} \leq \inf_{z_1,\ldots,z_n>0} \frac{\left(\sum_{i=1}^n \mu_1(i) \cdot z_i^{\frac{1}{\alpha}} \right)^{\alpha k}}{z_1^{k\cdot\nu_1(1)} \cdots z_n^{k\cdot\nu_1(n)}} \leq \prod_{i=1}^n \left(\frac{\mu_1(i)}{\nu_1(i)} \right)^{\alpha k\cdot\nu_1(i)}$$

where the last inequality is by plugging in $z_i = (\nu_1(i)/\mu_1(i))^\alpha$. Taking log and negating gives

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \geq -\log \prod_{i=1}^n \left(\frac{\mu_1(i)}{\nu_1(i)} \right)^{\alpha k\cdot\nu_1(i)} = \alpha k \sum_{i=1}^n \nu_1(i) \log \left(\frac{\nu_1(i)}{\mu_1(i)} \right) = \alpha k \cdot \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1} \parallel \mu D_{k\to 1}),$$

which establishes $(1/\alpha)$-entropic independence by Definition 24.5. □

Finally, we note that α-fractional log-concavity is preserved by scaling, which implies that the generating polynomial of any conditional distribution

$$g_{\mu_S} \propto \lim_{\lambda\to\infty} \frac{g_\mu\left(\lambda z_1, \ldots, \lambda z_{|S|}, \ldots, z_n\right)}{\lambda^{|S|}}.$$

is also α-fractionally log-concave, and thus μ_S is also $(1/\alpha)$-entropically independent by Lemma 24.9 and Lemma 24.11.

Therefore, we have proved the direction of Theorem 24.6 that we need, that an α-fractionally log-concave distribution is $(1/\alpha)$-entropic independent for all conditional distributions.

24.4 Local Entropy Contraction to Global Entropy Contraction

In this section, we prove Theorem 24.7, which also proves Theorem 23.19 for strongly log-concave distributions.

Proof of Theorem 24.7. The plan is to write both $\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$ and $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l})$ as a telescoping sum of terms of the form $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to i} \parallel \mu D_{k\to i}) - \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to(i-1)} \parallel \mu D_{k\to(i-1)})$.

Consider the following random process. Sample a set $S \sim \mu$ and uniformly at random permute its elements to obtain $X_1, \ldots, X_k$. Notice that any prefix $X_1, \ldots, X_i$ is distributed according to $\mu D_{k\to i}$. Consider the random variable

$$\tau_i := \frac{\nu D_{k\to i}(\{X_1, \ldots, X_i\})}{\mu D_{k\to i}(\{X_1, \ldots, X_i\})} \log \frac{\nu D_{k\to i}(\{X_1, \ldots, X_i\})}{\mu D_{k\to i}(\{X_1, \ldots, X_i\})}$$

Then

$$\mathbb{E}_{S\sim\mu}[\tau_i] = \mathbb{E}_{\{X_1,\ldots,X_i\}\sim\mu D_{k\to i}}[\tau_i] = \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to i} \parallel \mu D_{k\to i}).$$

Therefore, we can write both $\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$ and $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l})$ as telescoping sums

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) = \mathbb{E}[\tau_k] = \sum_{i=0}^{k-1} \big(\mathbb{E}[\tau_{i+1}] - \mathbb{E}[\tau_i]\big) \text{ and } \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l}) = \mathbb{E}[\tau_l] = \sum_{i=0}^{l-1} \big(\mathbb{E}[\tau_{i+1}] - \mathbb{E}[\tau_i]\big).$$

To prove entropy contraction, it is equivalent to proving that the last $k-l$ terms in the telescoping sum are sufficiently large compared to the rest.

Let $\Delta_i := \mathbb{E}[\tau_{i+1}] - \mathbb{E}[\tau_i]$ and $\beta_i := \frac{1}{\alpha(k-i)-1}$. As μ is α-fractionally log-concave, by Theorem 24.6, μ is $(1/\alpha)$-entropically independent, and thus it follows from Definition 24.5 that

$$\Delta_0 = \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1} \parallel \mu D_{k\to 1}) \le \frac{1}{\alpha k} \cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \quad \Longrightarrow \quad \Delta_0 \le \beta_0(\Delta_1 + \cdots + \Delta_{k-1}).$$

As conditioning preserves α-fractionally log-concavity, we can apply the same argument to each conditional distribution $\mu(\cdot \mid X_1, \ldots, X_i)$ and then take the expectation over $X_1, \ldots, X_i$ to get

$$\Delta_i \le \beta_i(\Delta_{i+1} + \ldots + \Delta_{k-1})$$

for each $1 \le i \le l-1$. Combining these inequalities, it follows inductively that for all $0 \le i \le l \le k - \frac{1}{\alpha}$,

$$\Delta_i \le \beta_i \cdot (\Delta_l + \ldots + \Delta_{k-1}) \cdot \prod_{j=i+1}^{l-1} (\beta_j + 1).$$

Hence,

$$\frac{\Delta_0 + \cdots + \Delta_{k-1}}{\Delta_l + \cdots + \Delta_{k-1}} = 1 + \frac{\Delta_0 + \cdots + \Delta_{l-1}}{\Delta_l + \cdots + \Delta_{k-1}} \le 1 + \sum_{i=0}^{l-1} \beta_i \cdot \prod_{j=i+1}^{l-1} (\beta_j + 1) = \prod_{i=0}^{l-1}(1 + \beta_i) = \prod_{j=k-l+1}^{k} \frac{j}{j - \frac{1}{\alpha}}.$$

Let $\Gamma(\cdot)$ be the Gamma function. If $1/\alpha$ is an integer, then the RHS is

$$\frac{\Gamma(k+1)/\Gamma(k+1-l)}{\Gamma(k+1-1/\alpha/\Gamma(k+1-l-1/\alpha)} = \frac{k!/(k-l)!}{(k-1/\alpha)!(k-l-1/\alpha)!} = \binom{k}{1/\alpha}\Big/\binom{k-l}{1/\alpha}.$$

This implies that

$$\frac{\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)}{\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) - \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l})} = \frac{\Delta_0 + \cdots + \Delta_{k-1}}{\Delta_l + \cdots + \Delta_{k-1}} \le \binom{k}{1/\alpha}\Big/\binom{k-l}{1/\alpha},$$

and rearranging gives the entropy contraction statement $\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l}) \le (1-\kappa) \cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$.

Finally, by the data processing inequality in information theory, apply the up operator would not increase the relative entropy, and so

$$\mathcal{D}_{\mathrm{KL}}(\nu P^{\nabla}_{k\leftrightarrow l} \parallel \mu P^{\nabla}_{k\leftrightarrow l}) = \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} U_{l\to k} \parallel \mu D_{k\to l} U_{l\to k}) \leq \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to l} \parallel \mu D_{k\to l}) \leq (1-\kappa)\cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

This implies that the relative entropy is exponentially decreasing and thus the we can upper bound the mixing time as in Theorem 23.14 with $\rho(P) \approx \kappa$. This proves the consequence of the modified log-Sobolev constant in bounding mixing time. It is left as in Problem 24.12 to prove that entropy contraction indeed implies a lower bound on the modified log-Sobolev constant. □

Problem 24.12 (Entropy Contraction Implies Modified Log-Sobolev Constant)**.** *Let μ be a probability distribution on $[n]$. Let P denote the transition matrix of an irreducible, reversible Markov chain on $[n]$ with stationary distribution μ. Suppose there exists some $\alpha \in (0,1]$ such that for all probability measures ν on $[n]$ which are absolutely continuous with respect to μ, we have*

$$\mathcal{D}_{\mathrm{KL}}(\nu P \parallel \mu P) \leq (1-\alpha)\cdot \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

Then the modified log-Sobolev constant of P is

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

24.5 Summary

Since the resolution of the matroid expansion conjecture using the connection to high-dimensional expanders, there are many recent developments in analyzing mixing time of Markov chains. It is quite amazing to see that the techniques can be extended to bounding the (notorious) modified log-Sobolev constants, with applications in proving optimal mixing times, in bounding correlations, and in proving concentration inequalities. In these most recent developements, the concepts from high-dimensional expanders have been bypassed and one could understand the results directly using probabilistic and analytical concepts (see the newer papers [AJK+21b, CE22]). This area is progressing and evolving very quickly, and the notes in the next offering may be entirely different.

24.6 References

[AJK+21a] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence in high-dimensional expanders: Modified log-sobolev inequalities for fractionally log-concave polynomials and the ising model. *CoRR*, abs/2106.04105, 2021. 217, 225, 226, 227

[AJK+21b] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence ii: Optimal sampling and concentration via restricted modified log-sobolev inequalities. *arXiv preprint arXiv:2111.03247*, 2021. 230

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. *arXiv preprint arXiv:2203.04163*, 2022. 230

[Ove22] Shayan Oveis Gharan. Course notes of "Modern spectral graph theory", lecture 12. 2022. https://courses.cs.washington.edu/courses/cse599s/22wi/. 153, 159, 226