A Local-to-Global Framework: Simplicial Complex

Zhidan Li

Contents

1 Markov Chains and Local Properties

Given a distribution μ over state space Ω, let P be a reversible Markov chain with respect to Ω. We define the mixing time of P at initial state $x \in \Omega$ as

$$t_{\text{mix}}(P, x, \varepsilon) = \inf\left\{t \geq 0 \,\middle|\, \mathcal{D}_{\text{TV}}\left(P^t(x, \cdot) \,\middle\|\, \mu\right) \leq \varepsilon\right\}.$$

The functional inequality is introduced to bound the mixing time of P. For two functions $f, g : \Omega \to \mathbb{R}$, define the Dirichlet form with assumption that all terms are well-defined to be

$$\mathcal{E}_P(f, g) := \langle f, (I - P)g \rangle_\mu = \int_{x \in \Omega} f(x)(I - P)g(x) \, \mathrm{d}\mu(x). \tag{1}$$

Definition 1.1 (Functional Inequalities). Given a reversible Markov chain P with respect to its stationary distribution μ over Ω, we define the *spectral gap* of P as

$$\text{Gap}(P) := \inf_{f:\Omega\to\mathbb{R}} \frac{\mathcal{E}_P(f, f)}{\mathbf{Var}_\mu(f)}$$

and we define the *modified log-Sobolev inequality constant* (MLSI) of P as

$$\rho_{\text{LS}}(P) := \inf_{f:\Omega\to\mathbb{R}_{\geq 0}} \frac{\mathcal{E}_P(f, \log f)}{\mathbf{Ent}_\mu[f]}$$

where the variance and the entropy of f with respect to μ are defined as

$$\mathbf{Var}_\mu(f) = \mathbf{E}_\mu\left[f^2\right] - \mathbf{E}_\mu[f]^2, \quad \mathbf{Ent}_\mu[f] = \mathbf{E}_\mu[f \log f] - \mathbf{E}_\mu[f] \log \mathbf{E}_\mu[f].$$

Moreover, for every reversible P, it holds that $\text{Gap}(P) = 1 - \lambda_2(P)$.

Previously several works have used the functional inequalities to bound the mixing time.

Lemma 1.2 (Theorem 12.4 in [LP17]). *There exists a universal constant $C > 0$ such that the followings hold for all $x \in \Omega$,*

$$t_{\mathrm{mix}}(P, x, \varepsilon) \le \frac{C}{\mathrm{Gap}(P)} \left(\log \frac{1}{\mu(x)} + \log \frac{1}{\varepsilon} \right),$$
$$t_{\mathrm{mix}}(P, x, \varepsilon) \le \frac{C}{\rho_{\mathrm{LS}}(P)} \left(\log \log \frac{1}{\mu(x)} + \log \frac{1}{\mu(x)} \right).$$

Dirichlet form and continuous-time random walk

The Dirichlet form is introduced in Bobkov and Tetali [BT06] to analyze the mixing time of Markov chains. To briefly see this, consider the following Markov process:

$$P_t = e^{-(I-P)t}, \quad \forall t \ge 0.$$

Let μ_0 be the initial distribution and $\mu_t = \mu_0 P_t$. Consider the function f_t supported on Ω as $f_t = \frac{\mathrm{d}\mu_t}{\mathrm{d}\mu}$. Thus $\langle f_t, \mathbf{1} \rangle_\mu = 1$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t = -(I-P) f_t, \quad \frac{\mathrm{d}}{\mathrm{d}t} \log f_t = -(I-P)\mathbf{1}.$$

Then,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{Var}_\mu(f_t) &= \frac{\mathrm{d}}{\mathrm{d}t} \left(\langle f_t, f_t \rangle_\mu - \langle f_t, \mathbf{1} \rangle_\mu^2 \right) \\
&= \left\langle \frac{\mathrm{d}}{\mathrm{d}t} f_t, f_t \right\rangle_\mu + \left\langle f_t, \frac{\mathrm{d}}{\mathrm{d}t} f_t \right\rangle_\mu - 2 \langle f_t, \mathbf{1} \rangle_\mu \left\langle \frac{\mathrm{d}}{\mathrm{d}t} f_1, \mathbf{1} \right\rangle_\mu \\
&= 2 \langle f_t, -(I-P) f_t \rangle_\mu - 2 \langle f_t, \mathbf{1} \rangle_\mu \langle -(I-P) f_t, \mathbf{1} \rangle_\mu \\
&= -2 \mathcal{E}_P(f_t, f_t).
\end{aligned}$$

Similarly for the relative entropy we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{D}_{\mathrm{KL}}(\mu_t \parallel \mu) &= \frac{\mathrm{d}}{\mathrm{d}t} \mathbf{Ent}_\mu[f_t] \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \langle f_t, \log f_t \rangle_\mu \\
&= \left\langle \frac{\mathrm{d}}{\mathrm{d}t} f_t, \log f_t \right\rangle_\mu + \left\langle f_t, \frac{\mathrm{d}}{\mathrm{d}t} \log f_t \right\rangle_\mu \\
&= \langle -(I-P) f_t, \log f_t \rangle_\mu + \langle f_t, -(I-P)\mathbf{1} \rangle_\mu \\
&= -2 \mathcal{E}_P(f_t, \log f_t).
\end{aligned}$$

The two inequalities drive us to bound the spectral gap and MLSI constant.

1.1 Spectral independence

Now we consider the case $\Omega \subseteq [q]^n$ for a positive integer $q \ge 2$. It makes common sense since we focus on the mixing rate of the Glauber dynamics for the Gibbs distribution of q-spin systems.

The local property named *spectral independence* is firstly introduced in Anari, Liu and Oveis Gharan [ALOG20] to evaluate the local dependence in hard-core models.

Definition 1.3 (Spectral Independence - Boolean Domain). Let μ be a distribution over $\Omega \subseteq \{-1, +1\}^n$. We define the *influence matrix* $\Psi_\mu \in \mathbb{R}^{n \times n}$ to be

$$\Psi_\mu(i, j) = \frac{1}{2} \mathbf{E}_{X \sim \mu} \left[X_i \mid X_j = 1 \right] - \frac{1}{2} \mathbf{E}_{X \sim \mu} \left[X_i \mid X_j = -1 \right], \quad \forall i, j \in [n].$$

For $\eta > 0$, we say μ is η-spectrally independent if $\left\| \Psi_\mu \right\|_{\mathrm{OP}} \le 1 + \eta$.

For arbitrary $q \geq 2$, Feng, Guo, Yin and Zhang [FGYZ22] extend the definition of the influence matrix of μ and introduce the generalized version of the spectral independence.

Definition 1.4 (Spectral Independence)**.** Let μ be a distribution over $\Omega \subseteq [q]^n$. For any $\Lambda \subseteq [n]$ and every feasible pinning $\tau \in [q]^\Lambda$, the *absolute influence matrix* $\Psi_\mu^\tau \in \mathbb{R}_{\geq 0}^{n\times n}$ is defined as, for every distinct $u, v \in [n]$,

$$\Psi_\mu^\tau(u,v) := \inf_{i,j\in[q]} \mathcal{D}_{\mathrm{TV}}\left(\mu_v^{\tau\cup\{u\leftarrow i\}} \,\middle\|\, \mu_v^{\tau\cup\{u\leftarrow j\}}\right).$$

For $\eta > 0$, we say μ is η-spectrally independent if for all $\Lambda \subseteq [n]$ and $\tau \in [q]^\Lambda$, the spectral radius of the absolute influence matrix satisfies $\rho(\Psi_\mu^\tau) \leq 1 + \eta$.

Remark 1.5. In some cases, we also define the influence matrix $\widetilde{\Psi_\mu}$ as

$$\widetilde{\Psi_\mu}((i,s),(j,t)) := \mathbf{Pr}_{\omega\sim\mu}\left[\omega(j) = t \mid \omega(i) = s\right] - \mathbf{Pr}_{\omega\sim\mu}\left[\omega(j) = t\right].$$

It is well-known that $\lambda_{\max}\left(\widetilde{\Psi_\mu}\right) \leq \rho(\Psi_\mu)$.

The following argument relates the influence matrix to the correlation of the distribution. This might explain the motivation and the intuition that we take the spectral independence into account and serve it as a local property of the distribution.

Lemma 1.6. *Given a distribution μ over $\Omega \subseteq \{-1,+1\}^n$, define the correlation matrix of μ as*

$$\mathbf{Cor}(\mu) := \mathrm{diag}\left(\mathbf{Cov}\left(\mu\right)\right)^{-1/2} \mathbf{Cov}\left(\mu\right) \mathrm{diag}\left(\mathbf{Cov}\left(\mu\right)\right)^{-1/2}.$$

Then $\Psi_\mu = \mathbf{Cov}\left(\mu\right) \mathrm{diag}\left(\mathbf{Cov}\left(\mu\right)\right)^{-1}$ *and*

$$\left\|\Psi_\mu\right\|_{\mathrm{OP}} = \left\|\mathbf{Cor}(\mu)\right\|_{\mathrm{OP}}.$$

Proof. Let X be a random variable drawn from μ. For $i, j \in [n]$, by calculation,

$$\begin{aligned}
\mathbf{Cov}\left(\mu\right)_{i,j} &= \mathbf{E}\left[X_i X_j\right] - \mathbf{E}\left[X_i\right]\mathbf{E}\left[X_j\right] \\
&= \mathbf{E}\left[X_i \mid X_j = 1\right]\mathbf{Pr}\left[X_j = 1\right]\left(1 - \mathbf{E}\left[X_j\right]\right) - \mathbf{E}\left[X_i \mid X_j = -1\right]\mathbf{Pr}\left[X_j = 1\right]\left(1 + \mathbf{E}\left[X_j\right]\right) \\
&= \Psi_\mu(i,j)\left(1 - \mathbf{E}\left[X_j\right]^2\right)
\end{aligned}$$

thus leading to the identity $\Psi_\mu = \mathbf{Cov}\left(\mu\right)\mathrm{diag}\left(\mathbf{Cov}\left(\mu\right)\right)^{-1}$. To prove the second identity, let $\mathbf{v}$ be an eigenvector of $\mathbf{Cor}(\mu)$ and its associated eigenvalue is λ. For simplicity let $D = \mathrm{diag}\left(\mathbf{Cov}\left(\mu\right)\right)$. Then,

$$\lambda\mathbf{v} = D^{-1/2}\mathbf{Cov}\left(\mu\right)D^{-1/2}\mathbf{v}.$$

Let $\mathbf{u} = D^{1/2}\mathbf{v}$. Thus we obtain

$$\lambda\mathbf{u} = \mathbf{Cov}\left(\mu\right)D^{-1/2}\mathbf{v} = \mathbf{Cov}\left(\mu\right)D^{-1}\mathbf{u} = \Psi_\mu\mathbf{u}.$$

Then we know Ψ_μ and $\mathbf{Cor}(\mu)$ share the same spectrum, meaning that their operator norms are equal. □

2 High-Dimensional Expander: Simplicial Complex

Now we introduce a framework relate the local property to the global rate of the mixing of Markov chains.

Definition 2.1 (Simplicial Complex)**.** A simplicial complex $\mathscr{C}$ is a non-empty downwards closed collection of sets (called faces) over a finite ground set of elements. It satisfies

- $\varnothing \in \mathscr{C}$;
- if $S \in \mathscr{C}$ and $T \subseteq S$, then $T \in \mathscr{C}$.

Additionally, we assume that $\mathscr{C}$ is pure, *i.e.*, for all maximal elements $S \in \mathscr{C}$, they share the same size denoted by $d = \mathrm{rank}(\mathscr{C})$. For all $S \in \mathscr{C}$, let $\mathrm{rank}(S) := |S|$. According to the rank function, we partition $\mathscr{C}$ into $d+1$ parts as: for every $0 \le k \le d$, define the k-skeleton as

$$\mathscr{C}(k) := \{S \in \mathscr{C} \mid \mathrm{rank}(S) = k\}.$$

For every face $S \in \mathscr{C}$, define the link at S as

$$\mathscr{C}_S := \{T \in \mathscr{C} \mid S \cap T = \varnothing, S \cup T \in \mathscr{C}\}$$

and for all $0 \le k \le d - \mathrm{rank}(S)$, define the k-skeleton at S as

$$\mathscr{C}_S(k) := \{T \in \mathscr{C}_S \mid \mathrm{rank}(T) = k\}.$$

2.1 Weight functions and random walks on the simplicial complex

Given a distribution μ over $\Omega = \mathscr{C}(d)$, we define the weight function $w : \mathscr{C} \to \mathbb{R}_{\ge 0}$ as

$$w(S) = \begin{cases} \mu(S) & S \in \mathscr{C}(d); \\ \sum_{T \supseteq S, T \in \mathscr{C}(k+1)} w(T) & S \in \mathscr{C}(k), k < d. \end{cases}$$

For every link $\mathscr{C}_S$ at face $S \in \mathscr{C}$, we define $w_S(T) = w(S \cup T)$ for every $T \in \mathscr{C}_S$.

To see the random walks on $\mathscr{C}$, firstly we introduce the distribution on it. For every $0 \le k \le d$, we define the distribution π_d on $\mathscr{C}(k)$ as

$$\pi_k(S) = \frac{w(S)}{\sum_{T \in \mathscr{C}(k)} w(T)}, \quad \forall S \in \mathscr{C}(k).$$

Similarly for the link at S, we can also define the distribution $\pi_{S,k}$ over $\mathscr{C}_S(k)$.

For $0 \le k \le d$ and every $S \in \mathscr{C}(k)$, by calculation,

$$w(S) = \frac{d!}{k!}\mu(S).$$

This leads to the identity

$$\pi_k(S) = \frac{w(S)}{\sum_{T \in \mathscr{C}(k)} w(T)} = \frac{\mu(S)}{\sum_{T \in \mathscr{C}(k)} \mu(T)} = \frac{1}{\binom{d}{k}}\mu(S).$$

For simplicity of notations and analysis, we assume that the dimension of all the matrices is $\mathscr{C}(1)$, and we add zeros to appropriate positions. For every $0 \le k \le d$, we define $\Pi_k := \mathrm{diag}(\pi_k)$ to be the diagonal matrix induced by π_k, and similarly define $\Pi_{S,k} \in \mathbb{R}^{\mathscr{C}_\tau(k) \times \mathscr{C}_\tau(k)}$ for all links at $S \in \mathscr{C}$ and $0 \le k \le d - \mathrm{rank}(S)$, and the inverse of them means taking inverse only on their non-zero entries. Additionally, we use the operator $\bar{\cdot}$ to denote the actual vector or matrix in the simplicial complex. Precisely speaking, for a matrix A supported on $\mathscr{C}_\tau(k) \times \mathscr{C}_\tau(k)$,

$$\overline{A}(S \cup T, S \cup R) := A(T, R), \quad \forall T, R \in \mathscr{C}_\tau(k)$$

and 0 otherwise, meanwhile for a vector $\mathbf{v}$ supported on $\mathscr{C}_\tau(k)$,

$$\overline{\mathbf{v}}(S \cup T) := \mathbf{v}(T), \quad \forall T \in \mathscr{C}_\tau(k)$$

and 0 otherwise.

There are two natural random walks on the simplicial complex $\mathscr{C}$: up-walk and down-walk.

- 'Up-Walk' $\mathrm{P}_k^{\uparrow}$: starting from $S \in \mathscr{C}(k)$, we add an element $x \in \mathscr{C}_S(1)$ as $\pi_{S,1}$.
- 'Down-Walk' $\mathrm{P}_k^{\downarrow}$: starting from $S \in \mathscr{C}(k)$, we remove an element $x \in S$ uniformly at random.

We write them in a explicit form: for $0 \le k \le d-1$, $S \in \mathscr{C}(k)$, $T \in \mathscr{C}(k+1)$,

$$\mathrm{P}_k^{\uparrow}(S,T) = \frac{w(T)}{w(S)} \mathbb{1}\left[S \subseteq T\right]$$

and for $1 \le k \le d$, $S \in \mathscr{C}(k)$, $T \in \mathscr{C}(k-1)$,

$$\mathrm{P}_k^{\downarrow}(S,T) = \frac{1}{k} \mathbb{1}\left[T \subseteq S\right].$$

Based on the two walks, we define the following up-down walk and down-up walk (note that they are all lazy random walks):

$$\begin{aligned} \mathrm{P}_k^{\Delta} &= \mathrm{P}_k^{\uparrow}\mathrm{P}_{k+1}^{\downarrow}, \quad \forall 0 \le k \le d-1, \\ \mathrm{P}_k^{\nabla} &= \mathrm{P}_k^{\downarrow}\mathrm{P}_{k-1}^{\uparrow}, \quad \forall 1 \le k \le d. \end{aligned}$$

For the up-down walks, usually we consider its non-lazy version $\mathrm{P}_k^{\wedge} := \frac{k+1}{k}\mathrm{P}_k^{\Delta} - \frac{1}{k}I$. For the link $\mathscr{C}_\tau$ at $\tau \in \mathscr{C}$, it is similar to define the random walks $\mathrm{P}_{\tau,k}^{\Delta}$, $\mathrm{P}_{\tau,k}^{\nabla}$ and $\mathrm{P}_{\tau,k}^{\wedge}$. Among all these walks, we pay quite a special attention to the local walk $\mathrm{P}_{\tau,1}^{\wedge}$ and $\mathrm{P}_{\tau,1}^{\nabla}$. Define the matrix $W_{\tau,2}$ supported on $\mathscr{C}_\tau(1) \times \mathscr{C}_\tau(1)$ as $W_{\tau,2}(x,y) = \pi_{\tau,2}(\{x,y\})$ for $x, y \in \mathscr{C}_\tau(1)$ and $\{x,y\} \in \mathscr{C}_\tau(2)$. By definition, it holds that

$$\mathrm{P}_{\tau,1}^{\wedge} = \frac{1}{2}\Pi_{\tau,1}^{-1}W_{\tau,2}, \tag{2}$$

$$\mathrm{P}_{\tau,1}^{\nabla} = \mathbf{1}\pi_{\tau,1}^{\top}. \tag{3}$$

Moreover, directly from the definition, for the distributions of the two adjacent layers, it holds that

$$\Pi_{k+1}\mathrm{P}_{k+1}^{\downarrow} = \left(\mathrm{P}_k^{\uparrow}\right)^{\top}\Pi_k, \quad \forall 0 \le k \le d-1. \tag{4}$$

Multiplying all-ones vector on both sides, we obtain,

$$\pi_{k+1}\mathrm{P}_{k+1}^{\downarrow} = \pi_k, \tag{5}$$

$$\pi_k\mathrm{P}_k^{\uparrow} = \pi_{k+1}. \tag{6}$$

For every $0 \le \ell \le k \le d$ and $\sigma \in \mathscr{C}(k)$, $\tau \in \mathscr{C}(\ell)$ with $\tau \subseteq \sigma$, by definition,

$$\begin{aligned} \pi_k(\sigma) &= \frac{1}{\binom{d}{k}}\mu(\sigma) \\ &= \frac{1}{\binom{d}{k}}\mu(\tau)\mu^{\tau}(\sigma \setminus \tau) \\ &= \binom{k}{\ell}\pi_\ell(\tau)\pi_{\tau,k-\ell}(\sigma \setminus \tau). \end{aligned}$$

2.2 Garland's method

The kernel of the local-to-global theorem is to establish the relationship between local walks and global walks. The Garland's method is implicit in the work of Oppenheim [Opp18] and we put it in a more direct and explicit form.

Lemma 2.2 (Garland's Method)**.** *The following identities hold:*

1. $\Pi_1 = \mathbf{E}_{\tau\sim\pi_1}\left[\Pi_{\tau,1}\right]$.
2. $\Pi_1\mathrm{P}_1^{\wedge} = \mathbf{E}_{\tau\sim\pi_1}\left[\Pi_{\tau,1}\mathrm{P}_{\tau,1}^{\wedge}\right]$.
3. $\Pi_1(\mathrm{P}_1^{\wedge})^2 = \mathbf{E}_{\tau\sim\pi_1}\left[\pi_{\tau,1}\pi_{\tau,1}^{\top}\right]$.

Proof. We prove these identities entry by entry.

1. For every $x \in \mathscr{C}(1)$, by definition,

$$\begin{aligned}\Pi_1(x,x) &= \pi_1(x) \\ &= \sum_{\{x,y\}\in\mathscr{C}(1)} \frac{1}{2}\pi_2(\{x,y\}) \\ &= \sum_{y\in\mathscr{C}(1)} \pi_1(y)\frac{\pi_2(\{x,y\})}{2\pi_1(y)} \\ &= \mathbf{E}_{y\sim\pi_1}\left[\Pi_{y,1}(x,x)\right].\end{aligned}$$

2. By (2), it holds that

$$\mathbf{E}_{\tau\sim\pi_1}\left[\Pi_{\tau,1}\mathrm{P}^{\wedge}_{\tau,1}\right] = \mathbf{E}_{\tau\sim\pi_1}\left[\frac{1}{2}W_{\tau,2}\right] = 2W_2.$$

   On the other hand, by definition,

$$\Pi_1\mathrm{P}^{\wedge}_1 = 2\Pi_1\mathrm{diag}\left(\pi_1\right)^{-1}W_2 = 2W_2.$$

   Then we conclude the identity.

3. For every $x, y \in \mathscr{C}(1)$,

$$\Pi_1(\mathrm{P}^{\wedge}_1)^2(x,y) = \sum_{\tau\in\mathscr{C}(1)} \pi_1(x)\pi_{x,1}(\tau)\pi_{\tau,1}(y).$$

   On the other hand,

$$\begin{aligned}\mathbf{E}_{\tau\sim\pi_1}\left[\pi_{\tau,1}\pi_{\tau,1}^{\top}\right](x,y) &= \sum_{\tau\in\mathscr{C}(1)} \pi_1(\tau)\pi_{\tau,1}(x)\pi_{\tau,1}(y) \\ &= \sum_{\tau\in\mathscr{C}(1)} \pi_1(x)\pi_{x,1}(\tau)\pi_{\tau,1}(y).\end{aligned}$$

   Thus we conclude the identity.

□

Additionally the following two identities between two skeletons are important.

Lemma 2.3. *The following identities hold*

1. $\Pi_k\mathrm{P}^{\wedge}_k = \mathbf{E}_{\tau\sim\pi_{k-1}}\left[\overline{\Pi_{\tau,1}\mathrm{P}^{\wedge}_{\tau,1}}\right]$.

2. $\Pi_k\mathrm{P}^{\vee}_k = \mathbf{E}_{\tau\sim\pi_{k-1}}\left[\overline{\Pi_{\tau,1}\mathrm{P}^{\vee}_{\tau,1}}\right] = \mathbf{E}_{\tau\sim\pi_{k-1}}\left[\overline{\pi_{\tau,1}\pi_{\tau,1}^{\top}}\right]$.

Proof. The proofs of two identities are similar.

1. By direct calculation,

$$\begin{aligned}\Pi_k\mathrm{P}^{\wedge}_k = \Pi_k \cdot \frac{1}{k}\sum_{\tau\in\mathscr{C}(k-1)} \overline{\mathrm{P}^{\wedge}_{\tau,1}} \\ &= \sum_{\tau\in\mathscr{C}(k-1)} \frac{1}{k}\Pi_k\overline{\mathrm{P}^{\wedge}_{\tau,1}} \\ &= \sum_{\tau\in\mathscr{C}(k-1)} \pi_{k-1}(\tau)\frac{\Pi_k}{k\pi_{k-1}(\tau)}\overline{\mathrm{P}^{\wedge}_{\tau,1}} \\ &= \sum_{\tau\in\mathscr{C}(k-1)} \pi_{k-1}(\tau)\overline{\Pi_{\tau,1}\mathrm{P}^{\wedge}_{\tau,1}}.\end{aligned}$$

2. Similarly to above, we have

$$\begin{aligned}\Pi_k \mathrm{P}_k^{\nabla} &= \Pi_k \cdot \frac{1}{k} \sum_{\tau \in \mathscr{C}(k-1)} \overline{\mathrm{P}_{\tau,1}^{\nabla}} \\ &= \sum_{\tau \in \mathscr{C}(k-1)} \pi_{k-1}(\tau) \frac{\Pi_k}{k\pi_{k-1}(\tau)} \overline{\mathrm{P}_{\tau,1}^{\nabla}} \\ &= \sum_{\tau \in \mathscr{C}(k-1)} \pi_{k-1}(\tau) \overline{\Pi_{\tau,1} \mathrm{P}_{\tau,1}^{\nabla}}.\end{aligned}$$

□

2.3 Trickling-down theorem

Based on the identities in Section 2.2, we establish more properties of the local walks.

Definition 2.4 (Local Spectral Expander). For a simplicial complex $\mathscr{C}$ equipped with distribution μ, for $0 \le k \le d-2$ and $\gamma_k \in [0,1]$, we say $\mathscr{C}(k)$ is a γ_k*-local spectral expander* if it holds that

$$\lambda_2(\mathrm{P}_{\tau,1}^{\wedge}) \le \gamma_k, \quad \forall \tau \in \mathscr{C}(k),$$

or equivalently,

$$\Pi_{\tau,1} \mathrm{P}_{\tau,1}^{\wedge} - \pi_{\tau,1} \pi_{\tau,1}^{\top} \preceq \gamma_k \left(\Pi_{\tau,1} - \pi_{\tau,1} \pi_{\tau,1}^{\top} \right).$$

Moreover, we say $\mathscr{C}$ is a $(\gamma_0, \dots, \gamma_{d-2})$*-local spectral expander* if $\mathscr{C}(k)$ is a γ_k-local spectral expander for all $0 \le k \le d-2$.

The following lemma shows, if $\mathscr{C}(k)$ is a local spectral expander, then we can see $\mathscr{C}(k-1)$ is also a local spectral expander.

Theorem 2.5 (Oppenheim's Trickling-Down Theorem, [Opp18]). *Suppose that $\mathscr{C}(k)$ is a γ-local spectral expander for some $1 \le k \le d-2$. Then $\mathscr{C}(k-1)$ is a $\frac{\gamma}{1-\gamma}$-local spectral expander (assuming the total connectivity of the random walk).*

Proof. When $k > 1$, we can only focus the link at each face in $\mathscr{C}(k)$ and this is the case $k = 1$. Then we assume that $k = 1$. By Lemma 2.2,

$$\begin{aligned}\Pi_1 \mathrm{P}_1^{\wedge} &= \mathbf{E}_{\tau \sim \pi_1} \left[\overline{\Pi_{\tau,1} \mathrm{P}_{\tau,1}^{\wedge}} \right] \\ &\preceq \mathbf{E}_{\tau \sim \pi_1} \left[\gamma \overline{\Pi_{\tau,1}} + (1-\gamma) \overline{\pi_{\tau,1} \pi_{\tau,1}^{\top}} \right] \\ &= \gamma \Pi_1 + (1-\gamma) \Pi_1 (\mathrm{P}_1^{\nabla})^2\end{aligned}$$

where the inequality comes from the fact that the local spectral expander means

$$\Pi_{\tau,1} - \Pi_{\tau,1} \mathrm{P}_{\tau,1}^{\wedge} \succeq (1-\gamma)(\Pi_{\tau,1} - \pi_{\tau,1} \pi_{\tau,1}^{\top}).$$

Now we consider the eigenvector $\mathbf{v}_2$ of P_1^{∇} with respect to the second largest eigenvalue λ_2. Then,

$$\lambda_2 \langle \mathbf{v}_2, \mathbf{v}_2 \rangle_{\pi_1} \le \gamma \langle \mathbf{v}_2, \mathbf{v}_2 \rangle_{\pi_1} + (1-\gamma) \lambda_2^2 \langle \mathbf{v}_2, \mathbf{v}_2 \rangle_{\pi_1}.$$

This means $(1-\lambda_2)((1-\gamma)\lambda_2 - \gamma) \le 0$. Since $\lambda_2 < 1$ (otherwise the bound is meaningless), we have $\lambda_2 \le \frac{\gamma}{1-\gamma}$. □

Observe that in Theorem 2.5, we only consider the second largest eigenvalue of the local walk. When we take more eigenvalues into account, the improved trickling-down theorem is introduced in Abdolazimi, Liu and Oveis Gharan [ALOG21].

Theorem 2.6 (Matrix Trickling-Down Theorem, [ALOG21]). *Given a simplicial complex $\mathscr{C}$ equipped with distribution μ, suppose that the following holds:*

1. $\lambda_2(\mathrm{P}_1^{\wedge}) < 1$, i.e., $\mathrm{P}_1^{\wedge}$ *is irreducible.*

2. *There exists a family of symmetric matrices* $\left\{M_\tau \in \mathbb{R}^{\mathscr{C}(1)\times\mathscr{C}(1)}\right\}_{\tau\in\mathscr{C}(1)}$ *such that*

$$\Pi_{\tau,1}\mathrm{P}_{\tau,1}^{\wedge} - \alpha\pi_{\tau,1}\pi_{\tau,1}^{\top} \preceq M_\tau \preceq \frac{1}{2\alpha+1}\Pi_{\tau,1}$$

for all $\tau \in \mathscr{C}(1)$.

Then for every $M \in \mathbb{R}^{\mathscr{C}(1)\times\mathscr{C}(1)}$ *satisfying* $M \preceq \frac{1}{2\alpha}\Pi_1$ *and* $\mathbf{E}_{\tau\sim\pi_{\tau,1}}[M_\tau] \preceq M - \alpha M\Pi_1^{-1}M$*, it holds that*

$$\Pi_1\mathrm{P}_1^{\wedge} - \left(2 - \frac{1}{\alpha}\right)\pi_1\pi_1^{\top} \preceq M.$$

In particular, $\lambda_2(\mathrm{P}_1^{\wedge}) \le \rho(\Pi_1^{-1}M)$.

Proof. We take expectation on all sides of the assumption and by Lemma 2.2,

$$\Pi_1\mathrm{P}_1^{\wedge} - \alpha\Pi_1(\mathrm{P}_1^{\wedge})^2 \preceq \mathbf{E}_{\tau\sim\pi_1}[M_\tau] \preceq \frac{1}{2\alpha+1}\Pi_1.$$

Therefore, $\Pi_1\mathrm{P}_1^{\wedge} - \alpha\Pi_1(\mathrm{P}_1^{\wedge})^2 \preceq M - \alpha M\Pi_1^{-1}M$. Set $Q = \mathrm{P}_1^{\wedge} - \beta\cdot\mathbf{1}\pi_1^{\top}$ with $\beta = 2 - \frac{1}{\alpha}$. Then we know

$$\Pi_1\mathrm{P}_1^{\wedge} - \alpha\Pi_1(\mathrm{P}_1^{\wedge})^2 = \Pi_1 Q - \alpha\Pi_1 Q^2.$$

Thus we know

$$\Pi_1 Q - \alpha\Pi_1 Q^2 \preceq M - \alpha M\Pi_1^{-1}M.$$

Since $\lambda_2(\mathrm{P}_{\tau,1}^{\wedge}) \le \frac{1}{2\alpha+1}$ for every $\tau \in \mathscr{C}(1)$, by Theorem 2.5, $\lambda_2(\mathrm{P}_1^{\wedge}) \le \frac{1}{2\alpha}$. Combined with $\beta = 2 - \frac{1}{\alpha} \ge 1 - \frac{1}{2\alpha}$ we have $Q \preceq \frac{1}{2\alpha}I$. By Lemma 2.3 in [ALOG21], we have $\Pi_1 Q \preceq M$. □

Commonly in use we apply the following proposition induced by Theorem 2.6.

Proposition 2.7. *Given a simplicial complex* $\mathscr{C}$ *equipped with distribution* μ*, if there exists a family of symmetric matrices* $\left\{M_\tau \in \mathbb{R}^{\mathscr{C}(1)\times\mathscr{C}(1)}\right\}_{\tau\in\mathscr{C}}$ *satisfying*

1. ***Base Cases:*** *For every* $\tau \in \mathscr{C}(d-2)$,

$$\Pi_{\tau,1}\mathrm{P}_{\tau,1}^{\wedge} - 2\pi_{\tau,1}\pi_{\tau,1}^{\top} \preceq M_\tau \preceq \frac{1}{5}\Pi_{\tau,1}.$$

2. ***Recursive Conditions:*** *For every* $\tau \in \mathscr{C}(d-k)$ *with* $k \ge 3$*, one of the followings holds:*

   - *The matrices satisfy*

$$M_\tau \preceq \frac{k-1}{3k-1}\Pi_{\tau,1}, \quad \mathbf{E}_{x\sim\pi_{\tau,1}}\left[M_{\tau\cup\{x\}}\right] \preceq M_\tau - \frac{k-1}{k-2}M_\tau\Pi_{\tau,1}^{-1}M_\tau.$$

   - $(\mathscr{C}_\tau, \pi_{\tau,k})$ *is the product of m pure weighted simplicial complexes* $(\mathscr{C}^{(1)}, \pi^{(1)}), \ldots, (\mathscr{C}^{(m)}, \pi^{(m)})$ *of dimension* $d_1, \ldots, d_m$ *respectively and,*

$$M_\tau = \sum_{j=1}^{m}\frac{d_j(d_j-1)}{k(k-1)}M_{\tau\cup\eta_j}$$

   where $\eta_j = \eta \setminus \mathscr{C}^{(j)}(1)$ *for an arbitrary* $\eta \in \mathscr{C}_\tau(k)$.

Then for every $\tau \in \mathscr{C}(d-k)$ *with* $k \ge 2$*, it holds that*

$$\Pi_{\tau,1}\mathrm{P}_{\tau,1}^{\wedge} - \frac{k}{k-1}\pi_{\tau,1}\pi_{\tau,1}^{\top} \preceq M_\tau \preceq \frac{k-1}{3k-1}\Pi_{\tau,1}.$$

In particular, $\lambda_2(\mathrm{P}_{\tau,1}^{\wedge}) \le \rho(\Pi_{\tau,1}^{-1}M_\tau)$ *for all* $\tau \in \mathscr{C}(d-k)$ *with* $k \ge 2$.

2.4 The local-to-global theorem

Now we are ready to introduce the method named *Alev-Lau's Local-to-Global Theorem* introduced in Alev and Lau [AL20].

Theorem 2.8 (Alev-Lau's Local-to-Global Theorem, [AL20]). *Assume that $\mathscr{C}$ is an $(\alpha_0, \ldots, \alpha_{d-2})$-local spectral expander. Then for any $1 \le k \le d$, it holds that*

$$\mathrm{Gap}(\mathrm{P}_k^{\nabla}) = \mathrm{Gap}(\mathrm{P}_{k-1}^{\Delta}) \ge \frac{1}{k}\prod_{i=0}^{k-2}(1-\alpha_i).$$

Proof. It suffices to show that for all $1 \le k \le d-1$,

$$\mathrm{Gap}(\mathrm{P}_{k+1}^{\nabla}) = \mathrm{Gap}(\mathrm{P}_k^{\Delta}) \ge \frac{k}{k+1}(1-\alpha_{k-1})\mathrm{Gap}(\mathrm{P}_k^{\nabla}). \tag{7}$$

Together with the hypothesis induction and $\mathrm{Gap}(\mathrm{P}_1^{\nabla}) = 1$ we can conclude the theorem.

To prove (7), firstly observe that $\mathrm{P}_{k+1}^{\nabla}$ and P_k^{Δ} share the same non-zero eigenvalues and thus their spectral gaps are the same. By Lemma 2.3, for every $1 \le k \le d-1$,

$$\Pi_k \mathrm{P}_k^{\wedge} - \Pi_k \mathrm{P}_k^{\nabla} = \mathbf{E}_{\tau\sim\pi_{k-1}}\left[\overline{\Pi_{\tau,1}\mathrm{P}_{\tau,1}^{\wedge}} - \overline{\pi_{\tau,1}\pi_{\tau,1}^{\top}}\right].$$

Since $\mathscr{C}$ is an $(\alpha_0, \ldots, \alpha_{d-2})$-local spectral expander, for every $\tau \in \mathscr{C}(k-1)$, the local walks satisfy:

$$\Pi_{\tau,1}\mathrm{P}_{\tau,1}^{\wedge} - \pi_{\tau,1}\pi_{\tau,1}^{\top} \preceq \alpha_{k-1}(\Pi_{\tau,1} - \pi_{\tau,1}\pi_{\tau,1}^{\top}).$$

Plugging it into above, by Lemmas 2.2 and 2.3, we obtain

$$\begin{aligned}\Pi_k \mathrm{P}_k^{\wedge} - \Pi_k \mathrm{P}_k^{\nabla} &\preceq \alpha_{k-1}\mathbf{E}_{\tau\sim\pi_{k-1}}\left[\overline{\Pi_{\tau,1} - \pi_{\tau,1}\pi_{\tau,1}^{\top}}\right]\\ &= \alpha_{k-1}(\Pi_k - \Pi_k \mathrm{P}_k^{\nabla}).\end{aligned}$$

This means

$$\Pi_k(I_{\mathscr{C}(k)} - \mathrm{P}_k^{\wedge}) \succeq (1-\alpha_{k-1})\Pi_k(I_{\mathscr{C}(k)} - \mathrm{P}_k^{\nabla})$$

thus leading to $\mathrm{Gap}(\mathrm{P}_k^{\wedge}) \ge (1-\alpha_{k-1})\mathrm{Gap}(\mathrm{P}_k^{\nabla})$. To finish the proof, note that

$$\mathrm{P}_k^{\Delta} = \frac{k}{k+1}\mathrm{P}_k^{\wedge} + \frac{1}{k+1}I_{\mathscr{C}(k)},$$

meaning that

$$\mathrm{Gap}(\mathrm{P}_k^{\Delta}) = \frac{k}{k+1}\mathrm{Gap}(\mathrm{P}_k^{\wedge}) \ge \frac{k}{k+1}(1-\alpha_{k-1})\mathrm{Gap}(\mathrm{P}_k^{\nabla}).$$

□

Remark 2.9. Note that, the random walk P_d^{∇} is the typical single-site Glauber dynamics P^{GD}. When μ is η-spectrally independent, it was shown in [ALOG20] that the corresponding simplicial complex $\mathscr{C}$ equipped with μ is a $(\frac{\eta}{d-1}, \frac{\eta}{d-2}, \ldots, \eta)$-local spectral expander. Then we know the spectral gap of P^{GD} is bounded by

$$\mathrm{Gap}(P^{\mathrm{GD}}) \ge \frac{1}{d}\prod_{i=0}^{k-2}\left(1 - \frac{\eta}{d-i-1}\right).$$

By Lemma 1.2 we can show the mixing rate of P^{GD}.

3 Variance and Entropy Contraction

Now we give an alternative view of Theorem 2.8. Given a simplicial complex $\mathscr{C}$ of dimension d, for $0 \leq \ell < k \leq d$, we define the following random walks

$$\begin{aligned}
\mathrm{P}^{\downarrow}_{k\to\ell} &= \mathrm{P}^{\downarrow}_{k}\mathrm{P}^{\downarrow}_{k-1}\dots\mathrm{P}^{\downarrow}_{\ell+1},\\
\mathrm{P}^{\uparrow}_{\ell\to k} &= \mathrm{P}^{\uparrow}_{\ell}\mathrm{P}^{\uparrow}_{\ell+1}\dots\mathrm{P}^{\uparrow}_{k-1},\\
\mathrm{P}^{\nabla}_{k\leftrightarrow\ell} &= \mathrm{P}^{\downarrow}_{k\to\ell}\mathrm{P}^{\uparrow}_{\ell\to k},\\
\mathrm{P}^{\Delta}_{\ell\leftrightarrow k} &= \mathrm{P}^{\uparrow}_{\ell\to k}\mathrm{P}^{\downarrow}_{k\to\ell}.
\end{aligned}$$

Then we consider the Dirichlet form of $\mathrm{P}^{\nabla}_{k\leftrightarrow\ell}$. By definition,

$$\begin{aligned}
\mathcal{E}_{\mathrm{P}^{\nabla}_{k\leftrightarrow\ell}}(f,f) &= \left\langle f,\left(I-\mathrm{P}^{\nabla}_{k\leftrightarrow\ell}\right)f\right\rangle_{\pi_k}\\
&= f^{\top}\Pi_k f - f^{\top}\Pi_k\mathrm{P}^{\nabla}_{k\leftrightarrow\ell}f\\
&= f^{\top}\Pi_k f - f^{\top}\Pi_k\mathrm{P}^{\downarrow}_{k\to\ell}\mathrm{P}^{\uparrow}_{\ell\to k}f\\
&\overset{(a)}{=} f^{\top}\Pi_k f - f^{\top}\left(\mathrm{P}^{\uparrow}_{\ell\to k}\right)^{\top}\Pi_\ell\mathrm{P}^{\uparrow}_{\ell\to k}f\\
&\overset{(b)}{=} \mathbf{Var}_{\pi_k}(f) - \mathbf{Var}_{\pi_\ell}\left(\mathrm{P}^{\uparrow}_{\ell\to k}f\right)
\end{aligned}$$

where (a) holds from (4) and (b) holds from (6). Similarly, for the Dirichlet form of $\mathrm{P}^{\Delta}_{\ell\leftrightarrow k}$, we have

$$\begin{aligned}
\mathcal{E}_{\mathrm{P}^{\Delta}_{\ell\leftrightarrow k}}(f,f) &= \left\langle f,\left(I-\mathrm{P}^{\Delta}_{\ell\leftrightarrow k}\right)f\right\rangle_{\pi_\ell}\\
&= f^{\top}\Pi_\ell f - f^{\top}\Pi_\ell\mathrm{P}^{\Delta}_{\ell\leftrightarrow k}f\\
&= f^{\top}\Pi_\ell f - f^{\top}\Pi_\ell\mathrm{P}^{\uparrow}_{\ell\to k}\mathrm{P}^{\downarrow}_{k\to\ell}f\\
&= \mathbf{Var}_{\pi_\ell}(f) - \mathbf{Var}_{\pi_k}\left(\mathrm{P}^{\downarrow}_{k\to\ell}f\right).
\end{aligned}$$

Follow the similar routine together with Jenssen's inequality, and we obtain the following identities and inequalities:

$$\mathcal{E}_{\mathrm{P}^{\nabla}_{k\leftrightarrow\ell}}(f,f) = \mathbf{Var}_{\pi_k}(f) - \mathbf{Var}_{\pi_\ell}\left(\mathrm{P}^{\uparrow}_{\ell\to k}f\right), \tag{8}$$

$$\mathcal{E}_{\mathrm{P}^{\Delta}_{\ell\leftrightarrow k}}(f,f) = \mathbf{Var}_{\pi_\ell}(f) - \mathbf{Var}_{\pi_k}\left(\mathrm{P}^{\downarrow}_{k\to\ell}f\right), \tag{9}$$

$$\mathcal{E}_{\mathrm{P}^{\nabla}_{k\leftrightarrow\ell}}(f,\log f) \geq \mathbf{Ent}_{\pi_k}[f] - \mathbf{Ent}_{\pi_\ell}\left[\mathrm{P}^{\uparrow}_{\ell\to k}f\right], \tag{10}$$

$$\mathcal{E}_{\mathrm{P}^{\Delta}_{\ell\leftrightarrow k}}(f,\log f) \geq \mathbf{Ent}_{\pi_\ell}[f] - \mathbf{Ent}_{\pi_k}\left[\mathrm{P}^{\downarrow}_{k\to\ell}f\right]. \tag{11}$$

The identities or inequalities as above show us that, when we want to show a Poincaré's inequality, it suffices to show the variance/entropy contraction.

Lemma 3.1 ([CGM21]). *Let $0 \leq \ell \leq k \leq d$ and $f^{(k)} : \mathscr{C}(k) \to \mathbb{R}_{\geq 0}$ be a function on $\mathscr{C}(k)$. Then*

$$\mathbf{Ent}_{\pi_k}\left[f^{(k)}\right] = \mathbf{E}_{\tau\sim\pi_\ell}\left[\mathbf{Ent}_{\pi_{\tau,k-\ell}}\left[f^{(k-\ell)}_{\tau}\right]\right] + \mathbf{Ent}_{\pi_\ell}\left[f^{(\ell)}\right]$$

where $f^{(k-\ell)}_{\tau}(\sigma) := f^{(k)}(\tau\cup\sigma)$ for every $\sigma\in\mathscr{C}_\tau(k-\ell)$, and $f^{(\ell)} := \mathrm{P}^{\uparrow}_{\ell}\mathrm{P}^{\uparrow}_{\ell+1}\dots\mathrm{P}^{\uparrow}_{k-1}f^{(k)}$.

Similarly, for all $f^{(k)} : \mathscr{C}(k) \to \mathbb{R}$, it holds that

$$\mathbf{Var}_{\pi_k}\left(f^{(k)}\right) = \mathbf{E}_{\tau\sim\pi_\ell}\left[\mathbf{Var}_{\pi_{\tau,k-\ell}}\left(f^{(k-\ell)}_{\tau}\right)\right] + \mathbf{Var}_{\pi_\ell}\left(f^{(\ell)}\right).$$

Proof. We prove the identity for variance and the identity for entropy is similar. Note that, as a simple extended version of Lemma 2.3, it holds that

$$\Pi_k = \mathbf{E}_{\tau\sim\pi_\ell}\left[\overline{\Pi_{\tau,k-\ell}}\right].$$

Without loss of generality, assume that $\mathbf{E}_{\pi_k}\left[f^{(k)}\right] = 0$. Then,

$$\begin{aligned}
\mathbf{Var}_{\pi_k}\left(f^{(k)}\right) &= \left(f^{(k)}\right)^\top \Pi_k f^{(k)} \\
&= \left(f^{(k)}\right)^\top \mathbf{E}_{\tau\sim\pi_\ell}\left[\overline{\Pi_{\tau,k-\ell}}\right] f^{(k)} \\
&= \mathbf{E}_{\tau\sim\pi_\ell}\left[\left(f_\tau^{(k-\ell)}\right)^\top \Pi_{\tau,k-\ell} f_\tau^{(k-\ell)}\right] \\
&= \mathbf{E}_{\tau\sim\pi_\ell}\left[\mathbf{Var}_{\pi_{\tau,k-\ell}}\left(f_\tau^{(k-\ell)}\right)\right] + \mathbf{E}_{\tau\sim\pi_\ell}\left[\mathbf{E}_{\pi_{\tau,k-\ell}}\left[\left(f_\tau^{(k-\ell)}\right)^2\right]\right] \\
&= \mathbf{E}_{\tau\sim\pi_\ell}\left[\mathbf{Var}_{\pi_{\tau,k-\ell}}\left(f_\tau^{(k-\ell)}\right)\right] + \mathbf{Var}_{\pi_\ell}\left(f^{(\ell)}\right).
\end{aligned}$$

For entropy the proof is similar and we can just assume that $\mathbf{E}_{\pi_k}\left[f^{(k)}\right] = 1$. □

3.1 Variance tensorizations

To establish a Poincaré inequality via spectral independence, it's time to introduce the *tensorization of variance*. The kernel of this method is law of total covariance.

Theorem 3.2 (Law of Total Covariance). *Let X, Y, Z be three random variables. Then it holds that*

$$\mathbf{Cov}(X, Y) = \mathbf{E}[\mathbf{Cov}(X, Y \mid Z)] + \mathbf{Cov}(\mathbf{E}[X \mid Z], \mathbf{E}[Y \mid Z]).$$

Proof. By law of total expectation, it holds that

$$\begin{aligned}
\mathbf{Cov}(X, Y) &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] \\
&= \mathbf{E}[\mathbf{E}[XY \mid Z]] - \mathbf{E}[\mathbf{E}[X \mid Z]]\mathbf{E}[\mathbf{E}[Y \mid Z]] \\
&= \mathbf{E}[\mathbf{Cov}(X, Y \mid Z) + \mathbf{E}[X \mid Z]\mathbf{E}[Y \mid Z]] - \mathbf{E}[\mathbf{E}[X \mid Z]]\mathbf{E}[\mathbf{E}[Y \mid Z]] \\
&= \mathbf{E}[\mathbf{Cov}(X, Y \mid Z)] + \mathbf{Cov}(\mathbf{E}[X \mid Z], \mathbf{E}[Y \mid Z]).
\end{aligned}$$

□

Note that when we consider $\sigma \sim \mu$ where μ is the distribution over $\Omega \subseteq [q]^n$ and let $X = Y = f(\omega), Z = \omega(\Lambda)$ for a fixed arbitrary subset $\Lambda \subseteq [n]$, it holds that

$$\mathbf{Var}_\mu(f) = \mathbf{E}_{\sigma_\Lambda\sim\mu_\Lambda}\left[\mathbf{Var}_{\mu^{\sigma_\Lambda}}\left(f^{\sigma_\Lambda}\right)\right] + \mathbf{Var}_{\sigma_\Lambda\sim\mu_\Lambda}\left(\mathbf{E}_{\mu^{\sigma_\Lambda}}\left[f^{\sigma_\Lambda}\right]\right).$$

Given $1 \le \ell \le n$, adding all identities for $\Lambda \in \binom{[n]}{\ell}$, it holds that

$$\mathbf{Var}_\mu(f) = \frac{1}{\binom{n}{\ell}} \sum_{\Lambda\subseteq[n],|\Lambda|=\ell} \left(\mathbf{E}_{\sigma_\Lambda\sim\mu_\Lambda}\left[\mathbf{Var}_{\mu^{\sigma_\Lambda}}\left(f^{\sigma_\Lambda}\right)\right] + \mathbf{Var}_{\sigma_\Lambda\sim\mu_\Lambda}\left(\mathbf{E}_{\mu^{\sigma_\Lambda}}\left[f^{\sigma_\Lambda}\right]\right)\right).$$

To illustrate it in the form of law of total variance, consider the pinning set $\mathcal{P}_\ell$ defined as:

$$\mathcal{P}_\ell := \left\{(\Lambda, \sigma_\Lambda) \;\middle|\; \Lambda \in \binom{[n]}{\ell}, \sigma_\Lambda \in [q]^\Lambda\right\}.$$

Consider the distribution μ_ℓ on $\mathcal{P}_\ell$ defined as

$$\mu_\ell(\Lambda, \sigma_\Lambda) := \frac{1}{\binom{[n]}{\ell}} \mathbf{Pr}_{\omega\sim\mu}\left[\omega(\Lambda) = \sigma_\Lambda\right], \quad \forall (\Lambda, \sigma_\Lambda) \in \mathcal{P}_\ell,$$

and the function $f^{(\ell)} : \mathcal{P}_\ell \to \mathbb{R}$ as

$$f^{(\ell)}(\Lambda, \sigma_\Lambda) := \mathbf{E}_{\mu^{\sigma_\Lambda}}\left[f^{\sigma_\Lambda}\right].$$

Then by direct calculation, we obtain

$$\mathbf{Var}_\mu(f) = \mathbf{Var}_{\mu_\ell}\left(f^{(\ell)}\right) + \mathbf{E}_{(\Lambda,\sigma_\Lambda)\sim\mu_\ell}\left[\mathbf{Var}_{\mu^{\sigma_\Lambda}}\left(f^{\sigma_\ell}\right)\right]. \tag{12}$$

We remark here that (12) is exactly what we have shown in Lemma 3.1.

Definition 3.3 (Variance Independence)**.** We say a distribution μ is η-variance independent if for all functions $f : \Omega \to \mathbb{R}$,

$$\left(1 - \frac{1+\eta}{n}\right)\mathbf{Var}_\mu(f) \le \mathbf{E}_{(i,s)\sim\mathcal{P}_1}\left[\mathbf{Var}_{\mu^{i\leftarrow s}}(f)\right],$$

or equivalently

$$\mathbf{Var}_{\mu_1}\left(f^{(1)}\right) \le \frac{1+\eta}{n}\mathbf{Var}_\mu(f).$$

Lemma 3.4. *Let μ be a distribution over $\Omega \subseteq [q]^n$. Suppose that μ is η-spectrally independent. Then μ is η-variance independent.*

Proof. We consider the term $\mathbf{Var}_{\mu_1}\left(f^{(1)}\right)$. By definition,

$$\begin{aligned}\mathbf{Var}_{\mu_1}\left(f^{(1)}\right) &= \left\langle f^{(1)}, f^{(1)}\right\rangle_{\mu_1} - \left\langle f^{(1)}, \mathbf{1}\right\rangle_{\mu_1}^2 \\ &= f^\top\left(\frac{1}{n}\sum_{(i,s)\in\mathcal{P}_1}\mu_i(s)\left(\mu^{i\leftarrow s}\right)\left(\mu^{i\leftarrow s}\right)^\top\right)f - \langle f, \mathbf{1}\rangle_\mu^2.\end{aligned}$$

Now we define the random walk $\mathcal{R}_{\mu,1}$ over Ω as:

$$\mathcal{R}_{\mu,1} = \frac{1}{n}\sum_{(i,s)\in\mathcal{P}_1}\frac{1}{\mu_i(s)}\left(\mathbf{1}^{i\leftarrow s}\right)\left(\mathbf{1}^{i\leftarrow s}\right)^\top \mathrm{diag}(\mu).$$

It's not hard to observe that its stationary distribution is μ. Then we know

$$\frac{\mathbf{Var}_{\mu_1}\left(f^{(1)}\right)}{\mathbf{Var}_\mu(f)} = \frac{\left\langle f, \mathcal{R}_{\mu,1}f\right\rangle_\mu - \langle f, \mathbf{1}\rangle_\mu^2}{\langle f, f\rangle_\mu - \langle f, \mathbf{1}\rangle_\mu^2} \le \lambda_2(\mathcal{R}_{\mu,1}) = \lambda_2\left(\mathrm{diag}(\mu)^{1/2}\mathcal{R}_{\mu,1}\mathrm{diag}(\mu)^{-1/2}\right).$$

Then it suffices to bound the second largest eigenvalue of $\mathcal{R}_{\mu,1}$. Although $\mathcal{R}_{\mu,1}$ is the transition matrix in $\mathbb{R}^{\Omega\times\Omega}$, it has a decomposition as

$$\mathrm{diag}(\mu)^{1/2}\mathcal{R}_{\mu,1}\mathrm{diag}(\mu)^{-1/2} = \frac{1}{n}U_{\mu,1}U_{\mu,1}^\top$$

where $U_{\mu,1} \in \mathbb{R}^{\Omega\times\mathcal{P}_1}$ has columns $\mu_i(s)^{-1/2}\mathrm{diag}(\mu)^{1/2}\,\mathbf{1}^{i\leftarrow s}$ for each $(i,s) \in \mathcal{P}_1$. Then we only need to consider the eigenvalue of the matrix $\frac{1}{n}U_{\mu,1}^\top U_{\mu,1}$. By definition, for every $(i,s),(j,t) \in \mathcal{P}_1$,

$$\left(\frac{1}{n}U_{\mu,1}^\top U_{\mu,1}\right)((i,s),(j,t)) = \frac{1}{n}\frac{\mathbf{Pr}_{\omega\sim\mu}\left[\omega(i) = s \wedge \omega(j) = t\right]}{\sqrt{\mathbf{Pr}_{\omega\sim\mu}\left[\omega(i) = s\right]}\sqrt{\mathbf{Pr}_{\omega\sim\mu}\left[\omega(i) = s\right]}}.$$

Note that, this is the symmetrized version of the random walk $Q_{\mu,1}$ with stationary distribution μ_1, *i.e.*,

$$Q_{\mu,1}((i,s),(j,t)) = \frac{1}{n}\mathbf{Pr}_{\omega\sim\mu}\left[\omega(j) = t \mid \omega(i) = s\right].$$

Thus we know

$$\lambda_2(\mathcal{R}_{\mu,1}) = \lambda_2(Q_{\mu,1}) = \lambda_{\max}(Q_{\mu,1} - \mathbf{1}\mu_1^\top).$$

Observe that $Q_{\mu,1} - \mathbf{1}\mu_1^\top$ is exactly $\frac{1}{n}\widetilde{\Psi_\mu}$ where $\widetilde{\Psi_\mu}$ is defined as Remark 1.5. Then we conclude $\lambda_2(\mathcal{R}_{\mu,1}) \le \frac{1+\eta}{n}$. □

Since for the Glauber dynamics P^{GD}, we have already known for every function $f : \Omega \to \mathbb{R}$,

$$\mathcal{E}_{P^{\mathrm{GD}}}(f,f) = \mathbf{E}_{i\sim[n]}\left[\mathbf{E}_{\tau\sim\mu_{[n]\setminus\{i\}}}\left[\mathbf{Var}_{\mu^{\tau}}(f)\right]\right],$$

the spectral independence immediately implies the mixing rate of P^{GD}.

Theorem 3.5 (A Reformulation of Theorem 2.8)**.** *Let μ be a distribution over $\Omega \subseteq [q]^n$. Suppose that μ is η-spectrally independent. Then Glauber dynamics for the distribution μ has a spectral gap $\Omega\left(n^{-(1+\eta)}\right)$, and thus has the mixing time $O(n^{2+\eta})$.*

Proof. We only need to show the spectral gap of P^{GD}. Since μ is η-spectrally independent, by Lemma 3.4, for $1 \le \ell \le n$ and every $(\Lambda, \sigma_\Lambda) \in \mathcal{P}_\ell$, μ^{σ_Λ} is η-variance independent. Then it holds that

$$\begin{aligned}
\mathbf{Var}_{\mu}(f) &\le \left(1-\frac{1+\eta}{n}\right)^{-1}\mathbf{E}_{(i,s)\sim\mu_1}\left[\mathbf{Var}_{\mu^{i\leftarrow s}}(f)\right]\\
&\le \left(1-\frac{1+\eta}{n}\right)^{-1}\left(1-\frac{1+\eta}{n-1}\right)^{-1}\mathbf{E}_{(\{i,j\},\sigma_{\{i,j\}})\sim\mu_2}\left[\mathbf{Var}_{\mu^{\{i,j\}\leftarrow\sigma_{\{i,j\}}}}(f)\right]\\
&\le \cdots\\
&\le \prod_{j=0}^{\ell-1}\left(1-\frac{1+\eta}{n-j}\right)^{-1}\mathbf{E}_{(\Lambda,\sigma_\Lambda)\sim\mu_\ell}\left[\mathbf{Var}_{\mu^{\sigma_\Lambda}}(f)\right]\\
&\lesssim \exp\left((1+\eta)\sum_{j=0}^{\ell-1}\frac{1}{n-j}\right)\mathbf{E}_{(\Lambda,\sigma_\Lambda)\sim\mu_\ell}\left[\mathbf{Var}_{\mu^{\sigma_\Lambda}}(f)\right]\\
&\lesssim \left(\frac{n}{n-k}\right)^{1+\eta}\mathbf{E}_{(\Lambda,\sigma_\Lambda)\sim\mu_\ell}\left[\mathbf{Var}_{\mu^{\sigma_\Lambda}}(f)\right].
\end{aligned}$$

Let $\ell = n-1$, and we conclude the result. □

We remark here that, when $\eta > 1$, it holds that $1 - \frac{1+\eta}{n-\ell} < 0$ for $\ell = n-1$. To avoid this case, alternatively we define: for every $0 \le \ell \le n$,

$$\eta_\ell = \max_{(\Lambda,\sigma_\Lambda)\in\mathcal{P}_\ell}\lambda_{\max}\left(\Psi_\mu^{\sigma_\Lambda}\right).$$

Usually η_ℓ has the upper bound

$$\eta_\ell \le \min\{\eta, C(n-\ell)\}$$

where $0 < C < 1$. However, this is only a technical issue and is not the heart of most cases.

3.1.1 Optimal spectral gap for sparse graphical models

For graphical models with constant degree, it is well-known that the mixing time of a single-site Markov chain is at least $\Omega_\Delta(n\log n)$ in Hayes and Sinclair [HS05]. To achieve an optimal mixing rate, we show how to improve the result in Theorem 3.5.

Definition 3.6 (Graphic Markov Property)**.** For a distribution μ, we say it has the *graphic Markov property* if there exists a graph $G = (V(G), E(G))$ such that μ is a Markov distribution with respect to G, *i.e.*, for every partition

$$V(G) = A \sqcup \Lambda \sqcup B$$

such that A is isolated with B by Λ, it holds that for every pinning σ_Λ on Λ, the distribution $\mu^{\sigma_\Lambda}_{A\sqcup B}$ is the product probability measure as $\mu^{\sigma_\Lambda}_{A\sqcup B} = \mu^{\sigma_\Lambda}_{A} \otimes \mu^{\sigma_\Lambda}_{B}$.

Also the following shattering lemma is of great importance.

Lemma 3.7 (Shattering Lemma for Sparse Graph). *Let $G = (V(G), E(G))$ be an n-vertex graph of maximum degree Δ. Then for every positive integer $\ell > 0$,*

$$\mathbf{Pr}_S\left[|S_v| = \ell\right] \le (2e\Delta\theta)^{\ell-1}$$

where S is a uniformly random subset of $V(G)$ of size $\lceil \theta n \rceil$, and S_v is the unique maximal connected component of $G[S]$ containing v.

Theorem 3.8. *Let μ be a distribution over $\Omega \subseteq [q]^n$. Suppose that μ is η-spectrally independent and graphic Markov. Then the Glauber dynamics for μ has spectral gap $\mathrm{Gap}\left(P^{\mathrm{GD}}\right) \ge \Omega(1/Cn)$ for some constant $C = C(\Delta, \eta) > 0$.*

Proof. Let $\ell = (1-\theta)n$ for some parameter $0 \le \theta \le 1$. By Lemma 3.7, we have

$$\begin{aligned}
\mathbf{Var}_\mu(f) &\le \theta^{-(1+\eta)} \mathbf{E}_{S\sim\binom{n}{\ell}}\left[\mathbf{E}_{\tau\sim\mu_{V\setminus S}}\left[\mathbf{Var}_{\mu^\tau}(f)\right]\right] \\
&\le \theta^{-(1+\eta)} \mathbf{E}_{S\sim\binom{n}{\ell}}\left[\mathbf{E}_{\tau\sim\mu_{V\setminus S}}\left[\sum_{U \text{ is the maximal connected component in } S} \mathbf{Var}_{\mu_U^\tau}(f)\right]\right] \\
&\le \theta^{-(1+\eta)} \sum_{v\in V} \mathbf{E}_{\tau\sim\mu_{-v}}\left[\mathbf{Var}_{\mu_v^\tau}(f)\right] \mathbf{E}_{S\sim\binom{n}{\ell}}\left[C_{|S_v|}\right] \\
&\le \theta^{-(1+\eta)} n \mathbf{E}_{v\sim V}\left[\mathbf{E}_{\tau\sim\mu_{-v}}\left[\mathbf{Var}_{\mu_v^\tau}(f)\right]\right] \sum_{k=1}^{\infty} (2e\Delta\theta)^{k-1} C_\ell.
\end{aligned}$$

when $\theta \le O(1/\Delta)$, it holds that $\mathbf{Var}_\mu(f) \le C(\Delta,\eta)n \cdot \mathbf{E}_{v\sim V}\left[\mathbf{E}_{\tau\sim\mu_{-v}}\left[\mathbf{Var}_{\mu_v^\tau}(f)\right]\right]$, thus leading to the result. □

Remark 3.9. Theorem 3.8 shows us the mixing rate of the Glauber dynamics for a distribution with spectral independence is $\widetilde{O}_{\Delta,\eta}(n^2)$.

3.2 Entropy tensorization and optimal mixing rate

To show the optimal mixing of the Markov chain, we consider the standard/modified log-Sobolev inequality constant.

Definition 3.10 (Marginal Boundedness). For a distribution μ on $\Omega \subseteq [q]^n$ and a parameter $\beta \in (0, 1/2]$, we say μ is β-marginally bounded if for every $\Lambda \subseteq [n]$, every feasible pinning σ_Λ on Λ and $i \in [n] \setminus \Lambda$, it holds that for every feasible $(i, s) \in [q]$,

$$\mu_i^{\sigma_\Lambda}(s) \ge \beta.$$

Theorem 3.11. *Let μ be a distribution over $\Omega \subseteq [q]^n$. Suppose that μ is η-spectrally independent, graphic Markov and β-marginally bounded. Then the Glauber dynamics for μ has the modified log-Sobolev inequality constant*

$$\rho_{\mathrm{LS}}(P^{\mathrm{GD}}) \ge \Omega_{\Delta,\eta,\beta}(1/n)$$

thus leading to the mixing time $O_{\Delta,\eta,\beta}(n \log n)$.

Similarly to the definition of variance independence Definition 3.3, we introduce the following concept of entropic independence firstly in Anari et al. [AJK+22].

Definition 3.12 (Entropic Independence [AJK+22]). For a distribution μ over $\Omega \subseteq [q]^n$, we say μ is η-spectrally independent if for all function $f : \Omega \to \mathbb{R}_{>0}$,

$$\left(1 - \frac{1+\eta}{n}\right)\mathbf{Ent}_\mu[f] \le \mathbf{E}_{i\sim[n]}\left[\mathbf{E}_{s\sim\mu_i}\left[\mathbf{Ent}_{\mu^{i\leftarrow s}}[f]\right]\right].$$

Note that it is equivalent to the following inequality

$$\mathbf{Ent}_{\mu_1}\left[f^{(1)}\right] \le \frac{1+\eta}{n}\mathbf{Ent}_\mu[f].$$

Often we write the entropic independence as the following form:

$$\mathcal{D}_{\mathrm{KL}}\left(\nu_1 \parallel \mu_1\right) \le \frac{1+\eta}{n}\mathcal{D}_{\mathrm{KL}}\left(\nu \parallel \mu\right), \quad \forall \text{probability measure } \nu \text{ over } \Omega.$$

Similarly to Theorem 3.5, the entropic independence also implies a factorization of entropy.

Proposition 3.13. *Let μ be a distribution over $\Omega \subseteq [q]^n$ and fix an integer $1 \le \ell \le n$. Suppose that there exists $\eta \le O(1)$ such that for every $\Lambda \subseteq [n]$ with $|\Lambda| \le n - \ell - 1$ and every pinning $\sigma_\Lambda \in [q]^\Lambda$, the conditional probability distribution μ^{σ_Λ} is η-entropically independent. Then for every function $f : \Omega \to \mathbb{R}_{>0}$,*

$$\mathbf{Ent}_\mu\left[f\right] \le C_\ell \mathbf{E}_{S\sim\binom{[n]}{\ell}}\left[\mathbf{E}_{\tau\sim\mu_{[n]\setminus S}}\left[\mathbf{Ent}_{\mu^\tau}\left[f\right]\right]\right]$$

where $C_\ell \lesssim \left(\frac{n}{\ell}\right)^{1+\eta}$.

The proof of Proposition 3.13 is similar to the proof of Theorem 3.5 and just replace all $\mathbf{Var}.(\cdot)$ with $\mathbf{Ent}.[\cdot]$.

When $\ell = 1$, it is easy to see the result of Proposition 3.13 establish an approximate tensorization of entropy. As a corollary, we can obtain the bound of $\rho_{\mathrm{LS}}(P^{\mathrm{GD}})$.

Corollary 3.14. *Let μ be a distribution over $\Omega \subseteq [q]^n$. Suppose that there exists $\eta \le O(1)$ such that for every $\Lambda \subseteq [n]$ and every pinning $\sigma_\Lambda \in [q]^\Lambda$, the conditional probability distribution μ^{σ_Λ} is η-entropically independent. Then $\rho_{\mathrm{LS}}(P^{\mathrm{GD}}) \ge \Omega(1/C)$ where $C \le n^{1+\eta}$.*

Together with the graphic Markov property and marginal boundedness, the spectral independence implies the entropic independence for the conditional probability measures under an arbitrary pinning.

Lemma 3.15 (Spectral Independence Implies Entropic Independence [CE22])**.** *Let μ be a distribution over $\Omega \subseteq \{-1,+1\}^n$. Suppose that μ is graphic Markov, μ-spectrally independence and β-marginally bounded. Then μ and all its conditional distributions are $O(\eta/\beta^2)$-entropically independent.*

Assuming Proposition 3.13 and lemma 3.15, we can prove the full version of Theorem 3.11. See [CLV21] for detailed proof.

3.2.1 From spectral independence and marginal boundedness to entropic independence

The most important step from rapid mixing to optimal mixing is to establish entropic independence from spectral independence and marginal boundedness. We follow a way to establish a 'local-to-global' framework, and under this kind of framework, we compare the 'local entropies' with 'local variance' via marginal boundedness.

Fix a function $f : \Omega \to \mathbb{R}$. For every $0 \le \ell \le n$, recall the function $f^{(\ell)}$ as

$$f^{(\ell)}(\sigma_\Lambda) := \mathbf{E}_{\omega\sim\mu}\left[f(\omega) \mid \omega(\Lambda) = \sigma_\Lambda\right], \quad \forall(\Lambda, \sigma_\Lambda) \in \mathcal{P}_\ell.$$

Here we omit the symbol denoting the subset Λ since it can be known from the pinning σ_Λ. For every feasible partial pinning τ_S on $S \subseteq [n]$, we define:

$$f^{(\tau_S,\ell)}(\sigma_\Lambda) = f^{(\ell+|S|)}(\tau_S \sqcup \sigma_\Lambda), \quad \forall(\Lambda, \sigma_\Lambda) \in \mathcal{P}_\ell, \Lambda \cap S = \emptyset.$$

Accordingly we define the distribution $\mu_\ell^{\tau_S}$ as

$$\mu_\ell^{\tau_S}(\sigma_\Lambda) = \frac{1}{\binom{n-|S|}{\ell}}\mathbf{Pr}_{\omega\sim\mu}\left[\omega(\Lambda) = \sigma_\Lambda \mid \omega(S) = \tau_S\right], \quad \forall(\Lambda, \sigma_\Lambda) \in \mathcal{P}_\ell, \Lambda \cap S = \emptyset.$$

For the sake of simplicity, we use $\mathbf{Ent}_{\tau_S,\ell}\left[\cdot\right]$ to denote $\mathbf{Ent}_{\mu_\ell^{\tau_S}}\left[\cdot\right]$.

Definition 3.16 (Local-Entropy Contraction)**.** For $0 \le \alpha \le 1$, we say μ satisfies α-local entropy contraction if for every function $f : \Omega \to \mathbb{R}_{>0}$, it holds that

$$\mathbf{Ent}_1\left[f^{(1)}\right] \le \frac{1}{2}\left(1 - \frac{\alpha}{n}\right)^{-1}\mathbf{Ent}_2\left[f^{(2)}\right].$$

Proposition 3.17 ([CLV21]). *Let μ be a distribution over $\Omega \subseteq [q]^n$. Suppose that μ is graphically Markov, η-spectrally independent and β-marginally bounded. Then μ satisfies α-local entropy contraction with $\alpha = O(\eta/\beta^2)$.*

Theorem 3.18 (Local-to-Global Entropy Contraction Theorem [CLV21]). *Suppose that there exists $0 \le \alpha \le 1$ such that for every $0 \le k \le n-2$, every $S \subseteq [n]$ and every feasible pinning $\tau_S \in [q]^S$, the probability measure μ^{τ_S} satisfies*

$$\mathbf{Ent}_{\tau_S,1}\left[f^{(\tau_S,1)}\right] \le \frac{1}{2}\left(1-\frac{\alpha}{n-k}\right)^{-1}\mathbf{Ent}_{\tau_S,2}\left[f^{(\tau_S,2)}\right].$$

Then for every $0 \le k \le \ell \le n$ and every function $f : \Omega \to \mathbb{R}_{>0}$,

$$\frac{\mathbf{Ent}_k\left[f^{(k)}\right]}{\beta_k} \le \frac{\mathbf{Ent}_\ell\left[f^{(\ell)}\right]}{\beta_\ell}$$

where $\beta_i := \sum_{j=0}^{i-1}\prod_{t=0}^{j-1}\left(1-\frac{2\alpha}{n-t}\right)$.

Proof. Note that it suffices to prove the case $\ell = k+1$. We prove it by induction. When $k = 0$, it is just the definition of the local-entropy contraction. Otherwise, by Lemma 3.1,

$$\begin{aligned}
\mathbf{Ent}_{k+1}\left[f^{(k+1)}\right] - \mathbf{Ent}_{k-1}\left[f^{(k-1)}\right] &= \mathbf{E}_{\sigma\sim\mu_{k-1}}\left[\mathbf{Ent}_{\sigma,2}\left[f^{(\sigma,2)}\right]\right] \\
&\ge 2\left(1-\frac{\alpha}{n-k+1}\right)\mathbf{E}_{\sigma\sim\mu_{k-1}}\left[\mathbf{Ent}_{\sigma,1}\left[f^{(\sigma,1)}\right]\right] \\
&= 2\left(1-\frac{\alpha}{n-k+1}\right)\left(\mathbf{Ent}_k\left[f^{(k)}\right] - \mathbf{Ent}_{k-1}\left[f^{(k-1)}\right]\right).
\end{aligned}$$

By induction hypothesis, $\frac{1}{\beta_{k-1}}\mathbf{Ent}_{k-1}\left[f^{(k-1)}\right] \le \frac{1}{\beta_k}\mathbf{Ent}_k\left[f^{(k)}\right]$. Then we obtain

$$\begin{aligned}
\mathbf{Ent}_{k+1}\left[f^{(k+1)}\right] &\ge 2\left(1-\frac{\alpha}{n-k+1}\right)\mathbf{Ent}_k\left[f^{(k)}\right] - \left(1-\frac{2\alpha}{n-k+1}\right)\mathbf{Ent}_{k-1}\left[f^{(k-1)}\right] \\
&\ge \left(2\left(1-\frac{\alpha}{n-k+1}\right) - \frac{\beta_k}{\beta_{k-1}}\left(1-\frac{2\alpha}{n-k+1}\right)\right)\mathbf{Ent}_k\left[f^{(k)}\right] \\
&= \left(1-\left(1-\frac{2\alpha}{n-k+1}\right)\left(\frac{\beta_k}{\beta_{k-1}}-1\right)\right)\mathbf{Ent}_k\left[f^{(k)}\right] \\
&= \frac{\beta_{k+1}}{\beta_k}\mathbf{Ent}_k\left[f^{(k)}\right].
\end{aligned}$$

□

Proof of Lemma 3.15. By Proposition 3.17, it holds that μ satisfies α-local entropy contraction with $\alpha = O(\eta/\beta^2)$. Then, by Theorem 3.18 with $k = 1$ and $\ell = n$, for all function $f : \Omega \to \mathbb{R}_{>0}$,

$$\mathbf{Ent}_1\left[f^{(1)}\right] \le \frac{1}{\beta_n}\mathbf{Ent}_\mu[f]$$

where

$$\beta_n = \sum_{j=0}^{n-1}\prod_{i=0}^{j-1}\left(1-\frac{2\alpha}{n-i}\right) \gtrsim \sum_{j=0}^{n-1}\left(\frac{n}{n-j}\right)^{-O(\eta/\beta^2)} \gtrsim \theta(1-\theta)^{O(\eta/\beta^2)}n, \quad \forall\theta\in(0,1).$$

It is easy to show the optimal choice of θ is $O(\beta^2/\eta)$. □

3.2.2 Optimal mixing rate without marginal boundedness

Recall that in Theorem 3.11, the marginal boundedness property seems to be necessary. However, in most graphic models, the marginal bound might be bad than what we expect. In this section, we will show how to derive an optimal mixing rate without the assumption marginal boundedness, but with a stricter assumption on spectral independence.

Definition 3.19 (Exponential Tilt)**.** For a vector $\theta \in \mathbb{R}^n$ and a distribution μ over $\{-1,+1\}^n$, define the exponential tilt $\mathcal{T}_\theta\mu$ as the distribution over $\{-1,+1\}^n$ by:

$$\mathcal{T}_\theta\mu(x) = \frac{\mu(x)\exp(\langle\theta,x\rangle)}{\sum_{y\in\{-1,+1\}^n}\mu(y)\exp(\langle\theta,y\rangle)}, \quad \forall x \in \{-1,+1\}^n .$$

Remark 3.20. The exponential tilt is introduced by the principle of maximum entropy. Consider the following optimization problem:

$$\inf_\nu \mathcal{D}_{\mathrm{KL}}(\nu \,\|\, \mu) \;\; s.t.\; \mathbf{E}_{x\sim\nu}[\varphi(x)] = m.$$

The principle of maximum entropy says the optimal distribution will be of the form

$$\mu_\theta(x) \propto \mu(x)\exp(\langle\theta,\varphi(x)\rangle)$$

for some vector $\theta \in \mathbb{R}^n$.

Theorem 3.21 ([AJK+22, CE22])**.** *Let μ be a distribution over $\{-1,+1\}^n$ and fix a parameter η. The followings are equivalent.*

- *For every $\theta \in \mathbb{R}^n$, the tilted distribution $\mathcal{T}_\theta\mu$ is η-spectrally independent.*
- *For every $\theta \in \mathbb{R}^n$, the tilted distribution $\mathcal{T}_\theta\mu$ is η-entropically independent.*

To prove Theorem 3.21, we introduce the logarithmic Laplace transform of μ. Given a distribution μ over $\{-1,+1\}^n$, define the function $\mathcal{L}_\mu : \mathbb{R}^n \to \mathbb{R}$ as

$$\mathcal{L}_\mu(\theta) := \log \mathbf{E}_{x\sim\mu}[\exp(\langle\theta,x\rangle)], \quad \forall\theta \in \mathbb{R}^n.$$

The following properties of $\mathcal{L}_\mu(\cdot)$ are proved by Bubeck and Eldan [BE19].

Proposition 3.22 ([BE19])**.** *Let μ be a distribution over $\mathbb{R}^n$. Then the followings hold:*

- *$\mathcal{L}_\mu(\cdot)$ is smooth and strictly convex.*
- *It holds that*

$$\nabla\mathcal{L}_\mu(\theta) = \mathbf{E}_{x\sim\mathcal{T}_\theta\mu}[x], \quad \nabla^2\mathcal{L}_\mu(\theta) = \mathbf{Cov}(\mathcal{T}_\theta\mu).$$

- *Its convex conjugate $\mathcal{L}^*_\mu(x) := \sup_\theta\left\{\langle x,\theta\rangle - \mathcal{L}_\mu(\theta)\right\}$ has the form*

$$\mathcal{L}^*_\mu(x) = \mathcal{D}_{\mathrm{KL}}\left(\mathcal{T}_{\theta^*(x)}\mu \,\middle\|\, \mu\right)$$

  where $\theta^*(x) = \nabla\mathcal{L}^*_\mu(x)$ is the optimizer of the convex conjugate. Furthermore, if all mappings are invertible, then

$$\nabla^2\mathcal{L}^*_\mu(x) = \mathbf{Cov}\left(\mathcal{T}_{\theta^*(x)}\mu\right)^{-1}.$$

Proof of Theorem 3.21. For convenience we use $\mathbf{b}(\cdot)$ to denote the mean of a distribution. Firstly observe that

$$\mu_1((i,+1)) = \frac{1}{n}\frac{1+\mathbf{b}(\mu)_i}{2}, \quad \mu_1((i,-1))\frac{1}{n}\frac{1-\mathbf{b}(\mu)_i}{2}.$$

Then it follows that

$$\mathcal{D}_{\mathrm{KL}}(\nu_1 \,\|\, \mu_1) = \frac{1}{n}\sum_{i=1}^n\left(\frac{1+\mathbf{b}(\nu)_i}{2}\log\frac{1+\mathbf{b}(\nu)_i}{1+\mathbf{b}(\mu)_i} + \frac{1-\mathbf{b}(\nu)_i}{2}\log\frac{1-\mathbf{b}(\nu)_i}{1-\mathbf{b}(\mu)_i}\right).$$

Define the function $H : [-1,+1]^n \times [-1,+1]^n \to \mathbb{R}$ as

$$H(x,y) := \frac{1}{n}\sum_{i=1}^n\left(\frac{1+x_i}{2}\log\frac{1+x_i}{1+y_i} + \frac{1-x_i}{2}\log\frac{1-x_i}{1-y_i}\right), \forall x,y \in [-1,+1]^n$$

Then it suffices to show for all distribution ν over $\{-1,+1\}^n$,

$$H(\mathbf{b}(\nu), \mathbf{b}(\mu)) \leq (1+\eta)\mathcal{D}_{\mathrm{KL}}\left(\nu \parallel \mu\right).$$

Then by the principle of maximum entropy, we only need to consider $\nu = \mathcal{T}_\theta \mu$. By Proposition 3.22, it is equivalent to

$$F_\mu(x) := (1+\eta)\mathcal{L}_\mu^*(x) - H(x, \mathbf{b}(\mu)) \geq 0, \quad \forall x \in \mathbb{R}^n.$$

Observe that $F_\mu(\mathbf{b}(\mu)) = 0$, and by calculation, $\nabla F_\mu(\mathbf{b}(\mu)) = 0$. On the other hand, note that

$$\nabla F_\mu(x) = (1+\eta)\mathbf{Cov}\left(\mathcal{T}_{\theta^*(x)}\mu\right)^{-1} - \mathrm{diag}\left(1-x^2\right)^{-1}.$$

Then $\nabla F_\mu(x) \preceq \mathbf{0}$ comes directly from the spectral independence of all exponential tilt distributions.

Conversely, assume that the entropic independence holds for all exponential tilt distributions. Observe that

$$F_\mu(x) - F_\mu(\mathbf{b}(\mu)) - \left\langle \nabla F_\mu(\mathbf{b}(\mu)), x - \mathbf{b}(\mu)\right\rangle = F_{\mathcal{T}_{\theta^*(x)}\mu}(x) \geq 0.$$

Then we know $F_\mu(x)$ is globally convex, meaning that the spectral independence holds for all tilts. □

References

[AJK+22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic Independence: Optimal Mixing of Down-Up Random Walks. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 1418–1430, New York, NY, USA, 2022. Association for Computing Machinery. 14, 17

[AL20] Vedat Levi Alev and Lap Chi Lau. Improved Analysis of Higher Order Random Walks and Applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, page 1198–1211, New York, NY, USA, 2020. Association for Computing Machinery. 9

[ALOG20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS2020, Durham, NC, USA, November 16-19, 2020*, pages 1319–1330. IEEE, 2020. 2, 9

[ALOG21] Dorna Abdolazimi, Kuikui Liu, and Shayan Oveis Gharan. A Matrix Trickle-Down Theorem on Simplicial Complexes and Applications to Sampling Colorings. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 161–172, 2021. 7, 8

[BE19] Sébastien Bubeck and Ronen Eldan. The Entropic Barrier: Exponential Families, Log-Concave Geometry, and Self-Concordance. *Mathematics of Operations Research*, 44(1):264–276, 2019. 17

[BT06] Sergey G. Bobkov and Prasad Tetali. Modified Logarithmic Sobolev Inequalities in Discrete Settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006. 2

[CE22] Yuansi Chen and Ronen Eldan. Localization Schemes: A Framework for Proving Mixing Bounds for Markov Chains (extended abstract). In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 110–122, Los Alamitos, CA, USA, Nov. 2022. IEEE Computer Society. 15, 17

[CGM21] Mary Cryan, Heng Guo, and Giorgos Mousa. Modified Log-Sobolev Inequalities for Strongly Log-Concave Distributions. *The Annals of Probability*, 49(1):506–525, 2021. 10

[CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal Mixing of Glauber Dynamics: Entropy Factorization via High-Dimensional Expansion. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 1537–1550, New York, NY, USA, 2021. Association for Computing Machinery. 15, 16

[FGYZ22] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the boolean domain. *ACM Trans. Algorithms*, 18(3), oct 2022. 3

[HS05] Thomas P. Hayes and Alistair Sinclair. A General Lower Bound for Mixing of Single-Site Dynamics on Graphs. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 511–520, 2005. 13

[LP17] David Asher Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society Providence, Rhode Island, Providence, Rhode Island, second edition edition, 2017. 2

[Opp18] Izhar Oppenheim. Local Spectral Expansion Approach to High Dimensional Expanders Part I: Descent of Spectral Gaps. *Discrete & Computational Geometry*, 59(2):293–330, 2018. 5, 7