Lecture 14: Entropic Independence and Optimal Mixing of Glauber Dynamics

October 31, 2023

In the previous lecture, we proved that $O(1)$-spectral independence implies an inverse polynomial spectral gap lower bound for Glauber dynamics, which we then improved to the optimal $\Omega(1/n)$ lower bound when our distribution is globally Markov w.r.t. a bounded-degree graph. In this lecture, we analyze contraction of KL-divergence using similar tools. In particular, we establish optimal $\Omega(1/n)$ lower bounds on the modified and standard log-Sobolev constants, as well as the optimal $O(n \log n)$-mixing of Glauber dynamics. The main tool in our discussion will be an analog of spectral independence for entropy known as *entropic independence* [Ana+22].

One of the main theorems of this lecture is the following, which we copied (and abridged) from the previous lecture.

Theorem 0.1 (Optimal Mixing for Sparse Graphical Models; [CLV21]). *Let μ be a probability measure on $\{\pm 1\}^n$. Suppose μ satisfies the following properties:*

(i) ***Spectral Independence:*** *There exists $\eta \leq O(1)$ such that for every $S \subseteq [n]$ and every pinning $\tau : S \to \{\pm 1\}$, the conditional distribution μ^τ is η-spectrally independent.*

(ii) ***Conditional Independence:*** *There exists a graph $G = (V, E)$ on vertex set $V = [n]$ of maximum degree Δ such that μ is* globally Markov w.r.t. G.

(iii) ***$\mathscr{B}$-Marginal Boundedness:*** *Suppose there exists a constant $0 < \mathscr{B} \leq 1/2$ such that for every $S \subseteq [n]$, every pinning $\tau : S \to \{\pm 1\}$, and every $i \in [n] \setminus S$, we have $\mu_i^\tau(+1), \mu_i^\tau(-1) \geq \mathscr{B}$.*

Then $\gamma(\mathsf{P}_{\mathsf{GD}}), \varrho(\mathsf{P}_{\mathsf{GD}}), \kappa(\mathsf{P}_{\mathsf{GD}}) \geq \Omega_{\eta,\Delta,\mathscr{B}}(1/n)$ and P_{GD} mixes in $O_{\eta,\Delta,\mathscr{B}}(n \log n)$-steps.

1 Entropic Independence

In the previous lecture, the key step to bounding the Poincaré constant was showing that spectral independence implies the following variance inequality

$$\left(1 - \frac{1+\eta}{n}\right) \cdot \mathrm{Var}_\mu(f) \leq \mathbb{E}_{i \sim [n]}\left[\mathbb{E}_{s \sim \mu_i}\left[\mathrm{Var}_{\mu^{i \leftarrow s}}(f)\right]\right] \qquad \forall f : \{\pm 1\}^n \to \mathbb{R}, \tag{1}$$

or equivalently,

$$\mathrm{Var}_{\mu_1}(f_1) \leq \frac{1+\eta}{n} \cdot \mathrm{Var}_\mu(f) \qquad \forall f : \{\pm 1\}^n \to \mathbb{R}. \tag{2}$$

Here, recall that μ_1, f_1 are supported on $[n] \times \{\pm 1\}$, and defined by marginalization: $\mu_1(i, s) = \frac{1}{n}\Pr_\mu[i \leftarrow s]$ and $f_1(i, s) = \mathbb{E}_{\sigma \sim \mu^{i \leftarrow s}}[f]$ for all $(i, s) \in [n] \times \{\pm 1\}$. These inequalities allowed us to inductively factorize the variance until we obtained $\mathbb{E}_{i \sim [n]}\left[\mathbb{E}_{\tau \sim \mu_{-i}}\left[\mathrm{Var}_{\mu^\tau}(f)\right]\right] = \mathcal{E}_{\mathsf{GD}}(f, f)$, the Dirichlet form appearing in the Poincaré Inequality. The entropic analog of these inequalities is what we need to factorize the entropy and establish bounds on the modified and standard log-Sobolev constants.

Definition 1 (Entropic Independence; [Ana+22]). *We say μ is η-*entropically independent *if*

$$\left(1 - \frac{1+\eta}{n}\right) \cdot \mathrm{Ent}_\mu(f) \leq \mathbb{E}_{i \sim [n]}\left[\mathbb{E}_{s \sim \mu_i}\left[\mathrm{Ent}_{\mu^{i \leftarrow s}}(f)\right]\right] \qquad \forall f : \{\pm 1\}^n \to \mathbb{R}, \tag{3}$$

or equivalently,

$$\operatorname{Ent}_{\mu_1}(f_1) \le \frac{1+\eta}{n} \cdot \operatorname{Ent}_{\mu}(f), \qquad \forall f : \{\pm 1\}^n \to \mathbb{R}_{\ge 0}, \tag{4}$$

In the literature, these inequalities also go under the name *(approximate) subadditivity of entropy* or *(approximate) Shearer Inequalities*; see e.g. [Bar+11; Bla+22]. They are more often written as

$$\mathscr{D}_{\mathrm{KL}}(\nu_1 \,\|\, \mu_1) \le \frac{1+\eta}{n} \cdot \mathscr{D}_{\mathrm{KL}}(\nu \,\|\, \mu), \qquad \forall \nu.$$

Proposition 1.1 (Entropic Independence $\implies$ Entropy Factorization (Informal); [Ana+22])**.** *Let μ be a probability measure on $\{\pm 1\}^n$, and fix $1 \le k \le n$. Suppose there exists $\eta \le O(1)$ such that for every $S \subseteq [n]$ with $|S| \le n-k-1$ and every pinning $\tau : S \to \{\pm 1\}$, the conditional measure μ^τ is η-entropically independent. Then*

$$\operatorname{Ent}_{\mu}(f) \le C_k \cdot \mathbb{E}_{S \sim \binom{[n]}{k}}\left[\mathbb{E}_{\tau \sim \mu_{V \setminus S}}\left[\operatorname{Ent}_{\mu^\tau}(f)\right]\right], \qquad \forall f : \{\pm 1\}^n \to \mathbb{R}_{\ge 0}, \tag{5}$$

with $C_k \lesssim \left(\frac{n}{k}\right)^{1+\eta}$.

Proof Sketch. Follow the proof of the first "spectral independence implies fast mixing" theorem in the previous lecture, using Eq. (3) and replacing every occurrence of $\operatorname{Var}(\cdot)$ with $\operatorname{Ent}(\cdot)$. □

Remark 1. In the case $k = 1$, Eq. (5) is often referred to as *approximate tensorization of entropy.*

Lemma 1.2 (Consequences of Entropy Factorization; see e.g. [CMT15])**.** *Let μ be a probability measure on on $\{\pm 1\}^n$. If μ satisfies Eq. (5) with $k = 1$, then Glauber dynamics w.r.t. μ has modified log-Sobolev constant $\varrho(\mathsf{P}_{\mathsf{GD}}) \ge \frac{1}{C_1}$. If in addition μ is $\mathscr{B}$-marginally bounded in the sense of Item (iii), then Glauber dynamics w.r.t. μ has standard log-Sobolev constant $\kappa(\mathsf{P}_{\mathsf{GD}}) \ge \frac{1}{C_{\mathscr{B}} \cdot C_1}$, where $C_{\mathscr{B}} = \frac{1 - 2\cdot\mathscr{B}}{\log\left(\frac{1}{\mathscr{B}} - 1\right)}$ is some constant depending on $\mathscr{B}$.*

1.1 Optimal Entropy Factorization for Glauber Dynamics

The last ingredient to prove Theorem 0.1 is to show that spectral independence (along with marginal boundedness) implies entropic independence. This is proved in the next section. Once we have entropic independence, all of the same entropy factorization arguments go through.

Theorem 1.3 ([CE22])**.** *Suppose μ is a distribution on $\{\pm 1\}^n$ satisfying the spectral independence and marginal boundedness assumptions Items (i) and (iii) from Theorem 0.1. Then μ and all of its conditional distributions are $O(\eta/\mathscr{B}^2)$-entropically independent.*

Proof Sketch of Theorem 0.1. From our assumptions combined with Theorem 1.3, we know that μ and its conditionals are $O(\eta/\mathscr{B}^2)$-entropically independent. We then apply Proposition 1.1 with $k = (1-\theta)n$ for a constant $0 \le \theta \le 1$, and then further factorize the entropy using the shattering property for sparse graphs for the remaining levels. This is exactly as we did in our proof of the optimal Poincaré Inequality. One can literally follow every step of the proof in the previous lecture and replace every occurrence of $\operatorname{Var}(\cdot)$ with $\operatorname{Ent}(\cdot)$. One should be slightly careful in the final factorization step, where we have $\operatorname{Ent}_{\mu_U^\tau}(f)$ for a pinning $\tau : V \setminus S \to \{\pm 1\}$ and U is a connected component of $G[S]$ (which is typically small by the Shattering Lemma); for a full implementation,[1] see [CLV21]. Ultimately, we obtain Eq. (5) with $C_1 \le O_{\eta,\Delta,\mathscr{B}}(n)$. Applying Lemma 1.2 finishes the proof. □

[1] One way is to bound the standard log-Sobolev constant of Glauber dynamics w.r.t. μ_U^τ, which turns out implies entropy factorization. We can apply off-the-shelf comparison inequalities between the standard log-Sobolev constant and the Poincaré constant, making use of marginal boundedness, and use spectral independence to bound the Poincaré constant. Notably, we can afford all sorts of exponential losses in these arguments because the probability that $|U| = \ell$ is exponentially decaying in ℓ, with a rate we can make as fast as we want by tuning θ.

2 Spectral Independence + Marginal Boundedness $\implies$ Entropic Independence: Proof of Theorem 1.3

We break the proof of Theorem 1.3 into two steps: We first reduce entropic independence to a more "local" entropy inequality which only involves "sufficiently averaged" functions like f_1. This averaging will allow us to compare "local entropies" with "local variances" by leveraging marginal boundedness. In particular, in our second step, we reduce this "local" entropy inequality to the analogous "local" variance inequality, which is guaranteed by spectral independence by the arguments provided in the previous lecture.

Let us now set up the required notation. Fix a function $f : \{\pm 1\}^n \to \mathbb{R}$. For $0 \le k \le n$, define the function f_k on pinnings of k coordinates by

$$f_k(\tau) \stackrel{\mathsf{def}}{=} \mathbb{E}_{\sigma\sim\mu}\left[f(\sigma) \mid \sigma_S = \tau\right], \qquad \forall \tau : S \to \{\pm 1\} \text{ s.t. } |S| = k,$$

and the distribution μ_k on pinnings by

$$\mu_k(\tau) \stackrel{\mathsf{def}}{=} \frac{1}{\binom{n}{k}} \Pr_{\sigma\sim\mu}[\sigma_S = \tau].$$

Note that $\mu_n = \mu$, $f_n = f$, and that if $f = \frac{d\nu}{d\mu}$ is the density of some other probability measure on $\{\pm 1\}^n$, then $f_k = \frac{d\nu_k}{d\mu_k}$ for all $0 \le k \le n$. Note that $\mathbb{E}_{\mu_k}[f_k] = \mathbb{E}_\mu[f]$ for all $0 \le k \le n$.

Remark 2. A dynamical way to view these distributions μ_k and functions f_k is to think of them as "projections" from the global distribution μ on $\{\pm 1\}^n$ and the global function $f : \{\pm 1\}^n \to \mathbb{R}$ down to partial configurations on k coordinates. In particular, if we let $\mathcal{D}^{n\searrow k}$ denote the Markov kernel which acts on distributions ν by sampling $\sigma \sim \nu$ and outputting the restriction σ_S for a uniformly random k-subset $S \sim \binom{[n]}{k}$, then $\mu_k = \mu \mathcal{D}^{n\searrow k}$. Similarly, we can let $\mathcal{U}^{k\nearrow n}$ denote the "dual" action w.r.t. μ, where given a distribution ν_k on pinnings on k coordinates, we first sample a random such pinning $\tau \sim \nu_k$, and then output a random complete configuration $\sigma \sim \mu$ conditioned on σ agreeing with τ. Then $\mu = \mu_k \mathcal{U}^{k\nearrow n}$ and $f_k = \mathcal{U}^{k\nearrow n} f$.

We can combine this notation with conditioning. For a pinning $\tau : S \to \{\pm 1\}$ and $0 \le k \le n - |S|$, we can define

$$f_k^\tau(\tau') = f_{k+|S|}(\tau \sqcup \tau'), \qquad \forall \tau' : T \to \{\pm 1\} \text{ s.t. } |T| = k, T \cap S = \emptyset,$$

and

$$\mu_k^\tau(\tau') = \frac{1}{\binom{n-|S|}{k}} \Pr_{\sigma\sim\mu}\left[\sigma_T = \tau' \mid \sigma_S = \tau\right], \qquad \forall \tau' : T \to \{\pm 1\} \text{ s.t. } |T| = k, T \cap S = \emptyset.$$

To reduce cumbersome notation, we write $\mathrm{Ent}_k^\tau(\cdot)$ instead of $\mathrm{Ent}_{\mu_k^\tau}(\cdot)$.

Definition 2 (Local Entropy Contraction)**.** *For $0 \le \alpha \le 1$, we say μ satisfies* α-local entropy contraction *if for every global function $f : \{\pm 1\}^n \to \mathbb{R}_{\ge 0}$, the induced projections f_1, f_2 satisfy*

$$\mathrm{Ent}_1(f_1) \le \frac{1}{2}\left(1 - \frac{\alpha}{n}\right)^{-1} \cdot \mathrm{Ent}_2(f_2).$$

With these notations in hand, we can formalize our two steps.

Theorem 2.1 ("Local-to-Global" Entropy Contraction; [CLV21])**.** *Suppose there exist constants $0 \le \alpha_0, \dots, \alpha_{n-2} \le 1$ such that for every $0 \le k \le n-2$, the following holds: For every pinning $\tau : S \to \{\pm 1\}$ s.t. $|S| = k$, the conditional measure μ^τ satisfies α-local entropy contraction, i.e.*

$$\mathrm{Ent}_1^\tau(f_1^\tau) \le \frac{1}{2}\left(1 - \frac{\alpha}{n-k}\right)^{-1} \cdot \mathrm{Ent}_2^\tau(f_2^\tau), \qquad \forall f : \{\pm 1\}^n \to \mathbb{R}_{\ge 0}.$$

Then for every $0 \le k \le \ell \le n$ and every global function $f : \mathrm{supp}(\mu) \to \mathbb{R}_{\ge 0}$,

$$\frac{\mathrm{Ent}_k(f_k)}{\beta_k} \le \frac{\mathrm{Ent}_\ell(f_\ell)}{\beta_\ell}, \qquad \textit{where} \qquad \beta_k \stackrel{\mathsf{def}}{=} \sum_{j=0}^{k-1} \prod_{i=0}^{j-1} \left(1 - \frac{2\alpha}{n-i}\right). \tag{6}$$

Furthermore, analogous bounds hold for every conditional measure μ^τ.

Proposition 2.2 ([CLV21]). *Suppose μ is a distribution on $\{\pm 1\}^n$ satisfying the spectral independence and marginal boundedness assumptions Items (i) and (iii) from Theorem 0.1. Then μ satisfies the α-local entropy contraction assumption of Theorem 2.1 with $\alpha = O(\eta/\mathscr{B}^2)$.*

Proof of Theorem 1.3. By Proposition 2.2, we have α-local entropy contraction with $\alpha = O(\eta/\mathscr{B}^2)$. Hence, applying Theorem 2.1 with $k = 1, \ell = n$, we obtain

$$\mathrm{Ent}_1(f_1) \leq \frac{1}{\beta_n} \cdot \mathrm{Ent}_\mu(f), \qquad \forall f : \{\pm 1\}^n \to \mathbb{R}_{\geq 0},$$

where

$$\beta_n = \sum_{j=0}^{n-1} \prod_{i=0}^{j-1} \left(1 - \frac{O(\eta/\mathscr{B}^2)}{n-i}\right) \gtrsim \sum_{j=0}^{n-1} \left(\frac{n}{n-j}\right)^{-O(\eta/\mathscr{B}^2)}.$$

We are done if we can show that $\beta_n \geq \Omega(\mathscr{B}^2 n/\eta)$. If we truncate the sum to $j = 0, \ldots, \theta n$ for some constant $0 \leq \theta \leq 1$, then we get a lower bound of

$$\beta_n \gtrsim \theta(1-\theta)^{O(\eta/\mathscr{B}^2)} \cdot n.$$

A straightforward calculus exercise reveals the optimal choice of θ is $O(\mathscr{B}^2/\eta)$, and so we're done. □

While we only applied Theorem 2.1 with $k = 1, \ell = n$, it will be useful in the proof to inductively establish the general case $0 \leq k \leq \ell \leq n$. Also note that the case of $\ell = n$ and k arbitrary, Theorem 2.1 gives an analog of Proposition 1.1 with the same conclusion but replacing the assumption of entropic independence with local entropy contraction in the sense of Definition 2. Even though local entropy contraction implies entropic independence by Theorem 2.1, [Ana+22] constructs simple examples where entropic independence holds but local entropy contraction *fails.*

2.1 Proof of Local-to-Global Entropy Contraction

The main tool we need is again the following lemma on entropy decomposition, which can be proved via direct calculation.

Lemma 2.3 (Law of Total Entropy). *Fix $f : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$. Then for every $0 \leq k \leq \ell \leq n$,*

$$\mathrm{Ent}_\ell(f_\ell) = \mathrm{Ent}_k(f_k) + \mathbb{E}_{\sigma \sim \mu_k}\left[\mathrm{Ent}_{\ell-k}^\sigma\left(f_{\ell-k}^\sigma\right)\right].$$

Proof of Theorem 2.1. It suffices to prove the special case $\ell = k+1$, since the general case Eq. (6) follows by chaining these together. We go by induction. The base case $k = 1$ follows immediately by the definition of local entropy contraction. By Lemma 2.3,

$$\begin{aligned} \mathrm{Ent}_{k+1}(f_{k+1}) - \mathrm{Ent}_{k-1}(f_{k-1}) &= \mathbb{E}_{\sigma \sim \mu_{k-1}}\left[\mathrm{Ent}_2^\sigma(f_2^\sigma)\right] \\ \mathrm{Ent}_k(f_k) - \mathrm{Ent}_{k-1}(f_{k-1}) &= \mathbb{E}_{\sigma \sim \mu_{k-1}}\left[\mathrm{Ent}_1^\sigma(f_1^\sigma)\right]. \end{aligned}$$

Local entropy contraction allows us to compare the right-hand sides of these two identities, and induction allows us to control the second difference $\mathrm{Ent}_k(f_k) - \mathrm{Ent}_{k-1}(f_{k-1})$. This suggests how we should implement the inductive step. In particular,

$$\begin{aligned} \mathrm{Ent}_{k+1}(f_{k+1}) - \mathrm{Ent}_{k-1}(f_{k-1}) &= \mathbb{E}_{\sigma \sim \mu_{k-1}}\left[\mathrm{Ent}_2^\sigma(f_2^\sigma)\right] && \text{(Lemma 2.3)} \\ &\geq 2\left(1 - \frac{\alpha}{n-k+1}\right) \cdot \mathbb{E}_{\sigma \sim \mu_{k-1}}\left[\mathrm{Ent}_1^\sigma(f_1^\sigma)\right] && \text{(Definition 2)} \\ &= 2\left(1 - \frac{\alpha}{n-k+1}\right) \cdot \left(\mathrm{Ent}_k(f_k) - \mathrm{Ent}_{k-1}(f_{k-1})\right) && \text{(Lemma 2.3)} \end{aligned}$$

Since $\frac{1}{\beta_{k-1}} \operatorname{Ent}_{k-1}(f_{k-1}) \leq \frac{1}{\beta_k} \operatorname{Ent}_k(f_k)$ by the inductive hypothesis,

$$\begin{aligned}
\operatorname{Ent}_{k+1}(f_{k+1}) &\geq 2\left(1 - \frac{\alpha}{n-k+1}\right) \cdot \operatorname{Ent}_k(f_k) - \left(1 - \frac{2\alpha}{n-k+1}\right) \cdot \operatorname{Ent}_{k-1}(f_{k-1}) \\
&\geq \left(2\left(1 - \frac{\alpha}{n-k+1}\right) - \left(1 - \frac{2\alpha}{n-k+1}\right) \cdot \frac{\beta_{k-1}}{\beta_k}\right) \cdot \operatorname{Ent}_k(f_k) \\
&\hfill \text{(Inductive Hypothesis)} \\
&= \left(1 - \left(1 - \frac{2\alpha}{n-k+1}\right) \cdot \left(\frac{\beta_{k-1}}{\beta_k} - 1\right)\right) \cdot \operatorname{Ent}_k(f_k) \\
&= \frac{\beta_{k+1}}{\beta_k} \cdot \operatorname{Ent}_k(f_k).
\end{aligned}$$

This completes the induction and the proof. □

2.2 Proof Sketch of Proposition 2.2

The key lemmas are the following. The first relates the deficit in local entropy contraction to the variance of f_1 w.r.t. μ_1. The second shows that when the distribution is marginally bounded, this variance $\operatorname{Var}_1(f_1)$ can be meaningfully compared with $\operatorname{Ent}_1(f_1)$ again up to constant factors. Combining them in a straightforward manner immediately implies Proposition 2.2.

Lemma 2.4 ([CGM21]). *Let μ be η-spectrally independent. Then*

$$\operatorname{Ent}_2(f_2) - 2 \cdot \operatorname{Ent}_1(f_1) \geq -\frac{\eta}{n-1} \cdot \frac{\operatorname{Var}_1(f_1)}{\mathbb{E}_{\mu_1}[f_1]}, \qquad \forall f : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}.$$

Lemma 2.5. *Suppose μ is $\mathscr{B}$-marginally bounded. Then for every global function $f : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$, the induced local function $f_1 : [n] \times \{\pm 1\} \to \mathbb{R}_{\geq 0}$ is "balanced" in the sense that*

$$f_1(i,s) \leq \frac{1}{\mathscr{B}} \cdot \mathbb{E}_{\mu_1}[f_1], \qquad \forall (i,s) \in [n] \times \{\pm 1\}. \tag{7}$$

For such functions,

$$\operatorname{Ent}_1(f_1) \leq \frac{\operatorname{Var}_1(f_1)}{\mathbb{E}_{\mu_1}[f_1]} \leq \frac{4}{\mathscr{B}^2} \cdot \operatorname{Ent}_1(f_1).$$

Proof Sketch of Lemma 2.5. We prove Eq. (7). The rough intuition behind the second conclusion is that $x \mapsto x \log x$ behaves quadratically in a neighborhood of 1, which is guaranteed by Eq. (7); we refer interested readers to [CLV21] for the rigorous proof of this comparison between variance and entropy. To justify Eq. (7), first note that we may assume by a global rescaling of f that $\mathbb{E}_\mu[f] = 1$, i.e. $f = \frac{d\nu}{d\mu}$ for some other distribution ν on $\{\pm 1\}^n$. Then $f_1 = \frac{d\nu_1}{d\mu_1}$ and $\mathbb{E}_{\mu_1}[f_1] = 1$. From here, we immediately have that for all $(i,s) \in [n] \times \{\pm 1\}$,

$$f_1(i,s) = \frac{\nu_1(i,s)/n}{\mu_1(i,s)/n} = \frac{\nu_i(s)}{\mu_i(s)} \leq \nu_i(s)/\mathscr{B} \leq 1/\mathscr{B} = \frac{1}{\mathscr{B}} \cdot \mathbb{E}_{\mu_1}[f_1].$$

□

Proof of Lemma 2.4. Since the desired inequality is scale invariant, we may assume $\mathbb{E}_\mu[f] = 1$, i.e. $f = \frac{d\nu}{d\mu}$ for some other distribution ν on $\{\pm 1\}^n$. Our goal is to reduce $\operatorname{Ent}_2(f_2) - 2 \cdot \operatorname{Ent}_1(f_1)$ down to the quadratic form of an appropriate correlation matrix which can be related to Ψ_μ. Towards this, observe that we can view μ_2 as a $2n \times 2n$ symmetric matrix given by

$$\mu_2((i,s),(j,t)) = \frac{1}{\binom{n}{2}} \cdot \Pr_{\sigma \sim \mu}[\sigma(i) = s, \sigma(j) = t],$$

whose rows and columns sum to $2 \cdot \mu_1$. Similarly, we can view $f_2 = \frac{d\nu_2}{d\mu_2}$ as a symmetric matrix; in particular, by the same reasoning as for μ_2, the rows and columns of $\nu_2 = f_2 \mu_2$ sum to $2 \cdot \nu_1 =$

$2 \cdot f_1\mu_1$. A quick calculation, with x, y denoting coordinate-assignments pairs in $[n] \times \{\pm 1\}$, reveals that

$$\begin{aligned}\mathrm{Ent}_1(f_1) &= \sum_{x \in [n] \times \{\pm 1\}} \mu_1(x) f_1(x) \log f_1(x) \\ &= \frac{1}{2} \cdot \mathbb{E}_{\{x,y\} \sim \mu_2} \left[f_2(x,y) \cdot \log \left(f_1(x) \cdot f_1(y) \right) \right],\end{aligned}$$

and so

$$\begin{aligned}\mathrm{Ent}_2(f_2) - 2 \cdot \mathrm{Ent}_1(f_1) &= \mathbb{E}_{\{x,y\} \sim \mu_2} \left[f_2(x,y) \cdot \left(\log f_2(x,y) - \log \left(f_1(x) \cdot f_1(y) \right) \right) \right] \\ &\geq \mathbb{E}_{\{x,y\} \sim \mu_2} \left[\left(f_2(x,y) - f_1(x) f_1(y) \right) \right] && (a \log \tfrac{a}{b} \geq a - b) \\ &= 1 - \left\langle f_1, \left(\frac{n}{n-1} Q_\mu - \frac{1}{n-1} \mathsf{Id} \right) f_1 \right\rangle_{\mu_1} && (\mathbb{E}_{\mu_2}[f_2] = 1) \\ &\geq - \left(\frac{n}{n-1} \lambda_2(Q_\mu) - \frac{1}{n-1} \right) \cdot \mathrm{Var}_1(f_1), && (\text{Poincaré Inequality for } \mathsf{Q}_\mu)\end{aligned}$$

where $Q_\mu \in \mathbb{R}^{2n \times 2n}$ is the matrix of conditional probabilities we saw in the previous lecture:

$$Q_\mu((i,s),(j,t)) = \frac{1}{n} \Pr_{\sigma \sim \mu} \left[\sigma(j) = t \mid \sigma(i) = s \right], \qquad \forall (i,s),(j,t) \in [n] \times \{\pm 1\}.$$

In particular, we already showed that $\lambda_2(Q_\mu) = \frac{\lambda_{\max}(\Psi_\mu)}{n}$, and so we're done. □

References

[Ana+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. Rome, Italy: Association for Computing Machinery, 2022, pp. 1418–1430 (cit. on pp. 1, 2, 4).

[Bar+11] Franck Barthe, Dario Cordero-Erausquin, Michel Ledoux, and Bernard Maurey. "Correlation and Brascamp–Lieb Inequalities for Markov Semigroups". In: *International Mathematics Research Notices* 2011.10 (2011), pp. 2177–2216 (cit. on p. 2).

[Bla+22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. "On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2022, pp. 3670–3692 (cit. on p. 2).

[CE22] Yuansi Chen and Ronen Eldan. "Localization Schemes: A Framework for Proving Mixing Bounds for Markov Chains". In: *arXiv preprint arXiv:2203.04163* (2022) (cit. on p. 2).

[CGM21] Mary Cryan, Heng Guo, and Giorgos Mousa. "Modified log-Sobolev inequalities for strongly log-concave distributions". In: *Annals of Probability* 49.1 (2021). A preliminary version of this paper appeared in Proceedings of the 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)., pp. 506–525 (cit. on p. 5).

[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. Virtual, Italy: Association for Computing Machinery, 2021, pp. 1537–1550 (cit. on pp. 1–5).

[CMT15] Pietro Caputo, Georg Menz, and Prasad Tetali. "Approximate tensorization of entropy at high temperature". en. In: *Annals of the Faculty of Sciences of Toulouse: Math 'ematique* Ser. 6, 24.4 (2015), pp. 691–716. DOI: 10.5802/afst.1460 (cit. on p. 2).