6.S891 Lecture 13: Spectral Independence

Kuikui Liu

October 24, 2023

In this lecture, we refocus our attention on Glauber dynamics. The goal is to prove that if the stationary measure μ satisfies a "limited correlations property" called *spectral independence* [ALO21], then Glauber dynamics mixes in polynomial-time. Furthermore, in the setting of Gibbs distributions of spin systems on bounded-degree graphs, the mixing time is actually the optimal $O(n \log n)$. The bulk of this lecture will consider simply the case of binary random variables, i.e. μ will be supported on the discrete hypercube $\{\pm 1\}^n$. This restriction is immaterial, and we will comment on how to extend everything to arbitrary *discrete product spaces*; see Section 5.

In the setting of $\{\pm 1\}^n$, for a given configuration $\sigma \in \{\pm 1\}^n$ and an index $i \in [n]$, define $\sigma^{\oplus i} \in \{\pm 1\}^n$ by setting $\sigma^{\oplus i}(i) = -\sigma(i)$ and $\sigma^{\oplus i}(j) = \sigma(j)$ for all $j \neq i$. Then Glauber dynamics P_{GD} evolves as follows: From the current configuration $\sigma \in \{\pm 1\}^n$, we

- select a uniformly random coordinate $i \sim [n]$, and
- transition to $\sigma^{\oplus i}$ with probability $\frac{\mu(\sigma^{\oplus i})}{\mu(\sigma)+\mu(\sigma^{\oplus i})}$.

Recall that P_{GD} is always reversible w.r.t. μ by design. Since we will establish functional inequalities for Glauber dynamics specifically, let us write down its Dirichlet form explicitly.

$$\begin{aligned}\mathcal{E}_{\mathsf{GD}}(f,g) &\overset{\mathsf{def}}{=} \frac{1}{2} \sum_{\sigma \in \{\pm 1\}^n} \mu(\sigma) \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{\mu\left(\sigma^{\oplus i}\right)}{\mu(\sigma) + \mu\left(\sigma^{\oplus i}\right)} \cdot \left(f(\sigma) - f\left(\sigma^{\oplus i}\right)\right) \cdot \left(g(\sigma) - g\left(\sigma^{\oplus i}\right)\right) \\ &= \mathbb{E}_{i \sim [n]}\left[\mathbb{E}_{\tau \sim \mu_{-i}}\left[\mathrm{Cov}_{\mu^{\tau}}(f,g)\right]\right].\end{aligned} \tag{1}$$

In the special case $f = g$, this in particular says that $\mathcal{E}_{\mathsf{GD}}(f,f) = \mathbb{E}_{i \sim [n]}\left[\mathbb{E}_{\tau \sim \mu_{-i}}\left[\mathrm{Var}_{\mu^{\tau}}(f)\right]\right]$.

1 Fast Mixing via Spectral Independence

Definition 1 (Spectral Independence; [ALO21]). *Let μ be a probability measure over $\{\pm 1\}^n$. Define the* (conditional) influence matrix $\Psi_\mu \in \mathbb{R}^{n \times n}$ *via*

$$\Psi_\mu(i \to j) \overset{\mathsf{def}}{=} \Pr_{\sigma \sim \mu}\left[j \leftarrow +1 \mid i \leftarrow +1\right] - \Pr_{\sigma \sim \mu}\left[j \leftarrow +1 \mid i \leftarrow -1\right], \qquad \forall i,j \in [n].$$

Note that $\Psi_\mu(i \to i) = 1$ for all $i \in [n]$. For $\eta \geq 0$, we say μ is η-spectrally independent *if $\lambda_{\max}\left(\Psi_\mu\right) \leq 1 + \eta$.*

In general, Ψ_μ is asymmetric, and may have both positive and negative entries. Nonetheless, it always has real eigenvalues by the following lemma, which also gives an equivalent characterization of spectral independence. The proof is a straightforward calculation, and is provided in Appendix A.

Fact 1.1. *Let $D_\mu \in \mathbb{R}^{n \times n}$ be the diagonal matrix with entries $D_\mu(i,i) = \mathrm{Var}_\mu(\sigma_i)$. Then $\Psi_\mu = D_\mu^{-1}\,\mathrm{Cov}(\mu)$, and μ is η-spectrally independent if and only if $\mathrm{Cov}(\mu) \preceq (1+\eta) \cdot D_\mu$.*

Example 1 (Product Measures). Suppose μ is uniform over $\{\pm 1\}^n$. Then by independence of the coordinates, we have that $\Psi_\mu = \mathsf{Id}$, and so μ is 0-spectrally independent. This is our "gold standard".

Example 2 (Extreme Bimodality). Suppose $\mu = \frac{1}{2}\delta_{+\mathbf{1}} + \frac{1}{2}\delta_{-\mathbf{1}}$. Then knowing the value of a single coordinate completely determines the values of all other coordinates. Hence, $\Psi_\mu = \mathbf{1}\mathbf{1}^\top$ and μ is $(n-1)$-spectrally independent. This distribution is particularly bad for local Markov chains because one must make a *global* sign flip in order to get from $+\mathbf{1}$ to $-\mathbf{1}$ (and vice versa).

Example 3 (Negatively Correlated (Homogeneous) Distributions). Suppose μ is supported on a slice $\binom{[n]}{k}$ of the Boolean cube $\{\pm 1\}^n \cong 2^{[n]}$, i.e. it is *homogeneous*. Further suppose μ is *negatively correlated*, i.e. $\Pr_{S\sim\mu}[j \in S \mid i \in S] \leq \Pr_{S\sim\mu}[j \in S \mid i \notin S]$ for all $i, j \in [n]$. For example, μ could be the uniform measure over spanning trees of a graph. Such distributions are extremely well-studied because they imply nice mixing [FM92; AOR16] and concentration phenomena [PP14], and also have been used extensively in the analysis of rounding schemes in approximation algorithms (see e.g. [Sri99; CVZ10]). For such distributions, all off-diagonal entries of the influence matrix Ψ_μ are nonpositive, and so

$$\begin{aligned}
\lambda_{\max}(\Psi_\mu) \leq \|\Psi_\mu\|_{\ell_\infty \to \ell_\infty} &= 1 + \max_i \sum_{j \neq i} |\Psi_\mu(i \to j)| \\
&= 1 + \max_i \left| \sum_{j\neq i} \Pr_\mu[j \in S \mid i \in S] - \sum_{j \neq i} \Pr_\mu [j \in S \mid i \notin S] \right| && \text{(Negative Correlation)} \\
&= 1 + \max_i \left| \underbrace{\mathbb{E}_\mu\left[|S| - 1 \mid i \in S\right]}_{=k-1} - \underbrace{\mathbb{E}_\mu\left[|S| \mid i \notin S\right]}_{=k} \right| && \text{(Linearity of Expectation)} \\
&= 2. && \text{(Homogeneity)}
\end{aligned}$$

Hence, such μ are 1-spectrally independent.

Our goal is to prove that spectral independence implies fast mixing of Glauber dynamics.

Theorem 1.2 (Spectral Independence $\implies$ Poincaré (Informal); [AL20; ALO21])**.** *Let μ be a probability measure on $\{\pm 1\}^n$. Suppose there exists $\eta \geq 0$ such that for every $S \subseteq [n]$ and every pinning $\tau : S \to \{\pm 1\}$, the conditional distribution μ^τ is η-spectrally independent. Then Glauber dynamics for μ has spectral gap $\Omega\left(n^{-(1+\eta)}\right)$ and mixing time $O\left(n^{2+\eta}\right)$.*

Remark 1. Technically, this statement is incorrect when $\eta > 1$, but there are straightforward ways to fix it. We chose to state it this way to keep it as simple as possible. See Remark 4 for further discussion.

Remark 2. This result was first proved in the context of the recently emerging study of *high-dimensional expanders*, where the goal is to extend the fruitful study of expander graphs to simplicial complexes/hypergraphs. It turns out, one can view μ as being a weighted simplicial complex such that its *down-up walk* is exactly Glauber dynamics. Local-to-global mixing theorems of this flavor were first established in a sequence of works [DK17; KM17; KO20].

If we make additional structural assumptions on the input distribution μ, we can do much better.

Theorem 1.3 (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:*

  For every partition $A \sqcup S \sqcup B$ such that every path from a vertex in A to a vertex in B goes through a vertex in the separator *S, and every pinning $\tau : S \to \{\pm 1\}$, the conditional measure μ^τ factorizes as $\mu_A^\tau \otimes \mu_B^\tau$. In other words, σ_A is independent from σ_B given σ_S.*

Then Glauber dynamics has spectral gap $\gamma(\mathsf{P}_{\mathsf{GD}}) \geq 1/Cn$ for some constant $C = C(\eta, \Delta) > 0$. Furthermore, if in addition μ satisfies:

- *(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 Glauber dynamics has standard and modified log-Sobolev constants $\varrho(\mathsf{P}_{\mathsf{GD}}), \kappa(\mathsf{P}_{\mathsf{GD}}) \geq 1/Cn$ for some constant $C = C(\eta, \Delta, \mathscr{B}) > 0$, and mixing time $O(n \log n)$.

We emphasize again that all of these results extend to arbitrary discrete product spaces (e.g. $[q]^n$). For the purposes of this lecture, we only prove the stated spectral gap results. We defer the proof of $O(n \log n)$ mixing to the next lecture, where we discuss *entropic independence*.

2 The High-Level Approach: Tensorization

Let us first consider how one would establish the Poincaré inequality for the simplest "ideal" setting: the uniform measure over $\{\pm 1\}^n$. There are many methods for doing this (e.g. Fourier analysis), but we focus on an inductive approach called *tensorization*: We first prove the desired statement in dimension 1, and then use independence of the coordinates to reduce the n-dimensional case to the 1-dimensional case via some sort of "decomposition". This is sometimes called the *martingale method* in the mathematical physics community, see e.g. [LY98].

To streamline the presentation, we only state the following variance inequality for product measures, which is enough for our purposes.

Lemma 2.1 (Variance Factorization from Perfect Independence)**.** *Let μ, ν be probability measures on finite state spaces Ω, Σ respectively, and let $\mu \otimes \nu$ denote the product measure on $\Omega \times \Sigma$ given by $(\mu \otimes \nu)(x, y) \stackrel{\mathsf{def}}{=} \mu(x) \cdot \nu(y)$. For $f : \Omega \times \Sigma \to \mathbb{R}$ and $\omega \in \Omega$ (resp. $\sigma \in \Sigma$), let $f^\omega : \Sigma \to \mathbb{R}$ (resp. $f^\sigma : \Omega \to \mathbb{R}$) be the specialization $f^\omega(\cdot) = f(\omega, \cdot)$ (resp. $f^\sigma(\cdot) = f(\cdot, \sigma)$). Then the following factorization inequality holds:*

$$\mathrm{Var}_{\mu \otimes \nu}(f) \leq \mathbb{E}_{\sigma \sim \nu}[\mathrm{Var}_\mu(f^\sigma)] + \mathbb{E}_{\omega \sim \mu}[\mathrm{Var}_\nu(f^\omega)], \qquad \forall f : \Omega \times \Sigma \to \mathbb{R}. \tag{2}$$

One key component in the proof of this inequality is the following decomposition of variance, which we will need throughout this lecture.

Lemma 2.2 (Law of Total Variance)**.** *Let ξ be a probability measure on a product space $\Omega \times \Sigma$ with marginals μ, ν on Ω, Σ, respectively. Then for every function $f : \Omega \times \Sigma \to \mathbb{R}$,*

$$\mathrm{Var}_\xi(f) = \mathbb{E}_{\omega \sim \mu}\left[\mathrm{Var}_{\xi^\omega}(f^\omega)\right] + \mathrm{Var}_\mu(f_\mu), \tag{3}$$

where $f_\mu : \Omega \to \mathbb{R}$ is the marginalization $f_\mu(\omega) \stackrel{\mathsf{def}}{=} \mathbb{E}_{\sigma \sim \xi^\omega}[f(\omega, \sigma)]$. The same holds with the roles of μ, ν reversed.

The proof of Lemma 2.2 is a straightforward calculation. Proving Lemma 2.1 using Lemma 2.2 is a standard exercise in convexity. Lemmas 2.1 and 2.2 both hold if we replace all occurrences of $\mathrm{Var}(\cdot)$ with $\mathrm{Ent}(\cdot)$, with the former being a special case of *Shearer's Inequalities* in information theory; see e.g. [CMT15] and references therein. These inequalities are well-known in the literature, and are sometimes called *tensorization* of variance/entropy for product measures (see e.g. [Led99; Ces01; GZ03; CMT15]). They can simultaneously be somewhat generalized to some restricted classes of Φ-*entropies* [Cha04] via *variational principles.*

3 Poincaré Inequality via Spectral Independence

Our goal is now to generalize the tensorization approach outlined in Section 2. The essence of this approach is to decompose the variance using Lemma 2.2. The key connection with spectral independence is the following weak "one-step version" of factorization of variance.

Lemma 3.1. *Let μ be a probability measure on $\{\pm 1\}^n$, and suppose μ itself is η-spectrally independent. Then for every function $f : \{\pm 1\}^n \to \mathbb{R}$,*

$$\left(1 - \frac{1 + \eta}{n}\right) \cdot \mathrm{Var}_\mu(f) \leq \mathbb{E}_{i \sim [n]}\left[\mathbb{E}_{s \sim \mu_i}\left[\mathrm{Var}_{\mu^{i \leftarrow s}}(f)\right]\right], \tag{4}$$

or equivalently,

$$\mathrm{Var}_{\mu_1}(f_1) \leq \frac{1 + \eta}{n} \cdot \mathrm{Var}_\mu(f), \tag{5}$$

where μ_1, f_1 are supported on $[n] \times \{\pm 1\}$, and given 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 $(i, s) \in [n] \times \{\pm 1\}$.

We first use it to complete the proof of rapid mixing via spectral independence.

Proof of Theorem 1.2. We repeatedly apply Eq. (4) from Lemma 3.1 to obtain

$$\begin{aligned}
\mathrm{Var}_{\mu}(f) &\leq \left(1-\frac{1+\eta}{n}\right)^{-1} \cdot \mathbb{E}_{i\sim[n]}\left[\mathbb{E}_{s\sim\mu_i}\left[\mathrm{Var}_{\mu^{i\leftarrow s}}(f)\right]\right] && (\eta\text{-spectral independence of } \mu)\\
&\leq \left(1-\frac{1+\eta}{n}\right)^{-1}\left(1-\frac{1+\eta}{n-1}\right)^{-1} \cdot \mathbb{E}_{\{i,j\}\sim\binom{[n]}{2}}\left[\mathbb{E}_{\tau\sim\mu_{ij}}\left[\mathrm{Var}_{\mu^{\tau}}(f)\right]\right] \\
&&& (\eta\text{-spectral independence of each } \mu^{i\leftarrow s})\\
&\leq \cdots && (\text{Induction})\\
&\leq \prod_{j=0}^{k-1}\left(1-\frac{1+\eta}{n-j}\right)^{-1} \cdot \mathbb{E}_{S\sim\binom{[n]}{k}}\left[\mathbb{E}_{\tau\sim\mu_S}\left[\mathrm{Var}_{\mu^{\tau}}(f)\right]\right]\\
&\lesssim \exp\left((1+\eta)\sum_{j=0}^{k-1}\frac{1}{n-j}\right) \cdot \mathbb{E}_{S\sim\binom{[n]}{k}}\left[\mathbb{E}_{\tau\sim\mu_S}\left[\mathrm{Var}_{\mu^{\tau}}(f)\right]\right]\\
&\lesssim \left(\frac{n}{n-k}\right)^{1+\eta} \cdot \mathbb{E}_{S\sim\binom{[n]}{k}}\left[\mathbb{E}_{\tau\sim\mu_S}\left[\mathrm{Var}_{\mu^{\tau}}(f)\right]\right],
\end{aligned}$$

for any $0 \leq k \leq n-1$. In particular, setting $k = n-1$ and observing that $\mathbb{E}_{S\sim\binom{[n]}{k}}\left[\mathbb{E}_{\tau\sim\mu_S}\left[\mathrm{Var}_{\mu^{\tau}}(f)\right]\right] = \mathcal{E}_{\mathsf{GD}}(f,f)$ yields the claim. □

Remark 3. Ultimately, what we have established here is that

$$\mathrm{Var}_{\mu}(f) \lesssim n^{1+\eta} \cdot \mathbb{E}_{i\sim[n]}\left[\mathbb{E}_{\tau\sim\mu_{-i}}\left[\mathrm{Var}_{\mu^{\tau}}(f)\right]\right] \qquad \forall f:\{\pm 1\}^n \to \mathbb{R}.$$

This is sometimes called *approximate tensorization of variance* [Ces01; CMT15], since by Lemma 2.1, perfect tensorization of variance when μ is a product measure exactly says that

$$\mathrm{Var}_{\mu}(f) \leq n \cdot \mathbb{E}_{i\sim[n]}\left[\mathbb{E}_{\tau\sim\mu_{-i}}\left[\mathrm{Var}_{\mu^{\tau}}(f)\right]\right] \qquad \forall f:\{\pm 1\}^n \to \mathbb{R}.$$

Remark 4. Technically, these inequalities are no longer correct if $\eta > 1$ since $1 - \frac{1+\eta}{n-k} < 0$ for $k = n-1$. The fully precise version is to let

$$\eta_k = \max_{S\in\binom{[n]}{k}} \max_{\tau:S\to\{\pm 1\}} \lambda_{\max}\left(\Psi_{\mu^{\tau}}\right),$$

which is always at most $n-k$. Then for most distributions of interest, we will actually have two bounds $\eta_k \leq \min\{\eta, C\cdot(n-k)\}$ for some universal constants $\eta \leq O(1)$ and $0 < C < 1$. The second bound $C\cdot(n-k)$ just makes everything work when $k = n - O(1)$, and will typically follow just from simple connectivity considerations. The real heart of the matter is showing that there exists $\eta \leq O(1)$ such that $\eta_k \leq \eta$ for all k.

Proof of Lemma 3.1. That the two claimed displays Eqs. (4) and (5) are equivalent follows immediately from a straightforward generalization of Lemma 2.2, which says that

$$\mathrm{Var}_{\mu}(f) = \mathrm{Var}_{\mu_1}(f_1) + \mathbb{E}_{\substack{i\sim[n]\\ s\sim\mu_i}}\left[\mathrm{Var}_{\mu^{i\leftarrow s}}(f)\right]. \tag{6}$$

Note that the expectation in the right-hand side is precisely the expectation under $(i,s)\sim\mu_1$. So, we prove the second inequality. To see the appearance of the influence matrix, we interpret $\mathrm{Var}_{\mu_1}(f_1)$ as the quadratic form of a some correlation matrix (appropriately scaled), and interpret $\mathrm{Var}_{\mu}(f)$ as essentially $\approx \|f\|_{\mu}^2$. This makes the ratio $\frac{\mathrm{Var}_{\mu_1}(f_1)}{\mathrm{Var}_{\mu}(f)}$ into a Rayleigh quotient, for which we can apply the variational characterization of eigenvalues along with η-spectral independence. Let us now formalize these calculations.

Viewing f and $\{\mu^{i\leftarrow s}\}_{i\in[n],s\in\{\pm 1\}}$ as 2^n-dimensional vectors indexed by elements of $\{\pm 1\}^n$, we may express $\mathrm{Var}_{\mu_1}(f_1)$ linear algebraically as

$$\begin{aligned}
\mathrm{Var}_{\mu_1}(f_1) &= f^{\top}\left(\frac{1}{n}\sum_{\substack{i\in[n]\\ s\in\{\pm 1\}}}\mu_i(s)\cdot\left(\mu^{i\leftarrow s}\right)\left(\mu^{i\leftarrow s}\right)^{\top}\right)f - \left\langle \mathbb{E}_{\substack{i\sim[n]\\ s\sim\mu_i}}\left[\mu^{i\leftarrow s}\right], f\right\rangle^2\\
&= \langle f, P_{\mu} f\rangle_{\mu} - \langle f, \mathbf{1}\rangle_{\mu}^2
\end{aligned}$$

where

$$P_\mu = \frac{1}{n} \sum_{\substack{i \in [n] \\ s \in \{\pm 1\}}} \frac{1}{\mu_i(s)} \cdot \left(\mathbf{1}^{i \leftarrow s}\right) \left(\mathbf{1}^{i \leftarrow s}\right)^\top \operatorname{diag}(\mu).$$

Similarly, $\operatorname{Var}_\mu(f) = \langle f, f \rangle_\mu - \langle f, \mathbf{1} \rangle_\mu^2$, and so

$$\sup_f \frac{\operatorname{Var}_{\mu_1}(f_1)}{\operatorname{Var}_\mu(f)} = \sup_{f \perp_\mu \mathbf{1}} \frac{\langle f, P_\mu f \rangle_\mu}{\langle f, f \rangle_\mu} = \lambda_2(P_\mu) = \lambda_2(M_\mu),$$

where $M_\mu = \operatorname{diag}(\mu)^{1/2} P_\mu \operatorname{diag}(\mu)^{-1/2}$ is the "symmetrized version" of the random walk matrix P_μ; note that M_μ has the same eigenvalues as P_μ. Now, even though M_μ is a huge $2^n \times 2^n$ matrix, it has extremely low rank, since it can be factorized as $M_\mu = \frac{1}{n} U_\mu U_\mu^\top$ where $U_\mu \in \mathbb{R}^{2^n \times 2n}$ has columns $\mu_i(s)^{-1/2} \operatorname{diag}(\mu)^{1/2} \mathbf{1}^{i \leftarrow s}$ for each $i \in [n], s \in \{\pm 1\}$. In particular, M_μ has the same eigenvalues as $\frac{1}{n} U_\mu^\top U_\mu \in \mathbb{R}^{2n \times 2n}$ (up to multiplicity of the zero eigenvalue), whose entries are given by

$$\left(\frac{1}{n} U_\mu^\top U_\mu\right)((i,s),(j,t)) \overset{\mathsf{def}}{=} \frac{1}{n} \frac{\left\langle \mathbf{1}^{i \leftarrow s}, \mathbf{1}^{j \leftarrow t} \right\rangle_\mu}{\sqrt{\mu_i(s) \cdot \mu_j(t)}} = \frac{1}{n} \frac{\Pr_{\sigma \sim \mu}[\sigma(i) = s, \sigma(j) = t]}{\sqrt{\Pr_{\sigma \sim \mu}[\sigma(i) = s] \cdot \Pr_{\sigma \sim \mu}[\sigma(j) = t]}}$$

This is again a symmetrized version of a random walk matrix $Q_\mu \in \mathbb{R}^{2n \times 2n}$ whose entries are given by

$$Q_\mu((i,s),(j,t)) \overset{\mathsf{def}}{=} \frac{1}{n} \Pr_{\sigma \sim \mu}[\sigma(j) = t \mid \sigma(i) = s].$$

To sum up our analysis thus far, we have proved that $\sup_f \frac{\operatorname{Var}_{\mu_1}(f_1)}{\operatorname{Var}_\mu(f)} = \lambda_2(Q_\mu)$. Hence, all that remains is to show that $\lambda_2(Q_\mu) = \frac{\lambda_{\max}(\Psi_\mu)}{n}$, which is by assumption at most $\frac{1+\eta}{n}$.

Note that Q_μ is a Markov chain on $[n] \times \{\pm 1\}$ with stationary measure μ_1. It follows that

$$\lambda_2(Q_\mu) = \frac{1}{n} \cdot \lambda_{\max}(\mathcal{I}_\mu)$$
$$\text{where} \qquad \mathcal{I}_\mu((i,s),(j,t)) = \Pr_{\sigma \sim \mu}[\sigma(j) = t \mid \sigma(i) = s] - \Pr_{\sigma \sim \mu}[\sigma(j) = t].$$

The matrix $\mathcal{I}_\mu$ already looks very similar to Ψ_μ. In particular, a quick calculation reveals that $\mathcal{I}_\mu$ can be expressed as a block matrix

$$\mathcal{I}_\mu - \mathsf{Id} = \begin{bmatrix} A_\mu & -A_\mu \\ B_\mu & -B_\mu \end{bmatrix}$$

where $A_\mu, B_\mu \in \mathbb{R}^{n \times n}$ are such that $A_\mu - B_\mu = \Psi_\mu - \mathsf{Id}$. It follows immediately that $\mathcal{I}_\mu - \mathsf{Id}$ has the same spectrum as $\Psi_\mu - \mathsf{Id}$ up to multiplicity of the zero eigenvalue, and so we're done. □

4 Optimal Spectral Gap for Sparse Graphical Models

We now sharpen our previous analysis in the setting of graphical models on bounded-degree graphs. For simplicity, we will prove a $\Omega_{\eta,\Delta}(1/n)$ lower bound on the spectral gap of Glauber dynamics without assuming marginal boundedness. This implies $O(n^2)$-mixing. We then give the full proof of $O(n \log n)$ mixing in the next lecture.

In the proof of Theorem 1.2, we saw that for every $0 \le k \le n-1$,

$$\operatorname{Var}_\mu(f) \le \left(\frac{n}{n-k}\right)^{1+\eta} \cdot \mathbb{E}_{S \sim \binom{[n]}{k}} \left[\mathbb{E}_{\tau \sim \mu_S} \left[\operatorname{Var}_{\mu^\tau}(f)\right]\right]. \tag{7}$$

If we instead took $k = (1-\theta)n$ for some constant $0 < \theta < 1$, then we'd have $\theta^{-(1+\eta)} = O(1)$. In particular, all of the loss comes from applying spectral independence after having pinned the assignments to linearly many coordinates. Hence, want to find a way to factorize the variance more efficiently in this regime. This is where conditional independence and sparsity come into play. If we pin as many as $k = \theta n$ many random vertices $S \in \binom{[n]}{k}$, then the underlying graph $G = (V, E)$

should *shatter* into many *small* connected components, between which there is pure independence. In particular, we will see that for most $S \in \binom{[n]}{k}$, we can apply Lemma 2.1 to break $\mathrm{Var}_{\mu^\tau}(f)$ into variances w.r.t. distributions on only $O(\log n)$ many vertices, even though in aggregate there are $(1-\theta)n$ total unpinned vertices. This shattering effect is crucial, and is formalized as follows; its proof is provided in Appendix A.

Lemma 4.1 (Shattering Lemma for Sparse Graphs). *Let $G = (V, E)$ be an n-vertex graph of maximum degree $\leq \Delta$. Then for every positive integer $\ell > 0$,*

$$\Pr_S[|S_v| = \ell] \leq (2e\Delta\theta)^{\ell-1},$$

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

Proof of Optimal Poincaré in Theorem 1.3. Letting $k = (1-\theta)n$ for a parameter $0 \leq \theta \leq 1$, we then have

$$\begin{aligned}
\mathrm{Var}_\mu(f) &\leq \theta^{-(1+\eta)} \cdot \mathbb{E}_{S\sim\binom{[n]}{k}}\left[\mathbb{E}_{\tau\sim\mu_{V\setminus S}}\left[\mathrm{Var}_{\mu^\tau}(f)\right]\right] \\
&\qquad\qquad\qquad \text{(Spectral Independence; see the proof of Theorem 1.2)} \\
&\leq \theta^{-(1+\eta)} \cdot \mathbb{E}_{S\sim\binom{[n]}{k}}\left[\mathbb{E}_{\tau\sim\mu_{V\setminus S}}\left[\sum_{\substack{\text{component } U \\ \text{in } G[V\setminus S]}} \mathrm{Var}_{\mu_U^\tau}(f)\right]\right] \\
&\qquad\qquad\qquad \text{(Conditional Independence, i.e. Lemma 2.1)} \\
&\leq \theta^{-(1+\eta)} \cdot \mathbb{E}_{S\sim\binom{[n]}{k}}\left[\mathbb{E}_{\tau\sim\mu_{V\setminus S}}\left[\sum_{v\in V\setminus S} C_{\mathsf{PI}}(|S_v|)\cdot \mathbb{E}_{\sigma\sim\mu_{V-v}^\tau}\left[\mathrm{Var}_{\mu_v^{\tau\sqcup\sigma}}(f)\right]\right]\right] \\
&\qquad\qquad\qquad \text{(Worst Poincaré Constant } C_{\mathsf{PI}}(\ell) \text{ over conditionals on } \ell \text{ coordinates)} \\
&= \theta^{-(1+\eta)} \cdot \sum_{v\in V} \mathbb{E}_{\tau\sim\mu_{-v}}\left[\mathrm{Var}_{\mu_v^\tau}(f)\right]\cdot \mathbb{E}_{S\sim\binom{[n]}{k}}\left[C_{\mathsf{PI}}(|S_v|)\right] \qquad \text{(Rearranging)} \\
&\leq \theta^{-(1+\eta)} n \cdot \mathbb{E}_{v\sim V}\left[\mathbb{E}_{\tau\sim\mu_{-v}}\left[\mathrm{Var}_{\mu_v^\tau}(f)\right]\right]\cdot \sum_{k=1}^{\infty}(2e\Delta\theta)^{\ell-1}C_{\mathsf{PI}}(\ell) \quad \text{(Shattering, Lemma 4.1)} \\
&\leq O_{\eta,\Delta}(n)\cdot \mathbb{E}_{i\sim[n]}\left[\mathbb{E}_{\tau\sim\mu_{-i}}\left[\mathrm{Var}_{\mu^\tau}(f)\right]\right].
\end{aligned}$$

To justify the final line, we choose $\theta \leq O(1/\Delta)$ so that $\theta^{-(1+\eta)} = \Delta^{O(1+\eta)}$ which is constant. Furthermore, $(2e\Delta\theta)^{\ell-1}$ is decaying exponentially fast in ℓ, so to ensure $\sum_{k=1}^{\infty}(2e\Delta\theta)^{\ell-1}C_{\mathsf{PI}}(\ell) \leq O(1)$, we can even afford an exponentially growing upper bound on the Poincaré constant $C_{\mathsf{PI}}(\ell)$ for the conditional measures of μ on ℓ coordinates. This permits essentially trivial bounds on the Poincaré constant which one can obtain using e.g. marginal boundedness and the fact that our Markov chain walks on a state space of size 2^ℓ; one could also just use Theorem 1.2. □

5 Beyond Binary Alphabets

Recall that a *discrete product space* is a finite domain Ω of the form $\prod_{i=1}^n \Sigma_i$ for some finite sets/alphabets $\Sigma_1, \ldots, \Sigma_n$. One can define a generalized influence matrix $\mathcal{I}_\mu$ indexed by coordinate-assignment pairs $\{(i,s) : i \in [n], s \in \Sigma_i\}$ by

$$\mathcal{I}_\mu((i,s),(j,t)) \overset{\mathsf{def}}{=} \Pr_{\sigma\sim\mu}[\sigma(j) = t \mid \sigma(i) = s] - \Pr_{\sigma\sim\mu}[\sigma(j) = t],$$

just as we did in the proof of Lemma 3.1. This doesn't quite give the same matrix as Ψ_μ in the binary setting, but it essentially has the same spectrum. So, we can again use $\lambda_{\max}(\mathcal{I}_\mu) \leq 1+\eta$ as our definition of spectral independence [ALO21; Che+21]. It also again admits a nice interpretation as an appropriate normalization of a covariance matrix (where instead we use $\{0,1\}$-indicator vectors of coordinate-assignment pairs). In some contexts, it can be convenient to use the following alternative version $\tilde{\mathcal{I}}_\mu \in \mathbb{R}_{\geq 0}^{n\times n}$ given by

$$\tilde{\mathcal{I}}_\mu(i \to j) \stackrel{\mathsf{def}}{=} \max_{s,t\in\Sigma_i} \left\| \mu_j^{i\leftarrow s} - \mu_j^{i\leftarrow t} \right\|_{\mathsf{TV}}.$$

This latter version was first proposed in [Fen+21]; note we always have $\lambda_{\max}(\mathcal{I}_\mu) \leq \rho\left(\tilde{\mathcal{I}}_\mu\right)$.

References

[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. Chicago, IL, USA: Association for Computing Machinery, 2020, pp. 1198–1211 (cit. on p. 2).

[ALO21] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model". In: *SIAM Journal on Computing* 0.0 (2021). Special Section FOCS 2020, FOCS20-1-FOCS20–37. DOI: 10.1137/20M1367696. eprint: https://doi.org/10.1137/20M1367696 (cit. on pp. 1, 2, 6).

[AOR16] Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. "Monte Carlo Markov Chain Algorithms for Sampling Strongly Rayleigh Distributions and Determinantal Point Processes". In: *COLT*. Vol. 49. Proceedings of Machine Learning Research. Columbia University, New York, New York, USA: PMLR, June 2016, pp. 103–115 (cit. on p. 2).

[Bor+13] Christian Borgs, Jennifer Chayes, Jeff Kahn, and László Lovász. "Left and right convergence of graphs with bounded degree". In: *Random Structures & Algorithms* 42.1 (2013), pp. 1–28 (cit. on pp. 8, 9).

[Ces01] Filippo Cesi. "Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields". In: *Probability Theory and Related Fields* 120.4 (2001), pp. 569–584 (cit. on pp. 3, 4).

[Cha04] Djalil Chafaï. "Entropies, convexity, and functional inequalities, On Φ-entropies and Φ-Sobolev inequalities". In: *Journal of Mathematics of Kyoto University* 44.2 (2004), pp. 325–363 (cit. on p. 3).

[Che+21] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Rapid Mixing for Colorings via Spectral Independence". In: *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '21. Virtual Event, Virginia: Society for Industrial and Applied Mathematics, 2021, pp. 1548–1557 (cit. on p. 6).

[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 p. 2).

[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 pp. 3, 4).

[CVZ10] Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. "Dependent Randomized Rounding via Exchange Properties of Combinatorial Structures". In: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2010, pp. 575–584 (cit. on p. 2).

[DK17] Irit Dinur and Tali Kaufman. "High Dimensional Expanders Imply Agreement Expanders". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. 2017, pp. 974–985. DOI: 10.1109/FOCS.2017.94 (cit. on p. 2).

[Fen+21] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. "Rapid Mixing from Spectral Independence beyond the Boolean Domain". In: *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '21. Virtual Event, Virginia: Society for Industrial and Applied Mathematics, 2021, pp. 1558–1577 (cit. on p. 7).

[FM92] Tomás Feder and Milena Mihail. "Balanced Matroids". In: *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*. STOC '92. Victoria, British Columbia, Canada: Association for Computing Machinery, 1992, pp. 26–38 (cit. on p. 2).

[GZ03] A. Guionnet and B. Zegarlinksi. "Lectures on Logarithmic Sobolev Inequalities". In: *Séminaire de Probabilités, XXXVI, volume 1801 of Lecture Notes in Math.* Springer, Berlin, 2003, pp. 1–134 (cit. on p. 3).

[KM17] Tali Kaufman and David Mass. "High Dimensional Random Walks and Colorful Expansion". In: *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Ed. by Christos H. Papadimitriou. Vol. 67. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017, 4:1–4:27. ISBN: 978-3-95977-029-3. DOI: 10.4230/LIPIcs.ITCS.2017.4 (cit. on p. 2).

[KO20] Tali Kaufman and Izhar Oppenheim. "High Order Random Walks: Beyond Spectral Gap". In: *Combinatorica* 40 (2020), pp. 245–281 (cit. on p. 2).

[Led99] Michel Ledoux. "Concentration of measure and logarithmic Sobolev inequalities". In: *Séminaire de Probabilités XXXIII (Strasbourg)* 33 (1999), pp. 120–216 (cit. on p. 3).

[LY98] Tzong-Yow Lee and Horng-Tzer Yau. "Logarithmic Sobolev inequality for some models of random walks". In: *The Annals of Probability* 26.4 (1998), pp. 1855–1873 (cit. on p. 3).

[PP14] Robin Pemantle and Yuval Peres. "Concentration of Lipschitz Functionals of Determinantal and Other Strong Rayleigh Measures". In: *Combinatorics, Probability and Computing* 23 (1 2014), pp. 140–160 (cit. on p. 2).

[Sri99] Aravind Srinivasan. "Approximation Algorithms Via Randomized Rounding: A Survey". In: *Series in Advanced Topics in Mathematics, Polish Scientific Publishers PWN.* Publishers, 1999, pp. 9–71 (cit. on p. 2).

A Unfinished Proofs

Proof of Fact 1.1. We have

$$\begin{aligned}\mathrm{Cov}_\mu(\sigma_i,\sigma_j) &= \mathbb{E}_\mu[\sigma_i\sigma_j] - \mathbb{E}_\mu[\sigma_i]\cdot\mathbb{E}_\mu[\sigma_j] \\ &= 4\left(\Pr_\mu[\sigma(i)=\sigma(j)=+1] - \Pr_\mu[\sigma(i)=+1]\cdot\Pr_\mu[\sigma(j)=+1]\right).\end{aligned}$$

Hence,

$$\begin{aligned}\left(D_\mu^{-1}\,\mathrm{Cov}(\mu)\right)(i,j) &= \frac{\Pr_\mu[\sigma(i)=\sigma(j)=+1] - \Pr_\mu[\sigma(i)=+1]\cdot\Pr_\mu[\sigma(j)=+1]}{\Pr_\mu[\sigma(i)=+1]\cdot\Pr_\mu[\sigma(i)=-1]} \\ &= \frac{\Pr_\mu[\sigma(j)=+1\mid\sigma(i)=+1] - \Pr_\mu[\sigma(j)=+1]}{\Pr_\mu[\sigma(i)=-1]} \\ &= \Psi_\mu(i\to j).\end{aligned}$$

The claim follows immediately. □

Proof of Lemma 4.1. We use the counting lemma of Borgs–Chayes–Kahn–Lovász [Bor+13], which recall says that in a graph of maximum degree $\leq \Delta$, for every vertex $v \in V$ and $\ell \in \mathbb{N}$, the number of connected induced subgraphs containing v with ℓ vertices is at most $(e\Delta)^{\ell-1}$. We saw this previously in the context of the cluster expansion. From this, if we write $k = \lceil \theta n \rceil$ for convenience, then we obtain

$$
\begin{aligned}
\Pr_S[|S_v| = \ell] &\leq \sum_{\substack{U \subseteq V \\ U \ni v, |U|=\ell \\ G[U] \text{ connected}}} \Pr_S[U \subseteq S] && \text{(Union Bound)} \\
&\leq \#\{U \subseteq V : U \ni v : |U| = \ell, G[U] \text{ connected}\} \cdot \frac{\binom{n-\ell}{k-\ell}}{\binom{n}{k}} && \text{([Bor+13])} \\
&\leq (e\Delta)^{\ell-1} \cdot \frac{(n-\ell)!}{n!} \cdot \frac{k!}{(k-\ell)!} \\
&\leq \frac{k}{n} \cdot (e\Delta)^{\ell-1} \cdot \left(\frac{k-1}{n-1}\right)^{\ell-1} \\
&\leq (2e\Delta\theta)^{\ell-1}, && \text{(Using } k = \lceil \theta n \rceil\text{)}
\end{aligned}
$$

as desired. □