TCS@UCSB Summer School on Spectral Independence	**August 8-12, 2022**

Lecture 5: *August 9th, 2022*
Lecturer: *Pietro Caputo*
Entropy Tensorization

These notes were prepared by Juspreet Singh Sandhu based on the lecture by Pietro Caputo. This is part of the 2022 Summer School on *New tools for optimal mixing of Markov chains: Spectral independence and entropy decay*, which was held at the University of California, Santa Barbara (UCSB) from August 8, 2022 to August 12, 2022. More information on the summer school is available at: https://sites.cs.ucsb.edu/~vigoda/School/

5.1 Functional Inequalities

We will take a functional analysis approach to bound the mixing time of the Markov chains of interest. We will consider the *Poincare constant* which measures the decay rate of variance and is closely related to the spectral gap of the chain, and also the (modified) log-Sobolev constant which captures the decay rate of relative entropy. The key idea in this lecture is to make connections between a particular bound on entropy through "tensorization" (or block-factorization) with the modified-Log Sobolev inequality (MLSI). The constants in these inequalities yield tight bounds on the mixing time of various families of Markov-chains.

5.1.1 Dirichlet Form

Let's first recall the definition of the Dirchlet form which measures the local variance. For an ergodic Markov chain with transition matrix P on state space Ω, let μ denote the unique stationary distribution. We assume Ω to be finite in these notes. The Dirichlet form is defined as the following:

Definition 5.1 (Dirichlet Form)**.**

$$\begin{aligned}\mathcal{D}_P(f,g) := \langle f, (\mathsf{Id}-P)g\rangle = \mathbb{E}_\mu(f(\mathsf{Id}-P)g) \\ = \sum_{\sigma,\eta\in\Omega} \mu(\sigma)P(\sigma,\eta)((f(\sigma)-f(\eta))(g(\sigma)-g(\eta)). \quad (5.1)\end{aligned}$$

When $f=g$ then the Dirichlet form measures the local variation of the function f where local is defined by the transition matrix P:

$$\mathcal{D}_P(f,f) = \sum_{\sigma,\eta\in\Omega} \mu(\sigma)P(\sigma,\eta)((f(\sigma)-f(\eta))^2.$$

Note that when $f=g$, this reduces to a quadratic form weighted by the measure μ and a symmetrization of P (baring parallels to the behavior of the weighted Laplacian of a graph, P^* is the adjoint (or the time-reversal) of P):

$$D_P(f,f) = \frac{1}{2}\sum_{\sigma,\tau}\mu(\sigma)\left[\frac{P+P^*}{2}\right](\sigma,\tau)(f(\sigma)-f(\tau))^2\,. \tag{5.2}$$

We refer to the symmetrization of P as Θ for the remainder of the note.

We now recall the definitions of variance and entropy from Pietro Caputo's first lecture.

Definition 5.2 (Variance)**.** *The variance of a function $f:\Omega\to\mathbb{R}$ with respect to μ is the following:*

$$\mathrm{Var}_A(f) := \mu[f-\mu(f)]^2 = \sum_{\sigma\in\Omega}\mu(\sigma)[f(\sigma)-\sum_{\eta\in\Omega}\mu(\eta)f(\eta)]^2 = \frac{1}{2}\sum_{\sigma,\eta\in\Omega}\mu(\sigma)\mu(\eta)(f(\sigma)-f(\eta))^2.$$

Note, variance measures the variation of f over all pairs of states whereas the Dirichlet form measures it over pairs of neighboring states.

Definition 5.3 (Entropy)**.** *The entropy of a function $f:\Omega\to\mathbb{R}$ with respect to μ is the following:*

$$\mathrm{Ent}[f] := \mu[f\log f]-\mu[f]\log(\mu[f]) = \sum_{\sigma\in\Omega}\mu(\sigma)f(\sigma)\log(f(\sigma)) - \sum_{\sigma\in\Omega}\mu(\sigma)f(\sigma)\log[\sum_{\eta\in\Omega}\mu(\eta)f(\eta)].$$

5.1.2 Log-Sobolev & Poincare Inequalities

To obtain tight mixing time bounds for Markov chains we will aim to compute *tight* bounds on the log-Sobolev inequality (LSI), modified log-Sobolev inequality (MLSI), and Poincare constant. With this motivation, we introduce the relevant inequalities.

Recall that for a Markov chain with transition matrix P, a stationary distribution is an eigenvector of P with eigenvalue 1. For an ergodic Markov chain, let $1>\lambda_2\geq\lambda_3\geq\cdots\geq\lambda_N>-1$ denote the eigenvalues of P where $N=|\Omega|$. The spectral gap of a reversible chain is defined as $\gamma=1-\lambda_2$. Note, if the chain is lazy, which means that $P(\sigma,\sigma)\geq 1/2$ for all $\sigma\in\Omega$, then the spectral gap equals the absolute spectral gap, i.e., $\gamma = 1-\max\{\lambda_2,|\lambda_N|\}$.

The Poincare constant captures the decay rate of variance, and it is equal to the spectral gap (see [LP17, Lemma 13.12]).

Lemma 5.4 (Poincare constant/Spectral gap)**.** *The spectral gap λ is the largest (universal) constant, such that, given any function $f:\Omega\to\mathbb{R}$,*

$$\lambda\cdot\mathsf{Var}[f]\leq\mathcal{D}_P(f,f)\,. \tag{5.3}$$

The above inequality is an example of a Poincare inequality as it relates the Dirichlet form (measuring local variation) to the variance, and the Dirichlet form corresponds to the derivative (with respect to time) of variance for the analogous continuous-time chain.

Note that the right-hand side in the Poincare inequality reduces to the ℓ_2-norm of the "symmetrized" gradient (as given by an appropriate Laplacian) when μ is a product measure—this observation can be made transparent by using the Efron-Stein decomposition to evaluate the Dirichlet form in such a case.

The log-Sobolev inequality is an entropy analog of the Poincare constant.

Lemma 5.5 (Log-Sobolev Inequality)**.** *The LSI constant δ_{LSI} is the largest (universal constant), such that, given any function $f : \Omega \to \mathbb{R}_{\geq 0}$,*

$$\delta_{LSI} \cdot \mathsf{Ent}_\mu[f] \leq \mathcal{D}_P(\sqrt{f}, \sqrt{f})\,. \tag{5.4}$$

An alternative version is the modified log-Sobolev inequality.

Lemma 5.6 (Log-Sobolev Inequality)**.** *The MLSI constant δ_{MLSI} is the largest (universal constant), such that, given any function $f : \Omega \to \mathbb{R}_{\geq 0}$,*

$$\delta_{MLSI} \cdot \mathsf{Ent}_\mu[f] \leq \mathcal{D}_P(f, \log(f))\,. \tag{5.5}$$

Whereas LSI has a nicer form on the RHS which often makes it easier to analyze, MLSI more directly captures the decay rate of the relative entropy and hence provides a tighter bound on the mixing time.

5.2 Functional Constants, Contractivity & Mixing

In order to relate the functional constants defined in Section 5.1.2 to the mixing time of a (discrete-time) Markov chain with transition matrix P, it is convenient to look at the continuous time analog. The kernel for the continuous time Markov chain is the following:

$$K_t = \exp(t(P - \mathsf{Id}))\,.$$

K_t can be seen as the average distribution one will be in at time t, provided P is applied in accordance with the arrivals of a Poisson process with rate t. Having established the continuous-time kernel K_t, one can associate the rate of contractivity for various quantities to be dependent on the functional constants in Section 5.1.2.

Fact 5.7 (TVD contraction via Spectral Gap). For any measure ν, the following holds,

$$\|\nu K_t - \mu\|_{TV} \leq \frac{1}{2\sqrt{\mu^*}} \exp(-\lambda t)\,. \tag{5.6}$$

In the above, $\mu^* = \min_{\omega \in \Omega}\{\mu(\omega) \mid \mu(\omega) \neq 0\}$.

Fact 5.8 (Entropy contraction via MLSI constant). For any measure ν, the following holds,

$$H(\nu K_t \mid \mu) \leq \exp(-\delta_{MLSI}\, t) H(\nu \mid \mu)\,. \tag{5.7}$$

The above inequalities will hold *iff* the corresponding functional inequalities will hold with the desired constants. To prove these relations, it suffices to take the time derivative of the appropriate inequality and relate the time derivative of the kernel K_t with the appropriate Dirichlet form.

For reversible Markov chains (i.e., $P = P^*$), the following relationship between the functional constants holds,

$$\delta_{LSI} \leq \delta_{MLSI} \leq 2\lambda\,. \tag{5.8}$$

To analyze the mixing of Markov Chains, it turns out that the modified log-Sobolev inequality is more informative than the standard log-Sobolev inequality. This can be observed by computing δ_{LSI} and δ_{MLSI} for the one-step mixing chain $P(\sigma,\tau) = \mu(\tau)$ for all $\sigma,\tau \in \Omega$. In this example, $\delta_{MLSI} = \Theta(1)$ but $\delta_{LSI} = \Theta(1/\log n)$, see [BT06, Example 3.10]. The lower bound on the log-Sobolev constant can be seen by the following general bound in terms of the spectral gap:

$$\frac{\lambda(1-2\mu^*)}{\log(\frac{1}{\mu^*}-1)} \leq \delta_{LSI}\,. \tag{5.9}$$

5.3 Block Dynamics & Approximate Tensorization

We now consider a family of Markov chains called *block dynamics* which are a generalization of the Glauber dynamics. Consider a collection of subsets of vertices. In each step of the chain we choose a random subset A from this collection and then we update all vertices in A simultaneously from the Gibbs distribution conditional on the current configuration on $\overline{A} = V \setminus A$.

Let $G = (V,E)$ be a graph. We denote by α a distribution on the subset of all blocks,

$$\alpha = \{\alpha_A \mid A \subset V\}\,. \tag{5.10}$$

Note this is a distribution and hence $\alpha_A \geq 0$ for all A and $\sum_A \alpha_A = 1$. This gives the following Markov chain, known as the block dynamics. From a configuration σ, choose a block A according to distribution α, and then resample the configuration on A from the Gibbs distribution conditional on σ on $A^c = V \setminus A$. Formally this is the following:

$$P_\alpha(\sigma,\tau) = \sum_{A\subset V} \alpha_A \cdot (\mu_A^{\sigma(A^c)}(\tau)\mathbb{1}(\tau_{A^c} = \sigma_{A^c}) = \mathbb{E}_\alpha[\mu_A^{\sigma(A^c)}(\tau)\mathbb{1}(\tau_{A^c} = \sigma_{A^c})]\,, \tag{5.11}$$

which says the probability of transitioning from configuration $\sigma \to \tau$ is given by the average (over α) conditional probability of μ over A, provided the pinned configurations are the same. A few explicit examples of such Markov chains are *Glauber dynamics* and *Even/Odd dynamics*.

Example 5.9 (Glauber Dynamics). Let $n = |V|$.

$$\alpha_A = \frac{1}{n}\mathbb{1}(|A| = 1)\,. \tag{5.12}$$

The Glauber dynamics replaces the configuration of one site (holding all others pinned) uniformly at random and hence it chooses A uniformly at random from the blocks of size one.

For bipartite graphs a natural dynamics updates one of the bipartitions in each step. We refer to this as the Even/Odd dynamics where the bipartitions are referred to as Even and Odd.

Example 5.10 (Even/Odd Dynamics).

$$\alpha_A = \frac{1}{2}\mathbb{1}(|A| = \text{even}) + \frac{1}{2}\mathbb{1}(|A| = \text{odd})\,. \tag{5.13}$$

An interesting "factorization" of entropy is given by the definition in terms of the block dynamics.

Definition 5.11 (Block Factorization of Entropy)**.** *Block Factorization $BF(C)$ holds if $\forall f \geq 0$, $\forall \alpha \in \mathcal{P}_V$, the following holds,*

$$\gamma(\alpha)\mathsf{Ent}_\mu[f] \leq C \cdot \sum_{A \subset V} \alpha_A \mathsf{Ent}_{\mu_A}[f]\,, \tag{5.14}$$

where

$$\gamma(\alpha) = \min_{x \in V} \sum_{A \subset V : x \in A} \alpha_A,$$

and $\mathsf{Ent}_{\mu_A}[f] = \sum_{\sigma \in \Omega} \mu(\sigma)\mathsf{Ent}_{\tau \sim \mu}[f(\tau)|\tau(\overline{A}) = \sigma(\overline{A})]$ is the average (over μ on all possible pinnings on $\overline{A} = V \setminus A$) conditional entropy of f on A, and P_V is the collection of distributions on subsets of V.

When μ is a product distribution, i.e., $\mu = \prod_{i=1}^m \mu_i$, then (5.14) is a weighted version of Shearer's Lemma and hence it holds with constant $C = 1$.

The usefulness of the above notion is that it yields a lower bound on the entropy contraction (as defined in Pietro Caputo's Lecture 1, see (1.9) there) in terms of $\gamma(\alpha)$.

Lemma 5.12 (Block Factorization with fixed $\alpha \implies$ Entropy Contraction)**.** *If Block Factorization of Entropy holds with constant C (for some* fixed *α), then entropy contracts with rate at least $\frac{\gamma(\alpha)}{C}$.*

The above immediately yields a mixing time bound,

$$t_{\text{mix}} \leq \frac{C}{\gamma(\alpha)} \log(\log(1/\mu^*))\,. \tag{5.15}$$

Therefore, the bound above is useful only if one has good control over C and so it is natural to consider what families of μ permit this—as stated above, a weighted version of Shearer's lemma allows one to get this for product measures μ and an elementary proof of it can be found here (proof of Lemma 1). Recent work has shown that even for more general μ, we can get good control on the constant C provided certain correlation-decay type conditions hold.

Remark 5.13 (BF(C) under Strong Spatial Mixing, [CP21])**.** For $G \subset \mathbb{Z}^d$ and $\Omega = \{\pm 1\}^{|V|}$ block factorization of entropy holds with some constant $C > 0$ provided there is *strong spatial mixing*.

The above result holds for any Gibbs measure defined on $\Omega = \{\pm 1\}^{|V|}$, see [CP21, Theorem 2.3] for more details.

A special case of block factorization of particular interest for the Glauber dynamics is when Definition 5.11 holds for $\alpha_A = \mathsf{Unif}(\binom{V}{1})$. This is called 1-uniform block factorization, and is known as *approximate tensorization*.

Definition 5.14 (Approximate Tensorization)**.** *Approximate Tensorization $AT(C)$ holds if $\forall f \geq 0$, the following holds,*

$$\frac{1}{n}\mathsf{Ent}_\mu[f] \leq C \cdot \sum_{x \in V} \mathsf{Ent}_x[f]\,, \tag{5.16}$$

where $\mathsf{Ent}_x[f] = \sum_{\sigma \in \Omega} \mu(\sigma)\mathsf{Ent}_{\tau \sim \mu}[f(\tau)|\tau(V \setminus \{x\}) = \sigma(V \setminus \{x\})$ is the average (over μ on all possible pinnings on $V \setminus \{x\}$) conditional entropy of f.

Approximate tensorization is the special case of interest of block factorization for the Glauber dynamics. By (5.15), approximate tensorization implies $O(n \log n)$ mixing time for the Glauber dynamics.

The main result in [CLV21] which will be presented in Kuikui Liu's lectures is that spectral independence implies approximate tensorization and hence optimal mixing of the Glauber dynamics.

Remark 5.15 (Approximate tensorization under Spectral Independence, [CLV21]). For any $G = (V, E)$ with $|V| = n$, $\Delta(G) = O(1)$ and $\Omega = \{\pm 1\}^n$, if the Gibbs distribution satisfies spectral independence with constant C, then it satisfies *approximate tensorization* with some $C' > 0$.

In the above result, "the local-to-global" expansion framework of HDXs is used to prove $\theta(n)$-uniform block factorization but, surprisingly, it turns out that (under reasonable constraints on the underlying graphs) this is sufficient to get approximate tensorization with constant C. The local-to-global expansion framework assumes the notion of *spectral independence*.

More generally, a recent result [BCC+22] shows that, in fact,

$$\text{Spectral Independence} \implies \text{BF}(C)\,,\, C = O(1)\,,$$

whenever the underlying graph has maximum degree $\Delta(G) = O(1)$.

Critical in proving the *full* block-factorization statement above is a trick that allows one to reduce the case of arbitrary α to that of ℓ-uniform α. The strategy is as follows:

1. Reduce proving the statement on $G = (V, E)$ to proving it for G divided into multiple partitions, where the number of partitions are not too large and each partition is an independent set. Since $\Delta(G) = O(1)$ holds, one can obtain control over the number of partitions.
2. Based on step 1, it suffices to prove block factorization of entropy for α which is uniform over the partitions. For each partition A, since A is an independent set then $\mu_A^{\sigma_{A^c}}$ is a product measure.
3. In the case of the product measure, one can obtain BF(1) via Shearer's Lemma.
4. Use *spectral independence* to "stitch" these bounds to obtain BF(C) for the complete graph with arbitrary α.

References

[BCC+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 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692, 2022.

[BT06] S. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006.

[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 Symposium on Theory of Computing (STOC)*, pages 1537–1550, 2021.

[CP21] P. Caputo and D. Parisi. Block factorization of the relative entropy via spatial mixing. *Communications in Mathematical Physics*, 388:793–818, 2021.

[LP17] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, 2017.