Localization schemes

Simons Institute - Beyond the Boolean Cube Workshop

Yuansi Chen

joint work with Ronen Eldan

Duke University

Outline

1. Spectral independence via coordinate-by-coordinate localization
2. Glauber dynamics mixing in Ising model via Eldan's stochastic localization
3. Glauber dynamics mixing in hardcore model via negative fields localization

Setup: sampling from a target measure

Given a target measure μ (possibly unnormalized), on a state space $\mathcal{X} = \{-1, +1\}^n$ or $\mathbb{R}^n$, we want to **draw samples** $X \sim \mu$.

Glauber dynamics for sampling μ on $\{-1, +1\}^n$

At current state $x \in \{-1, +1\}^n$, draw index i uniformly from $[n]$

- move to $y = x \oplus e_i$ with probability $\frac{\mu(y)}{\mu(y)+\mu(x)}$
- otherwise, stay at x

Denote this transition kernel $P_{x \to y}$.

Glauber dynamics for sampling μ on $\{-1,+1\}^n$

At current state $x \in \{-1,+1\}^n$, draw index i uniformly from $[n]$

- move to $y = x \oplus e_i$ with probability $\frac{\mu(y)}{\mu(y)+\mu(x)}$
- otherwise, stay at x

Denote this transition kernel $P_{x\to y}$.

Mixing time: starting from initial measure μ_{ini}, let $\mu_{\text{ini}}P^k$ denote the measure at time, how many iterations does it take so that

$$\text{TV}(\mu, \mu_{\text{ini}}P^k) \le \epsilon?$$

Mixing time analysis via functional inequalities (1)

Define the **Dirichlet form**

$$\mathcal{E}_P(f,g) = \langle (I-P)f, g \rangle_\mu$$

Poincaré inequality (or spectral gap)

$$\lambda \operatorname{Var}_\mu(f) \leq \mathcal{E}_P(f,f), \quad \forall f$$

For reversible lazy Markov chain, it implies variance decay:

$$\operatorname{Var}_\mu Pf \leq (1-\lambda) \operatorname{Var}_\mu f, \quad \forall f$$

Take $f = \frac{\mu_{\text{ini}} P^k}{\mu}$, we can bound chi-squared divergence decay, leading to mixing time

$$\frac{1}{\lambda}\left(\log \frac{1}{\mu_{\text{ini,min}}} + \log \frac{1}{\epsilon}\right)$$

Mixing time analysis via functional inequalities (2)

Modified Log-Sobolev inequality (MLSI)

$$\rho_{\mathsf{MLSI}} \mathrm{Ent}_{\mu}(f) \leq \mathcal{E}_P(f, \log f), \quad \forall f \geq 0$$

We can bound KL-divergence decay, leading to mixing time

$$\frac{1}{\rho_{\mathsf{MLSI}}} \left(\log \log \frac{1}{\mu_{\mathrm{ini,min}}} + \log \frac{1}{\epsilon} \right)$$

From now on, we focus on functional inequalities

- Target measure μ
- $2^n \times 2^n$ Markov transition kernel P
- To prove mixing time, it suffice to prove

$$\lambda \operatorname{Var}_\mu(f) \leq \mathcal{E}_P(f, f)$$

For product measure, it is easy.
Other than that, for what kind of target measure, can we prove spectral gap?

Coordinate-by-coordinate localization

Spectral independence [Anari, Liu, Oveis Gharan '20]

Define the $n \times n$ pairwise influence matrix Ψ_μ

$$\Psi_\mu[i,j] = \mathbb{P}_{x\sim\mu}(x_j = +1 \mid x_i = +1) - \mathbb{P}_{x\sim\mu}(x_j = +1 \mid x_i = -1)$$

μ is η-spectrally independent if

$$\|\Psi_\mu\|_2 \leq \eta$$

Spectral independence [Anari, Liu, Oveis Gharan '20]

Define the $n \times n$ pairwise influence matrix Ψ_μ

$$\Psi_\mu[i,j] = \mathbb{P}_{x\sim\mu}(x_j = +1 \mid x_i = +1) - \mathbb{P}_{x\sim\mu}(x_j = +1 \mid x_i = -1)$$

μ is η-spectrally independent if

$$\|\Psi_\mu\|_2 \leq \eta$$

A sufficient condition for proving spectral gap: if all conditionals of μ (the law of $X \mid X_i = \pm 1$ and $X \mid X_i = \pm 1, X_j = \pm 1$, etc.) are η-spectrally independent, then spectral gap

$$\lambda \geq \prod_{i=0}^{n-2} \left(1 - \frac{\eta}{n-i}\right)$$

Spectral independence is a condition on covariance

Since

$$\mathsf{Cov}_\mu = \mathsf{diag}(\mathsf{Cov}_\mu)(\Psi_\mu + \mathbb{I}_n)$$

we have

$$\mathsf{Cov}_\mu \preceq (1+\eta)\,\mathsf{diag}(\mathsf{Cov}_\mu) \Leftrightarrow \|\Psi_\mu + \mathbb{I}_n\|_2 \leq 1+\eta.$$

Spectral independence is a condition on covariance

Since

$$\mathsf{Cov}_\mu = \mathsf{diag}(\mathsf{Cov}_\mu)(\Psi_\mu + \mathbb{I}_n)$$

we have

$$\mathsf{Cov}_\mu \preceq (1+\eta)\,\mathsf{diag}(\mathsf{Cov}_\mu) \Leftrightarrow \|\Psi_\mu + \mathbb{I}_n\|_2 \leq 1+\eta.$$

Constraining the covariance makes sense, but

Q1: why do we have to put assumptions on all conditionals?

...trickling down, HDX, local-to-global

Q2: what are other ways to put assumptions to prove spectral gap, when direct proof is difficult?

What are localization schemes?

A localization scheme is a mapping from measure ν to a stochastic process $(\nu_t)_{t \geq 0}$ such that

- $\nu_0 = \nu$
- For any measurable A, $\nu_t(A)$ is a martingale (in other words, $\mathbb{E}[\nu_t(A) \mid \{\nu_\tau(A), \tau \leq s\}] = \nu_s(A), \forall 0 \leq s \leq t$)

What are localization schemes?

A localization scheme is a mapping from measure ν to a stochastic process $(\nu_t)_{t\geq 0}$ such that

- $\nu_0 = \nu$
- For any measurable A, $\nu_t(A)$ is a martingale (in other words, $\mathbb{E}[\nu_t(A) \mid \{\nu_\tau(A), \tau \leq s\}] = \nu_s(A), \forall 0 \leq s \leq t$)

Our main standpoint:

- You pick a localization scheme
- Study the evolution of the variance $\mathrm{Var}_{\nu_t}(f)$ along the process $(\nu_t)_t$
- Put assumptions to approximately conserve variance, then you can prove spectral gap!

Spectral independence assumption comes from coordinate-by-coordinate localization

Coordinate-by-coordinate localization

Start from ν on $\{-1,+1\}^n$. Let $(k_1,\ldots,k_n)$ be a random permutation of $[n]$, and X is a random draw from ν, independent of the rest. Define

$$\nu_i = \text{law of } \left\{X \mid X_{k_1},\ldots,X_{k_i}\right\}$$

Spectral independence assumption comes from coordinate-by-coordinate localization

Coordinate-by-coordinate localization

Start from ν on $\{-1, +1\}^n$. Let $(k_1, \ldots, k_n)$ be a random permutation of $[n]$, and X is a random draw from ν, independent of the rest. Define

$$\nu_i = \text{law of } \left\{X \mid X_{k_1}, \ldots, X_{k_i}\right\}$$

We claim that

In [Anari, Liu, Oveis Gharan '20], η-spectrally independence for every conditional of ν is a condition to conserve variance along the coordinate-by-coordinate localization

$$\left(1 - \frac{\eta}{n-i}\right) \mathbb{E}[\mathrm{Var}_{\nu_i}(f)] \leq \mathbb{E}[\mathrm{Var}_{\nu_{i+1}}(f)]$$

Derivation: approximate conservation of variance

Similarly,

- Semi-log-concavity [Eldan, Shamir '20]
- Fractional log-concavity [Alimohammadi, Anari, Shiragur, Vuong '21]
- Entropic independence [Anari, Jain, Koehler, Pham, Vuong '21]

which bounds covariance of all tilted measures,

are sufficient conditions to approximately conserve entropy
along the coordinate-by-coordinate localization
so that one could prove MLSI

Beyond coordinate-by-coordinate localization?

Let's first take a tour **beyond the Boolean cube** to $\mathbb{R}^n$, where Eldan first introduced stochastic localization [Eldan '13]

Eldan’s stochastic localization

Eldan's stochastic localization [Eldan '13]

Given an density ν on $\mathbb{R}^n$, the density at time t is the solution of the SDE

$$d\nu_t(x) = (x - \mathbf{b}(\nu_t))^\top C_t^{\frac{1}{2}} dW_t \cdot \nu_t(x), \quad \forall x \in \mathbb{R}^n$$

where $\mathbf{b}(\nu_t)$ is the mean of ν_t and W_t is the Brownian motion. Take $C_t = \mathbb{I}_n$ to simplify explanation.

Explicit form of the random measure at time t

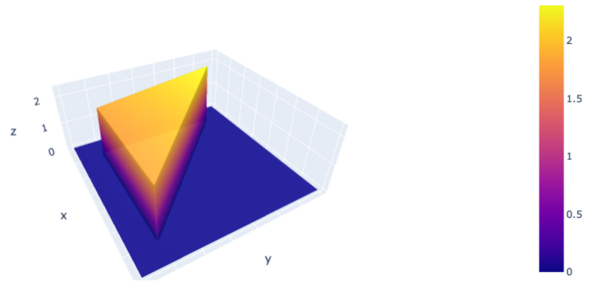
ν_t has an explicit form

$$\nu_t(x) = \frac{1}{Z(c_t, t)} \exp\left(-\frac{t}{2}|x|^2 + c_t^\top x\right)\nu(x)$$
$$dc_t = dW_t + \mathbf{b}(\nu_t)dt$$

At time t, the initial density is multiplied by a Gaussian with $1/t$ variance, while the center of the Gaussian is random.

Demonstration of Eldan's stochastic localization in 2 dimension

Initialized with uniform distribution over a convex set ($n = 2$)



Stochastic localization are used in high dimensional probability

Say we want to show a “property A” of the density ν

- **Transform** via stochastic localization
- **Prove** “property A” for ν_t (usually easier)
- **Relate** “property A” of ν_t to that of ν (via SDE analysis)

Stochastic localization are used in high dimensional probability

Say we want to show a "property A" of the density ν

- **Transform** via stochastic localization
- **Prove** "property A" for ν_t (usually easier)
- **Relate** "property A" of ν_t to that of ν (via SDE analysis)

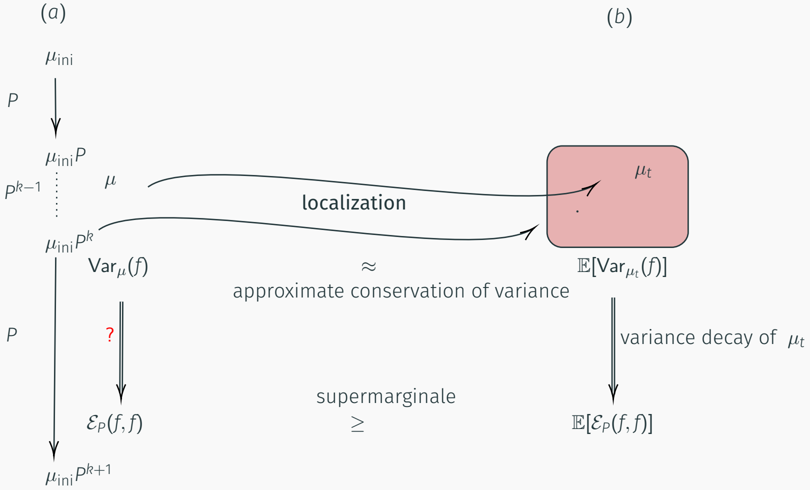
See survey paper in 2022 ICM proceedings [Eldan], "property A" can be

- isoperimetric inequality (e.g. KLS conjecture [KLS '95])
- concentration of Lipschitz functions in Gaussian space
- noise stability inequality
- Poincaré inequality ...

Focus on sampling

1. The desired functional inequality is then our "property A"
2. Hopefully, this "property A" is easier to prove for the process at some time t
3. We put assumptions to make the approximate conservation of variance analysis go through

Use of localization schemes for sampling proofs



Ising model

The probability measure on $\{-1, +1\}^n$ defined as

$$\mu(x) \propto \exp(\langle x, Jx \rangle + \langle h, x \rangle)$$

is called Ising model with interaction matrix $J \in \mathbb{R}^{n \times n}$ and external field $h \in \mathbb{R}^n$.

Glauber dynamics on Ising model

Theorem

Let $\nu_{\tau,v}(x) \propto \mu(x)\exp(-\tau\langle x, Jx\rangle + \langle v, x\rangle)$ If

$$\mathrm{Cov}_{\nu_{\tau,v}} \preceq \alpha(\tau)\mathbb{I}_n, \quad \forall\tau \in [0,1], \forall v$$

Then the MLSI constant of Glauber dynamics

$$\rho_{\mathrm{MLSI}} \geq \frac{1}{n}\exp\left(-2\|J\|_2\int_0^1 \alpha(\tau)d\tau\right)$$

Glauber dynamics on Ising model

Theorem

Let $\nu_{\tau,v}(x) \propto \mu(x) \exp(-\tau \langle x, Jx \rangle + \langle v, x \rangle)$ If

$$\mathrm{Cov}_{\nu_{\tau,v}} \preceq \alpha(\tau)\mathbb{I}_n, \quad \forall \tau \in [0,1], \forall v$$

Then the MLSI constant of Glauber dynamics

$$\rho_{\mathrm{MLSI}} \geq \frac{1}{n} \exp\left(-2 \|J\|_2 \int_0^1 \alpha(\tau) d\tau\right)$$

For J be a positive-definite matrix with $\|J\|_2 < \frac{1}{2}$ and $v \in \mathbb{R}^n$, adapting Bauerschmidt, Dagallier '22, we have

$$\|\mathrm{Cov}\,(\nu_{\tau,v})\|_2 \leq \frac{1}{1 - 2(1-\tau)\|J\|_2},$$

leading to $\rho_{\mathrm{MLSI}} \geq \frac{1}{n}(1 - 2\|J\|_2)$.

Tightness

- The condition $\|J\|_2 \leq \frac{1}{2}$ is tight in general, as it is tight for Curie-Weiss model
- However, for the Sherrington-Kirkpatrick model, which assumes $J = \frac{\beta}{2} A$ where A is drawn from $\mathrm{GOE}(n)$. The above approach only gets fast mixing of Glauber dynamics for $\beta < \frac{1}{4}$, while the conjectured phase transition is at $\beta < 1$.

What happens when we apply Eldan's stochastic localization?

Take control matrix $C_t = (2J)$, for $t \in [0, 1]$,

$$\nu_t(x) \propto \mu(x) \exp(-t \langle x, Jx \rangle + \langle c_t, x \rangle)$$
$$\propto \exp((1 - t) \langle x, Jx \rangle + \langle h + c_t, x \rangle)$$

where $c_t = C_t^{\frac{1}{2}} dW_t + \mathbf{b}(\nu_t) dt$.

What happens when we apply Eldan's stochastic localization?

Take control matrix $C_t = (2J)$, for $t \in [0, 1]$,

$$\nu_t(x) \propto \mu(x) \exp(-t \langle x, Jx \rangle + \langle c_t, x \rangle)$$
$$\propto \exp((1-t) \langle x, Jx \rangle + \langle h + c_t, x \rangle)$$

where $c_t = C_t^{\frac{1}{2}} dW_t + \mathbf{b}(\nu_t) dt$.

At time $t = 1$, ν_t becomes a product measure (so easy to show MLSI).

Let's take a look at the evolution of entropy

Evolution of entropy along Eldan's SL

For $f : \mathcal{X} \to \mathbb{R}_+$

$$d\mathrm{Ent}_{\nu_t}[f] = -\frac{1}{2}\mathbb{E}_{\nu_t}[f]\left|C_t^{\frac{1}{2}}(\mathbf{b}(\omega_t) - \mathbf{b}(\nu_t))\right|^2 dt + \text{martingale}$$

where ω_t is the probability measure $\propto f\nu_t$.

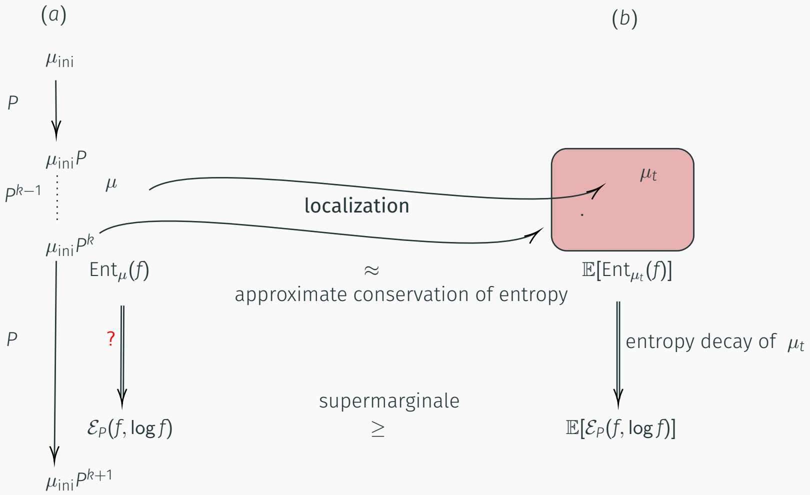
Additionally, if $\mathrm{Cov}(\mathcal{T}_v\nu_t) \preceq A_t, \forall v$, then

$$\frac{1}{2}\mathbb{E}_{\nu_t}[f]\left|C_t^{\frac{1}{2}}(\mathbf{b}(\omega_t) - \mathbf{b}(\nu_t))\right|^2 \leq \left\|C_t^{\frac{1}{2}}A_tC_t^{\frac{1}{2}}\right\|_2 \mathrm{Ent}_{\nu_t}[f]$$

Solving the equation, we obtain approximate conservation of entropy

$$\mathbb{E}[\mathrm{Ent}_{\nu_t}[f]] \geq e^{-2\|J\|_2 \int_0^t \alpha(\tau)d\tau}\mathrm{Ent}_{\nu_0}[f]$$

Use of localization schemes for entropy decay



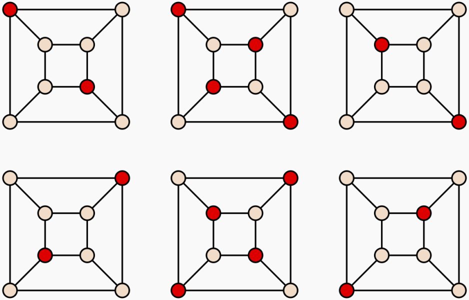
Negative-fields localization

The hardcore model

Given a graph $G = (V, E)$ with $|V| = n$, a hardcore model with fugacity λ on $\{-1, +1\}^n$ is

$$\mu(\sigma) \propto \lambda^{|I_\sigma|},$$

where $\mu(\sigma) > 0$ if the set $I_\sigma = \{v \in V \mid \sigma_v = +1\}$ coorresponds to an independent set of G.



Negative-fields localization

Given a measure ν on $\{-1,1\}^n$, the process $\{\nu_t\}_{t\geq 0}$ evolves as

- For $x \in \{-1,1\}^n$, ν_t solves the SDE

$$d\nu_s(x) = \nu_s(x) \langle x - \mathbf{b}(\nu_s), dJ_s \rangle ,$$

where

$$dJ_{s,i} = -ds + \frac{1}{1+\mathbf{b}(\nu_s)_i} N_{s,i}$$

where $N_{s,i}$ is a Poisson point process with intensity $1+\mathbf{b}(\nu_s)_i$

Inspired by field dynamics in Chen, Feng, Yin, and Zhang '21

How does the measure ν_t look like?

- At time t, define $A_t = \left\{i \in \{1, \ldots, n\} \mid N_{t,i} \geq 1\right\}$. Since $N_{t,i}$ is non-decreasing, A_t is an almost surely non-descreasing process of subsets of $\{1, \ldots, n\}$.
- We can write ν_t as

$$\nu_t = \mathcal{T}_{-t\vec{\mathbf{1}}} \mathcal{R}_{A_t} \nu$$

"ν_t is the density obtained by pinning all coordinates in A_t to $+1$ and then tilt by $-t\vec{\mathbf{1}}$"

How does the measure ν_t look like?

- At time t, define $A_t = \left\{i \in \{1, \ldots, n\} \mid N_{t,i} \geq 1\right\}$. Since $N_{t,i}$ is non-decreasing, A_t is an almost surely non-descreasing process of subsets of $\{1, \ldots, n\}$.
- We can write ν_t as

$$\nu_t = \mathcal{T}_{-t\vec{1}} \mathcal{R}_{A_t} \nu$$

"ν_t is the density obtained by pinning all coordinates in A_t to $+1$ and then tilt by $-t\vec{1}$"

What is remaining?

- The mixing analysis on measures with large tilts are well-known in [Erbar, Henderson, Menz and Tetali '17]
- We need to study the evolution of the process: this is where we use properties of the hardcore model to ensure approximate conservation of entropy.

Summary

- Introduced localization schemes to analyze mixing
- For each localization scheme,
  - we can study the evolution of variance (or entropy)
  - assumptions to ensure the approximate conservation of variance (or entropy) are usually the key assumptions
- Designing Localization schemes allows us to take advantage of our insights about target distributions
  - Recover results of spectral independence/fractional log-concavity
  - Optimal $O(n \log n)$ Glauber dynamics mixing bound for Ising models in the uniqueness regime under any external fields
  - $O(n \log n)$ Glauber dynamics mixing bound for the hardcore model in the tree-uniqueness regime

Thank you!