A Local-to-Global Framework: Localization Schemes

Zhidan Li

Abstract

This is a technique handbook of the newly developed framework — localization schemes — related to the topic of local-to-global. Briefly saying, we focus on the tools that are useful in the field of approximate counting and sampling.

1 Localization Schemes and Markov Chains

Now we introduce another framework to show the local-to-global theorem. This framework, named the *localization schemes*, is highly related to the recent breakthrough of the famous Kannan-Lovász-Simonovits Conjecture, and deeply studied in Chen and Eldan [CE22] to analyze the mixing time of the Markov chains.

We fix a state space Ω equipped with a σ-algebra Σ. Usually we assume that $\Sigma = 2^{\Omega}$ when Ω is finite and $\Sigma = \mathrm{Borel}(\Omega)$ when Ω is a continuous space, and then we omit Σ. Let $\mathcal{M}(\Omega)$ be the space of all probability measures on Ω.

Definition 1.1 (Localization Process)**.** A *localization process* $(\mu_t)_{t\geq 0}$ on the state space Ω is a stochastic process satisfying

(P1) Almost surely μ_t is a probability measure on Ω for all $t \geq 0$.

(P2) For every measurable $A \subseteq \Omega$, the process $(\mu_t(A))_{t\geq 0}$ is a martingale.

(P3) For every measurable $A \subseteq \Omega$, the process $(\mu_t(A))_{t\geq 0}$ almost surely converges to either 0 or 1 as $t \to \infty$.

For convenience, we use Θ_t to denote the distribution of μ_t for every $t \geq 0$.

Definition 1.2 (Localization Scheme)**.** A *localization scheme* $\mathcal{L}$ on Ω is a mapping assigning to each probability measure $\mu \in \mathcal{M}(\Omega)$ a localization process $(\mu_t)_{t\geq 0}$ with $\mu_0 = \mu$. In this case, we say $(\mu_t)_t$ is the localization process associated with μ via the localization scheme $\mathcal{L}$.

For convenience, for every $t \geq 0$, let $\Gamma_{\mu,t}$ be the collection of all possible probability measures at time t, and $\Theta_{\mu,t}$ be the probability measure on $\Gamma_{\mu,t}$ where for every $\nu \in \Gamma_{\mu,t}$, $\Theta_{\mu,t}(\nu)$ equals the probability such that $\mu_t = \nu$ with $\mu_0 = \mu$ under $\mathcal{L}$. Usually μ is clear, and we will drop the subscript μ.

1.1 Doob localization schemes

Now we set a kind of fancy-looking localization scheme as an example. Given a state space $\Omega \subseteq \mathbb{R}^n$ with a σ-algebra Σ, we say that a filtration $\mathcal{F}_t$ of Ω is *precise* if

$$\lim_{t\to\infty} \mathcal{F}_t = \Sigma.$$

Then let $\mathcal{M}_F(\Omega)$ be the space of probability measures over the set of precise filtrations on Ω.

Fix a probability measure $\Theta \in \mathcal{M}_F(\Omega)$. For a probability measure μ on Ω, we construct a localization process $(\mu_t)_{t\geq 0}$ as

$$\mu_t(A) = \mathbf{E}_{(\mathcal{F}_t)\sim\Theta}\left[\mathbf{Pr}_{X\sim\mu}\left[X \in A \mid \mathcal{F}_t\right]\right], \quad \forall A \subseteq \Omega.$$

Such a localization scheme is named a *Doob localization scheme*. It has been shown that this kind of localization scheme is strongly related to the *diffusion process* from an information-theoretic view. See [Mon23] for detailed arguments.

1.2 Markov dynamics associated with the localization process

In this part we associate a localization process $(\mu_t)_{t\geq 0} = \mathcal{L}(\mu)$ with a Markov dynamics reversible with respect to the distribution $\mu \in \mathcal{M}(\Omega)$.

Definition 1.3 (Markov Chains Associated with Localization Processes)**.** Let $(\mu_t)_{t\geq 0}$ be a localization process on Ω associated with μ via a localization scheme $\mathcal{L}$ and $\tau > 0$ be a stopping time. The Markov dynamics $P = P^{(\mathcal{L},\tau)}$ associated with $(\mu_t)_{t\geq 0}$ and τ is defined as

$$P(x, A) = \mathbf{E}_{\Theta_t}\left[\frac{\mu_\tau(x)\mu_\tau(A)}{\mu(x)}\right], \quad \forall x \in \Omega, A \in \Sigma.$$

Remark 1.4. An optional way to view Definition 1.3 is, let X, Y be two random variables taking values in $\Omega\times\Omega$ satisfying

$$\mathbf{Pr}\left[X \in A, Y \in B\right] = \mathbf{E}\left[\mu_\tau(A)\mu_\tau(B)\right], \quad \forall A, B \in \Sigma.$$

Then we define the kernel as

$$P(x, A) = \mathbf{Pr}\left[Y \in A \mid X = x\right].$$

Fact 1.5. *Let $P = P^{(\mathcal{L},\tau)}$ be the transition kernel defined as Definition 1.3. Then P is reversible with respect to μ.*

Proof. For every $x \in \Omega$, it almost surely holds that

$$P(x, \Omega) = \mathbf{E}_{\Theta_t}\left[\frac{\mu_\tau(x)\mu_\tau(\Omega)}{\mu(x)}\right] = \mathbf{E}_{\Theta_t}\left[\frac{\mu_\tau(x)}{\mu(x)}\right] = 1.$$

Then we know $P(x, \cdot)$ is a probability measure on Ω almost surely. Also for every $A, B \in \Sigma$, it holds that

$$\begin{aligned}
\int_{x\in A} P(x, B)\ \mathrm{d}\mu(x) &= \int_{x\in A} \mathbf{E}\left[\frac{\mathrm{d}\mu_\tau(x)}{\mathrm{d}\mu(x)}\mu_\tau(B)\right]\ \mathrm{d}\mu(x) \\
&= \mathbf{E}\left[\int_{x\in\Omega} \mu_\tau(B)\ \mathrm{d}\mu_\tau(x)\right] \\
&= \mathbf{E}\left[\mu_\tau(A)\mu_\tau(B)\right] \\
&= \int_{y\in B} P(y, A)\ \mathrm{d}\mu(y).
\end{aligned}$$

Therefore we know P is reversible with respect to μ. □

To view the Markov chain more clearly, consider the following two-step transition: at the current state $x \in \Omega$,

- firstly we draw a probability measure $\nu \in \Gamma_\tau$ following probability $\Theta_\tau(\nu) \cdot \frac{\nu(x)}{\mu(x)}$;
- then we draw the next state $y \sim \nu$.

Define the transition operators $\mathcal{D}_\mu^{(t)} : \Omega \times \Gamma_t \to \mathbb{R}$ and $\mathcal{U}_\mu^{(t)} : \Gamma_t \times \Omega \to \mathbb{R}$ as

$$\mathcal{D}_\mu^{(t)}(x, \nu) = \Theta_\tau(\nu) \cdot \frac{\nu(x)}{\mu(x)}, \quad \mathcal{U}_\mu^{(t)}(\nu, x) = \nu(x), \quad \forall x \in \Omega, \nu \in \Gamma_t.$$

It is easy to see $P^{(\mathcal{L},\tau)} = \mathcal{D}_\mu^{(t)}\mathcal{U}_\mu^{(t)}$.

1.3 Functional inequalities

Recall the Dirichlet form of a random walk P with stationary distribution μ: for two functions $f, g : \Omega \to \mathbb{R}$,

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

and the spectral gap and modified log-Sobolev inequality constant of P:

$$\mathrm{Gap}(P) := \inf_{f:\Omega\to\mathbb{R}} \frac{\mathcal{E}_P(f,f)}{\mathbf{Var}_\mu[f]}, \quad \rho_{\mathrm{LS}}(P) := \inf_{f:\Omega\to\mathbb{R}_{>0}} \frac{\mathcal{E}_P(f,\log f)}{\mathbf{Ent}_\mu[f]}.$$

The following identity and inequality illustrate the connection between the functional inequalities and the variance or entropy of the localization process.

Proposition 1.6. *Let $P = P^{(\mathcal{L},\tau)}$ be a transition kernel associated with a localization process $(\mu_t)_{t\ge 0} = \mathcal{L}(\mu)$ and $\tau > 0$. Then it holds that*

$$\mathcal{E}_P(f,f) = \mathbf{E}_{\Theta_\tau}\left[\mathbf{Var}_{\mu_\tau}[f]\right], \quad \mathcal{E}_P(f,\log f) \ge \mathbf{E}_{\Theta_\tau}\left[\mathbf{Ent}_{\mu_\tau}[f]\right].$$

for every function f supported on Ω when the Dirichlet forms are well-defined.

Proof. We prove them one by one. By calculation,

$$\begin{aligned}
\mathcal{E}_P(f,f) &= \int_{x\in\Omega} f(x)(I-P)f(x)\ \mathrm{d}\mu(x) \\
&= \int_{x\in\Omega} \left(f(x)^2 - f(x)(Pf)(x)\right)\ \mathrm{d}\mu(x) \\
&= \int_{x\in\Omega} f(x)^2\ \mathrm{d}\mu(x) - \int_{x\in\Omega} f(x)\left(\int_{y\in\Omega} f(y)\ \mathrm{d}P(x,y)\right)\ \mathrm{d}\mu(x) \\
&= \mathbf{E}_\mu\left[f^2\right] - \int_{x\in\Omega}\int_{y\in\Omega} f(x)f(y)\mathbf{E}_{\Theta_\tau}\left[\frac{\mathrm{d}\mu_\tau(x)}{\mathrm{d}\mu(x)}\ \mathrm{d}\mu_\tau(y)\right]\ \mathrm{d}\mu(x) \\
&= \mathbf{E}_\mu\left[f^2\right] - \mathbf{E}_{\Theta_\tau}\left[\int_{x\in\Omega} f(x)\left(\int_{y\in\Omega} f(y)\ \mathrm{d}\mu_\tau(y)\right)\ \mathrm{d}\mu(x)\right] \\
&= \mathbf{E}_{\Theta_\tau}\left[\mathbf{E}_{\mu_\tau}\left[f^2\right] - \mathbf{E}_{\mu_\tau}[f]^2\right] \\
&= \mathbf{E}_{\Theta_\tau}\left[\mathbf{Var}_{\mu_\tau}[f]\right]
\end{aligned}$$

where the identity $\mathbf{E}_\mu\left[f^2\right] = \mathbf{E}_{\Theta_\tau}\left[\mathbf{E}_{\mu_\tau}\left[f^2\right]\right]$ holds from the martingality of the process.

For the MLSI constant, by calculation, we know

$$\begin{aligned}
\mathcal{E}_P(f,\log f) &= \int_{x\in\Omega} f(x)\left((I-P)\log f\right)(x)\ \mathrm{d}\mu(x) \\
&= \int_{x\in\Omega} \left(f(x)\log f(x) - f(x)(P\log f)(x)\right)\ \mathrm{d}\mu(x) \\
&= \int_{x\in\Omega} f(x)\log f(x)\ \mathrm{d}\mu(x) - \int_{x\in\Omega} f(x)\left(\int_{y\in\Omega} \log f(y)\ \mathrm{d}P(x,y)\right)\ \mathrm{d}\mu(x) \\
&= \mathbf{E}_\mu[f\log f] - \int_{x\in\Omega}\int_{y\in\Omega} f(x)\log f(y)\mathbf{E}_{\Theta_\tau}\left[\frac{\mathrm{d}\mu_\tau(x)}{\mathrm{d}\mu(x)}\ \mathrm{d}\mu_\tau(y)\right]\ \mathrm{d}\mu(x) \\
&= \mathbf{E}_\mu[f\log f] - \mathbf{E}_{\Theta_\tau}\left[\int_{x\in\Omega} f(x)\left(\int_{y\in\Omega} \log f(y)\ \mathrm{d}\mu_\tau(y)\right)\ \mathrm{d}\mu(x)\right] \\
&= \mathbf{E}_{\Theta_\tau}\left[\mathbf{E}_{\mu_\tau}[f\log f] - \mathbf{E}_{\mu_\tau}[f]\,\mathbf{E}_{\mu_\tau}[\log f]\right] \\
&\ge \mathbf{E}_{\Theta_\tau}\left[\mathbf{E}_{\mu_\tau}[f\log f] - \mathbf{E}_{\mu_\tau}[f]\log \mathbf{E}_{\mu_\tau}[f]\right] \\
&= \mathbf{E}_{\Theta_\tau}\left[\mathbf{Ent}_{\mu_\tau}[f]\right]
\end{aligned}$$

where the inequality holds from the Jensen's inequality $\log \mathbf{E}_\pi [f] \geq \mathbf{E}_\pi [\log f]$ for every distribution π on Ω and every test function $f : \Omega \to \mathbb{R}_{>0}$. □

2 Linear-Tilt Localization Processes

Now we introduce a family of localization processes that lies at the core of the analysis of the mixing time. For a distribution π on Ω, we use $\mathbf{b}(\pi)$ to denote the mass center of π, *i.e.*,

$$b(\pi) = \int_{x \in \Omega} x \, \mathrm{d}\pi(x).$$

Definition 2.1 (Linear-Tilt Localization Processes)**.** For a localization process $(\mu_t)_{t\geq 0}$, we say it is a *linear-tilt localization process* if:

- **(Discrete version)** For all $t \in \mathbb{N}$ and $x \in \Omega$,

$$\mu_{t+1}(x) = \mu_t(x) \left(1 + \langle x - \mathbf{b}(\mu_t), Z_t \rangle\right) \tag{1}$$

  where Z_t is a random variable with $\mathbf{E}[Z_t \mid \mu_t] = 0$. Or,

- **(Continuous version)** For all $t \geq 0$ and $x \in \Omega$,

$$\mathrm{d}\mu_t(x) = \mu_t(x) \langle x - \mathbf{b}(\mu_t), Z_t \rangle \tag{2}$$

  where Z_t is a random variable with $\mathbf{E}[Z_t \mid \mu_t] = 0$.

For convenience, we say $(Z_t)_{t\geq 0}$ is the driving factor of $(\mu_t)_{t\geq 0}$.

We will focus on two localization schemes: (1) the coordinate-by-coordinate localization schemes; and (2) the stochastic localization schemes driven by standard Brownian motion.

2.1 The coordinate-by-coordinate localization schemes

Given a distribution μ over $\Omega \subseteq \mathbb{R}^n$, we construct a discrete-time localization process $(\mu)_{t\geq 0}$ as follows:

- Firstly we pick a permutation $k_1, \ldots, k_n$ of $[n]$ uniformly at random.
- Let $X \sim \mu$. For $t \geq 0$, we set μ_t to be the law of X conditional on $X_{k_1}, \ldots, X_{k_i}$ where $i = \min\{n, \lfloor t \rfloor\}$.

Now we show the observation that the dynamics associated with the coordinate-by-coordinate localization process is the well-known *Glauber dynamics*.

Fact 2.2. *Given a coordinate-by-coordinate localization scheme $\mathcal{L}$ over $\Omega \subseteq \mathbb{R}^n$ and an integer $\tau = n - 1$, the Markov chain $P = P^{(\mathcal{L},\tau)}$ associated with $(\mu_t)_{t\geq 0} = \mathcal{L}(\mu)$ and τ is the single-site Glauber dynamics denoted by* P^{GD} *with stationary distribution μ.*

Proof. We verify the fact by definition. For every $x \in \Omega$ and $i \in [n]$, define $L_{x,i} := \left\{z \in \Omega \,\middle|\, \forall j \in [n] \setminus \{i\}, z_j = x_j\right\}$. It's not hard to see that it suffices to show the case $\|x - y\|_0 = 1$.

Assume that x, y only differ at the coordinate $i \in [n]$, *i.e.*, $x_i \neq y_i$ and $x_j = y_j$ for every $j \in [n] \setminus \{i\}$. Then by definition,

$$\begin{aligned}
P(x,y) &= \mathbf{E}_{\Theta_{n-1}}\left[\frac{\mu_{n-1}(x)\mu_{n-1}(y)}{\mu(x)}\right] \\
&= \sum_{j\in[n]} \frac{1}{n}\mathbf{E}\left[\frac{\mu_{n-1}(x)\mu_{n-1}(y)}{\mu(x)} \;\middle|\; k_n = j\right] \\
&= \frac{1}{n}\mathbf{E}\left[\frac{\mu_{n-1}(x)\mu_{n-1}(y)}{\mu(x)} \;\middle|\; k_n = i\right] \\
&= \frac{1}{n}\mathbf{Pr}\left[\mathrm{supp}(\mu_{n-1}) = L_{x,i}\right]\mathbf{E}\left[\frac{\mu_{n-1}(x)\mu_{n-1}(y)}{\mu(x)} \;\middle|\; k_n = i, \mathrm{supp}(\mu_{n-1}) = L_{x,i}\right] \\
&= \frac{1}{n}\frac{\mu(L_{x,i})\mu(x)\mu(y)}{\mu(x)\mu(L_{x,i})^2} \\
&= \frac{1}{n}\frac{\mu(y)}{\mu(L_{x,i})}.
\end{aligned}$$

When $\|x-y\|_0 \geq 2$, it is easy to see $P(x,y) = \mathrm{P}^{\mathrm{GD}}(x,y) = 0$. Thus we conclude the statement. □

Remark 2.3. When $\tau = n-\ell$, the corresponding Markov kernel associated with the coordinate-by-coordinate localization process and τ is the ℓ-uniform block dynamics $\mathrm{P}^{\ell-\mathrm{GD}}$.

2.1.1 The coordinate-by-coordinate localization process as a linear-tilt process

In this part, we will show the coordinate-by-coordinate localization process $(\mu_t)_{t\geq 0}$ is a linear-tilt localization process. Fix a probability measure μ on $\Omega = \{-1,+1\}^n$. We pick a permutation $k_1,\dots,k_n$ of $[n]$ uniformly at random. Let $U_1,\dots,U_n$ be independent random variables uniformly distributed in $[-1,+1]$.

Let $\mu_0 = \mu$. For $i = 0,1,\dots,n$, we define

$$\mu_{i+1}(x) = \mu_i(x)\left(1 + \langle x - \mathbf{b}(\mu_i), Z_i\rangle\right), \quad \forall x \in \Omega$$

where Z_i is a $\sigma(\mu_0,\dots,\mu_i)$-measurable random variable defined as

$$Z_i := \mathbf{e}_{k_{i+1}} \times \begin{cases} \frac{1}{1+\mathbf{b}(\mu_i)_{k_{i+1}}} & \mathbf{b}(\mu_i)_{k_{i+1}} \geq U_{i+1}, \\ \frac{-1}{1-\mathbf{b}(\mu_i)_{k_{i+1}}} & \mathbf{b}(\mu_i)_{k_{i+1}} \leq U_{i+1}, \end{cases}$$

where $(\mathbf{e}_1,\dots,\mathbf{e}_n)$ is the standard basis of $\mathbb{R}^n$.

It is not hard to see $\mathbf{E}\left[Z_i \mid \mu_i\right] = 0$, and

$$\begin{aligned}
\mu_{i+1}(\Omega) &= \int_{x\in\Omega} \mathrm{d}\mu_{i+1}(x) \\
&= \int_{x\in\Omega} \left(1 + \langle x - \mathbf{b}(\mu_i), Z_i\rangle\right)\,\mathrm{d}\mu_i(x) \\
&= \mu_i(\Omega) + \left\langle \int_{x\in\Omega} (x - \mathbf{b}(\mu_i))\,\mathrm{d}\mu_i(x), Z_i\right\rangle \\
&= \mu_i(\Omega)
\end{aligned}$$

meaning that $\mu_i(\Omega) = 1$ for each $i \in [n]$. To show μ_{i+1} is a pinning of μ_i, firstly note that the marginal distribution of the coordinate k_{i+1} is

$$\mathbf{Pr}_{X\sim\mu_t}\left[X_{k_{i+1}} = 1\right] = \frac{1+\mathbf{b}(\mu_i)_{k_{i+1}}}{2}, \quad \mathbf{Pr}_{X\sim\mu_t}\left[X_{k_{i+1}} = 1\right] = \frac{1-\mathbf{b}(\mu_i)_{k_{i+1}}}{2}.$$

By the definition of Z_i, when x is not identical to the pinned value, the inner product will be -1 and the probability will vanish.

2.2 Stochastic localization schemes driven by standard Brownian motion

Now we introduce a kind of linear-tilt localization scheme named the *stochastic localization scheme* firstly constructed by Eldan [Eld13]. Fix a probability measure μ on $\Omega \subseteq \mathbb{R}^n$. Let $(B_t)_{t\geq 0}$ be the standard Brownian motion in $\mathbb{R}^n$ adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. Let $(C_t)_{t\geq 0}$ be a stochastic process adapted to $(\mathcal{F}_t)_{t\geq 0}$ taking values in $n \times n$ positive semidefinite matrices. We define a measure-valued stochastic process $(\mu_t)_{t\geq 0}$ by $\frac{\mathrm{d}\mu_t}{\mathrm{d}\mu}(x) = F_t(x)$ as,

$$F_0(x) = 1, \quad \mathrm{d}F_t(x) = F_t(x)\left\langle x - \mathbf{b}(\mu_t), C_t \, \mathrm{d}B_t \right\rangle, \qquad \forall x \in \Omega. \tag{3}$$

Proposition 2.4. *If $\int_{t=0}^{\infty} C_t^2 \, \mathrm{d}t = \infty$, then $(\mu_t)_{t\geq 0}$ is a localization process. Moreover,*

$$\frac{\mathrm{d}\mu_t}{\mathrm{d}\mu_t}(x) = F_t(x) = \frac{1}{Z_t}\exp\left(-\frac{1}{2}\left\langle \Sigma_t x, x\right\rangle + \left\langle \mathbf{y}_t, x\right\rangle\right)$$

where Z_t is a normalizing factor to ensure that $\int_{x\in\Omega} F_t(x)\, \mathrm{d}\mu(x) = 1$ and $(\Sigma_t)_{t\geq 0}, (\mathbf{y}_t)_{t\geq 0}$ are stochastic processes adapted to $\mathcal{F}_t$ in the form of

$$\mathrm{d}\mathbf{y}_t = C_t \, \mathrm{d}B_t + C_t^2 \mathbf{b}(\mu_t)\, \mathrm{d}t, \quad \mathrm{d}\Sigma_t = C_t^2 \, \mathrm{d}t.$$

Proof. We prove the proposition by solving (3). Consider the stochastic process $(\log F_t(x))_{t\geq 0}$. By Itô's formula,

$$\begin{aligned}
\mathrm{d}\log F_t(x) &= \frac{\mathrm{d}F_t(x)}{F_t(x)} - \frac{\mathrm{d}[F(x)]_t}{2F_t(x)^2} \\
&= \left\langle x - \mathbf{b}(\mu_t), C_t \, \mathrm{d}B_t\right\rangle - \frac{1}{2}\|C_t(x - \mathbf{b}(\mu_t))\|_2^2 \, \mathrm{d}t.
\end{aligned}$$

This leads to the form

$$F_t(x) = \frac{1}{Z_t}\exp\left(-\frac{1}{2}\left\langle \Sigma_t x, x\right\rangle + \left\langle \mathbf{y}_t, x\right\rangle\right)$$

where $Z_t, \Sigma_t, \mathbf{y}_t$ are described as the proposition. Also we know $\mu_t(x) \geq 0$ for every $x \in \Omega$. By definition,

$$\begin{aligned}
\mathrm{d}\mu_t(\Omega) = &\ \mathrm{d}\int_{x\in\Omega} \mathrm{d}\mu_t(x) \\
&= \int_{x\in\Omega} F_t(x)\left\langle x - \mathbf{b}(\mu_t), C_t \, \mathrm{d}B_t\right\rangle \, \mathrm{d}\mu(x) \\
&= \left\langle \int_{x\in\Omega} (x - \mathbf{b}(\mu_t))\, \mathrm{d}\mu_t(x), C_t \, \mathrm{d}B_t \right\rangle \\
&= 0.
\end{aligned}$$

Then we know $\mu_t(\Omega) = 1$ for every $t \geq 0$ almost surely. Thus we know μ_t is almost surely a probability measure on Ω. The martingality comes directly from the definition, and to see the convergence of the process, note that when $\Sigma_t \to \infty$, by the form of F_t it will be a Dirac measure. □

When $C_t \equiv Q^{-1/2}$, we know the law of $\mathbf{y}_t$ by El Alaoui and Montanari [EAM22].

Theorem 2.5 ([EAM22])**.** *Fix a probability measure μ on Ω and a positive semidefinite matrix $Q \in \mathbb{R}^{n\times n}$. Let $(\mu_t)_{t\geq 0}$ be a stochastic localization process starting from μ driven by $C_t \equiv Q^{-1/2}$. Define the stochastic process $(\Sigma_t)_{t\geq 0}, (\mathbf{y}_t)_{t\geq 0}$ as above. Then*

$$\Sigma_t = tQ^{-1}, \ \mathbf{y}_t/t \sim \mu * \mathcal{N}(0, \Sigma_t), \qquad \forall t \geq 0.$$

2.3 Variance contraction via linear-tilt localization processes

Now we show how to bound the spectral gap of the Glauber dynamics P^{GD}. The following property named the *variance conservation* is the key in our analysis.

Definition 2.6 (Variance Conservation - Discrete)**.** Given a time-discrete localization process $(\mu_t)_{t\in\mathbb{N}}$ on Ω satisfying $(\kappa_1, \kappa_2, \ldots)$-*variance conservation* up to time $t \in \mathbb{N}$, if for every test function $f : \Omega \to \mathbb{R}$,

$$\mathbf{E}\left[\mathbf{Var}_{\mu_i}[f] \,\middle|\, \mu_{i-1}\right] \ge (1-\kappa_i)\mathbf{Var}_{\mu_{i-1}}[f], \quad \forall 1 \le i \le t.$$

Proposition 2.7. *Let $(\mu_t)_{t\in\mathbb{N}}$ be a time-discrete localization process on Ω satisfying $(\kappa_1, \kappa_2, \ldots)$-variance conservation up to time $t \in \mathbb{N}$. Let P be the random walk associated with $(\mu_t)_{t\in\mathbb{N}}$ and time t. Then its spectral gap* Gap(P) *satisfies*

$$\mathrm{Gap}(P) \ge \prod_{i=1}^{t}(1-\kappa_i).$$

Proof. By Proposition 1.6, it suffices to show for every test function $f : \Omega \to \mathbb{R}^n$,

$$\frac{\mathbf{E}_{\Theta_t}\left[\mathbf{Var}_{\mu_t}[f]\right]}{\mathbf{Var}_{\mu}[f]} \ge \prod_{i=1}^{t}(1-\kappa_i).$$

Note that $\mu_0 = \mu$. Then by direct calculation,

$$\begin{aligned}
\frac{\mathbf{E}_{\Theta_t}\left[\mathbf{Var}_{\mu_t}[f]\right]}{\mathbf{Var}_{\mu}[f]} &= \mathbf{E}_{\Theta_t}\left[\frac{\mathbf{Var}_{\mu_t}[f]}{\mathbf{Var}_{\mu_0}[f]}\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[\ldots\mathbf{E}\left[\frac{\mathbf{Var}_{\mu_t}[f]}{\mathbf{Var}_{\mu_{t-1}}[f]} \,\middle|\, \mu_{t-1}\right]\ldots\right]\frac{\mathbf{Var}_{\mu_1}[f]}{\mathbf{Var}_{\mu_0}[f]} \,\middle|\, \mu_0\right] \\
&\ge \prod_{i=1}^{t}(1-\kappa_i)
\end{aligned}$$

where the last inequality holds from Definition 2.6. ◻

Now it's time for us to show the variance contraction for a linear-tilt localization process $(\mu_t)_{t\in\mathbb{N}}$. The first step is to show the form of the evolution of its variance.

Lemma 2.8. *Let $(\mu_t)_{t\in\mathbb{N}}$ be a time-discrete linear-tilt localization process and $(Z_t)_{t\in\mathbb{N}}$ be its driving factor. Then for every test function $f : \Omega \to \mathbb{R}$ and $t \in \mathbb{N}$,*

$$\mathbf{E}\left[\mathbf{Var}_{\mu_{t+1}}[f] \,\middle|\, \mu_t\right] = \mathbf{Var}_{\mu_t}[f] - \langle V_t, C_t V_t\rangle$$

where

$$V_t := \int_{x\in\Omega}(x - \mathbf{b}(\mu_t))f(x)\,\mathrm{d}\mu_t(x),\ C_t := \mathbf{Cov}\left(Z_t \mid \mu_t\right).$$

Proof. Fix a test function $f : \Omega \to \mathbb{R}$. By direct calculation,

$$\begin{aligned}
\mathbf{E}\left[\mathbf{Var}_{\mu_{t+1}}[f] \,\middle|\, \mu_t\right] &= \mathbf{E}\left[\int_\Omega f(x)^2\,\mathrm{d}\mu_{t+1}(x) - \left(\int_\Omega f(x)\,\mathrm{d}\mu_{t+1}(x)\right)^2 \,\middle|\, \mu_t\right] \\
&= \int_\Omega f(x)^2\,\mathrm{d}\mu_t(x) - \mathbf{E}\left[\left(\int_\Omega f(x)\left(1 + \langle x - \mathbf{b}(\mu_t), Z_t\rangle\right)\,\mathrm{d}\mu_t(x)\right)^2 \,\middle|\, \mu_t\right] \\
&= \int_\Omega f(x)^2\,\mathrm{d}\mu_t(x) - \left(\int_\Omega f(x)\,\mathrm{d}\mu_t(x)\right)^2 - \mathbf{E}\left[\left(\int_\Omega f(x)\langle x - \mathbf{b}(\mu_t), Z_t\rangle\,\mathrm{d}\mu_t(x)\right)^2 \,\middle|\, \mu_t\right] \\
&= \mathbf{Var}_{\mu_t}[f] - \mathbf{E}\left[\langle V_t, Z_t\rangle^2 \,\middle|\, \mu_t\right] \\
&= \mathbf{Var}_{\mu_t}[f] - V_t^\top\mathbf{E}\left[Z_t^\top Z_t \,\middle|\, \mu_t\right]V_t \\
&= \mathbf{Var}_{\mu_t}[f] - V_t^\top C_t V_t.
\end{aligned}$$

□

Proposition 2.9. *Let $(\mu_t)_{t\in\mathbb{N}}$ be a time-discrete linear-tilt localization process and $(Z_t)_{t\in\mathbb{N}}$ be its driving factor. Then $(\mu_t)_{t\in\mathbb{N}}$ satisfies $(\kappa_1, \kappa_2, \ldots)$-variance conservation where*

$$\kappa_{t+1} = 1 - \left\| C_t^{1/2} \mathbf{Cov}\left(\mu_t\right) C_t^{1/2} \right\|_{\mathrm{OP}}, \quad \forall t \in \mathbb{N}.$$

Proof. Firstly it is not hard to see that it suffices to show the case $\mathbf{E}_\mu[f] = \mathbf{E}_{\mu_t}[f] = 0$. By Lemma 2.8, we only need to bound the term $\langle V_t, C_t V_t \rangle$. By definition,

$$\begin{aligned}
\langle V_t, C_t V_t \rangle &= \left\| C_t^{1/2} V_t \right\|_2^2 \\
&= \sup_{\theta : \|\theta\|_2 = 1} \left\langle C_t^{1/2} V_t, \theta \right\rangle^2 \\
&= \sup_{\theta : \|\theta\|_2 = 1} \left(\int_\Omega \langle C_t (x - \mathbf{b}(\mu_t)), \theta \rangle f(x) \, \mathrm{d}\mu_t(x) \right)^2 \\
&\le \sup_{\theta : \|\theta\|_2 = 1} \mathbf{Var}_{\mu_t}[f] \int_\Omega \langle C_t (x - \mathbf{b}(\mu_t)), \theta \rangle^2 f(x) \, \mathrm{d}\mu_t(x) \\
&= \left\| C_t^{1/2} \mathbf{Cov}\left(\mu_t\right) C_t^{1/2} \right\|_{\mathrm{OP}} \mathbf{Var}_{\mu_t}[f]
\end{aligned}$$

where the inequality is held by the Cauchy-Schwarz inequality. □

2.3.1 Variance conservation via the coordinate-by-coordinate localization process

Now we show the main result of rapid mixing via the spectral independence by Anari, Liu and Oveis Gharan [ALOG20].

Lemma 2.10. *Fix a disbribution μ on $\Omega \subseteq \{-1, +1\}^n$. Let $(\mu_t)_{t\in\mathbb{N}}$ be a coordinate-by-coordinate localization process starting from μ. Then $(\mu_t)_{t\in\mathbb{N}}$ satisfies $(\kappa_1, \kappa_2, \ldots)$-variance conservation up to n such that*

$$\kappa_{t+1} = 1 - \frac{\|\mathbf{Cor}\left(\mu_t\right)\|_{\mathrm{OP}}}{n - t}, \quad \forall 0 \le t < n$$

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

Proof. By Proposition 2.9, it suffices to show

$$C_t^{1/2} \mathbf{Cov}\left(\mu_t\right) C_t^{1/2} = \frac{\mathbf{Cor}\left(\mu_t\right)}{n - t}.$$

By direct calculation, for every unpinned $i \in [n]$,

$$\begin{aligned}
C_t(i, i) &= \mathbf{Cov}\left(Z_t \mid \mu_t\right)_{i,i} \\
&= \frac{1}{n - t} \frac{1}{1 - \mathbf{b}(\mu_t)_i^2} \\
&= \frac{1}{(n - t) \mathbf{Cov}\left(\mu_t\right)_{i,i}}.
\end{aligned}$$

Then the identity holds. □

Since we have already know $\left\|\Psi_{\mu_t}\right\|_{\mathrm{OP}} = \|\mathbf{Cor}\left(\mu_t\right)\|_{\mathrm{OP}}$, we can establish the result of [ALOG20].

Lemma 2.11 (A Reformulation of the Main Result in [ALOG20])**.** *Given an $(\eta_0, \ldots, \eta_n)$-spectrally independent Gibbs distribution μ of some hardcore model over the state space $\Omega \subseteq \{-1, +1\}^n$, the spectral gap of the ℓ-uniform block dynamics is at least*

$$\mathrm{Gap}(\mathrm{P}^{\ell-\mathrm{GD}}) \ge \prod_{t=0}^{n-\ell-1} \left(1 - \frac{\eta_t}{n - t}\right).$$

3 Entropic Contraction

To show the entropic decay of Markov dynamics associated with a localization scheme, we consider the notions named *entropic stability* put in Anari et al. [AJK+21].

For a probability measure μ on $\Omega \subseteq \mathbb{R}^n$ and a vector $v \in \mathbb{R}^n$, we define the *exponential tilt* $\mathcal{T}_v\mu$ of μ as

$$\frac{\mathrm{d}\mathcal{T}_v\mu}{\mathrm{d}\mu}(x) := \frac{e^{\langle v,x\rangle}}{\int_{y\in\Omega} e^{\langle v,y\rangle}\,\mathrm{d}\mu(y)}, \quad \forall x \in \Omega.$$

Definition 3.1 (Entropic Stability)**.** For a probability measure μ on $\Omega \subseteq \mathbb{R}^n$, a function $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ and $\alpha > 0$, we say that μ is α-entropically stable with respect to ψ if

$$\psi(\mathbf{b}(\mathcal{T}_v\mu), \mathbf{b}(\mu)) \leq \alpha \mathcal{D}_{\mathrm{KL}}\left(\mathcal{T}_v\mu \parallel \mu\right), \quad \forall v \in \mathbb{R}^n.$$

Remark 3.2. The function ψ is usually a kind of distance on $\mathbb{R}^n$.

By the principle of maximum entropy, the following corollary comes immediately.

Corollary 3.3. *Suppose that a probability measure μ on Ω is α-entropically stable with respect to ψ. Then for every probability ρ that is absolutely continuous with respect to μ, it holds that*

$$\psi(\mathbf{b}(\rho), \mathbf{b}(\mu)) \leq \alpha \mathcal{D}_{\mathrm{KL}}\left(\rho \parallel \mu\right).$$

3.1 Entropic contraction via stochastic localization

Assuming the entropic stability of the probability measure μ on $\Omega \subseteq \mathbb{R}^n$, we can conveniently show the entropic contraction via stochastic localization processes.

Proposition 3.4. *Suppose that μ is a probability measure on $\Omega \subseteq \mathbb{R}^n$ and $(\mu_t)_{t\geq 0}$ is a stochastic localization process driven by $(C_t)_{t\geq 0}$ starting from $\mu_0 = \mu$. Fix a real $T > 0$. Assume that, almost surely for every $t \in [0,T]$, μ_t is α_t-entropically stable with respect to $\psi(x,y) = \frac{1}{2}\|C_t(x-y)\|^2$. Then we have the following approximate entropic conservation bound*

$$\mathbf{E}\left[\mathbf{Ent}_{\mu_T}\left[f\right]\right] \geq \exp\left(-\int_0^T \alpha_t \,\mathrm{d}t\right)\mathbf{Ent}_{\mu}\left[f\right]$$

for every function $f : \Omega \to \mathbb{R}_{>0}$.

Proof. Fix a test function $f : \Omega \to \mathbb{R}_{>0}$. For every $t \geq 0$, define a probability measure ρ_t absolutely continuous with respect to μ_t as

$$\frac{\mathrm{d}\rho_t}{\mathrm{d}\mu_t}(x) := \frac{f(x)}{\int_{y\in\Omega} f(y)\,\mathrm{d}\mu_t(y)}, \quad \forall x \in \Omega.$$

Consider the martingale $(M_t)_{t\geq 0}$:

$$M_t := \mu_t(f) = \int_{x\in\Omega} f(x)F_t(x)\,\mathrm{d}\mu(x).$$

Then by elementary calculation,

$$\begin{aligned}\mathrm{d}M_t &= \int_{x\in\Omega} f(x)F_t(x)\left\langle x - \mathbf{b}(\mu_t), C_t\,\mathrm{d}B_t\right\rangle\,\mathrm{d}\mu(x)\\ &= M_t\left\langle C_t(\mathbf{b}(\rho_t) - \mathbf{b}(\mu_t)),\ \mathrm{d}B_t\right\rangle.\end{aligned}$$

Then by Itô's formula,

$$\mathrm{d}M_t \log M_t = \frac{1}{2} M_t \|C_t(\mathbf{b}(\rho_t) - \mathbf{b}(\mu_t))\|^2 \, \mathrm{d}t + \mathrm{d}(\text{martingale}).$$

Then we finally know

$$\mathrm{d}\mathbf{Ent}_{\mu_t}[f] = -\frac{1}{2} M_t \|C_t(\mathbf{b}(\rho_t) - \mathbf{b}(\mu_t))\|^2 \, \mathrm{d}t + \mathrm{d}(\text{martingale}).$$

Since μ_t is α_t-entropically stable with respect to $\psi(x, y) = \frac{1}{2}\|C_t(x - y)\|^2$, it holds that

$$\mathrm{d}\mathbf{Ent}_{\mu_t}[f] \geq -\alpha_t \mathbf{Ent}_{\mu_t}[f] \, \mathrm{d}t + \mathrm{d}(\text{martingale}).$$

Then we consider the process $\left(\exp\left(\int_0^s \alpha_s \, \mathrm{d}s\right) \mathbf{Ent}_{\mu_t}[f]\right)_{t \geq 0}$. By Itô's formula,

$$\begin{aligned}
\mathrm{d}\left(e^{\int_0^t \alpha_s \, \mathrm{d}s} \mathbf{Ent}_{\mu_t}[f]\right) &= -\frac{1}{2} e^{\int_0^t \alpha_s \, \mathrm{d}s} M_t \|C_t(\mathbf{b}(\rho_t) - \mathbf{b}(\mu_t))\|^2 \, \mathrm{d}t + \alpha_t e^{\int_0^t \alpha_s \, \mathrm{d}s} \mathbf{Ent}_{\mu_t}[f] \, \mathrm{d}t + \mathrm{d}\,(\text{martingale}) \\
&\geq -\alpha_t e^{\int_0^t \alpha_s \, \mathrm{d}s} \mathbf{Ent}_{\mu_t}[f] \, \mathrm{d}t + \alpha_t e^{\int_0^t \alpha_s \, \mathrm{d}s} \mathbf{Ent}_{\mu_t}[f] \, \mathrm{d}t + \mathrm{d}\,(\text{martingale}) \\
&= \mathrm{d}\,(\text{martingale}).
\end{aligned}$$

which implies $e^{\int_0^t \alpha_s \, \mathrm{d}s} \mathbf{Ent}_{\mu_t}[f]$ is a sub-martingale. Thus we conclude

$$\mathbf{E}\left[\mathbf{Ent}_{\mu_T}[f]\right] \geq \exp\left(-\int_0^T \alpha_t \, \mathrm{d}t\right) \mathbf{Ent}_{\mu}[f].$$

□

4 Annealing via Localization Schemes

To utilize the power of localization schemes, we show how to anneal via two localization schemes.

Definition 4.1 (Concatenation of Localization Schemes)**.** Given two localization schemes $\mathcal{L}_i, \mathcal{L}_f$ on a space Ω, a probability measure μ on Ω and a real $T > 0$, define the localization process $(\mu_t)_{t \geq 0}$ in the *concatenation of* $\mathcal{L}_i, \mathcal{L}_f$ associated with μ at time T as

$$\mu_t := \begin{cases} \mu_t^{(i)} & t \leq T \\ \mu_{t-T}^{(f)} & t \geq T \end{cases} \quad \forall t \geq 0$$

where $\left(\mu_t^{(i)}\right)_{t \geq 0}$ is the localization process obtained by applying $\mathcal{L}_i$ to μ and $\left(\mu_t^{(f)}\right)_{t \geq 0}$ is the localization process obtained by applying $\mathcal{L}_f$ to $\mu_T^{(i)}$. Such a localization scheme is denoted by $\mathrm{concate}(\mathcal{L}_i, \mathcal{L}_f, T)$.

Lemma 4.2 (Variance Contraction via Annealing)**.** *Under the above settings, fix a time $\tau > 0$ additionally. Let $P^t := P^{(\mathcal{L}_f, \tau)}(\mu_t^{(i)})$ for every $t > 0$. If*

- *For every function $f : \Omega \to \mathbb{R}$,*

$$\mathbf{E}\left[\mathbf{Var}_{\mu_T}[f]\right] \geq \varepsilon \mathbf{Var}_{\mu}[f].$$

- *Almost surely we have*

$$\mathrm{Gap}(P^T) \geq \delta.$$

Then almost surely it holds that

$$\mathrm{Gap}(P^0) \geq \varepsilon\delta.$$

Similarly, we have an entropic version of Lemma 4.2.

Lemma 4.3 (Entropic Contraction via Annealing)**.** *Under the above settings, fix a time $\tau > 0$ additionally. Let $P^t := P^{(\mathcal{L}_f,\tau)}(\mu_t^{(i)})$ for every $t > 0$. If*

- *For every function $f : \Omega \to \mathbb{R}_{>0}$,*

$$\mathbf{E}\left[\mathbf{Ent}_{\mu_T}[f]\right] \geq \varepsilon \mathbf{Ent}_{\mu}[f].$$

- *For every function $f : \Omega \to \mathbb{R}_{>0}$,*

$$\mathbf{E}\left[\mathbf{Ent}_{\mu_{T+\tau}}[f] \,\middle|\, \mu_T\right] \geq \delta \mathbf{Ent}_{\mu_T}[f].$$

Then almost surely it holds that

$$\rho_{\mathrm{LS}}(P^0) \geq \varepsilon\delta.$$

The proof of Lemmas 4.2 and 4.3 is the observation that the Dirichlet form is a super-martingale. The details of this observation can be found in [CE22] and for simplicity we omit here.

References

[AJK+21] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic Independence I: Modified Log-Sobolev Inequalities for Fractionally Log-Concave Distributions and High-Temperature Ising Models, 2021. 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. 8

[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. 1, 11

[EAM22] Ahmed El Alaoui and Andrea Montanari. An Information-Theoretic View of Stochastic Localization. *IEEE Transactions on Information Theory*, 68(11):7423–7426, 2022. 6

[Eld13] Ronen Eldan. Thin Shell Implies Spectral Gap Up to Polylog via a Stochastic Localization Scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013. 6

[Mon23] Andrea Montanari. Sampling, Diffusions, and Stochastic Localization, 2023. 1