Sampling: stochastic localization, parallel sampling, auxiliary method

Ruihan Xu

August 11, 2025

1 Introduction

1.1 stochastic localization for KLS conjecture

The stochastic Localization scheme was initially developed by Eldan [7] to construct a series of probability densities evolving from a known $f(x)$ that can be utilized in proving a $poly(\log n)$ growth of the Cheeger constant, defined by

$$G_n := (\inf_{\mu} \inf_{A \subset \mathbb{R}^n} \frac{\mu^+(A)}{\mu(A)})^{-1}$$

where μ runs over all isotropic log-concave measures in $\mathbb{R}^n$ and A runs over all Borel sets with $\mu(A) \leq \frac{1}{2}$.

To prove $G_n = poly(\log n)$ for universal $f(x)$, we first verify the result on the martingale $\{f_t\}$, which has better and better log-concavity as t evolves and thus is easier to analyze, and use the martingale condition to transfer the result back to $f(x)$. The constructed measure $f_t(x)$ has some good properties:

1. Martingale $\{f_t\}$ satisfies $f(x) = \mathbb{E}[f_t(x)]$ for all $x \in \mathbb{R}^n$ and t.
2. When t large enough and the covariance matrix A_t of f_t is properly bounded ($\|A_t\|_{op} < CK_n^2(\log n)e^{-ct}$, $K_n := \sup_\mu \| \int_{\mathbb{R}^n} x_1 x \otimes x d\mu(x)\|$ for isotropic log-concave μ), f_t can be expressed as
$$f_t(x) = \exp(-|\frac{x}{CK_n\sqrt{\log n}}|^2 - \phi_t(x)), \quad \phi_t \text{ is convex.}$$

With a well-known concentration inequality for certain log-concave functions and a theorem connecting the concentration with a bound of the Cheeger constant, the problem is almost done.

Proposition 1 *There exists a universal constant $\Theta > 0$ such that the following holds: Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $K > 0$. Suppose that*

$$d\mu(x) \propto e^{-\phi(x) - \frac{1}{2K^2}|x|^2} dx$$

is a probability measure whose barycenter lies at the origin. Then

1. *For all Borel set $A \subset \mathbb{R}^n$ with $0.1 < \mu(A) < 0.9$, one has $\mu(A_{K\Theta}) > 0.95$, where $A_{K\Theta}$ is the $K\Theta$-extension of A.*

2. *For all $\theta \in S^{n-1}$, one has $\int_{\mathbb{R}^n} \langle x, \theta \rangle^2 d\mu(x) \leq \Theta K^2$.*

Theorem 1 (E. Mailman) *Suppose a log-concave probability measure μ satisfies the following: there exist constants $0 < \lambda < \frac{1}{2}$ and $\Theta > 0$ such that for every measurable set $E \subset \mathbb{R}^n$ with $\mu(E) \geq \frac{1}{2}$ one has*

$$\mu\big(E_\Theta\big) \geq 1 - \lambda,$$

where E_Θ denotes the Θ-extension of E. Then μ obeys the isoperimetric inequality

$$\frac{\mu^+(E)}{\mu(E)} \geq \frac{1 - 2\lambda}{\Theta}.$$

The theorem indicates that if we can control $\mu(E_\Theta/E)$ from below, then $G_n = O(\frac{1}{\Theta})$, and the proposition indicates that $f_t(E_\Theta/E)$ is bounded from below with $\Theta = \frac{1}{K_n\sqrt{\log n}}$ for large t. Finally, since f_t is martingale, one can show that $f(E_\Theta/E)$ can be also bounded from below with $\Theta = \frac{1}{K_n\sqrt{\log n}}$, and the problem is solved.

2 Stochastic localization in the linear tilt form

2.1 linear tilt localization

Start from distribution $\nu_0 = \nu$, define in discrete time

$$\nu_{t+1}(x) = \nu_t(x)(1 + \langle x - b(\nu_t), Z_t \rangle)$$

where $b(\nu_t)$ is the barycenter/expectation of ν_t and Z_t is a random variable such that $\mathbb{E}[Z_t|\nu_t] = 0$ and guarantees that ν_{t+1} is a probability measure. If we let the time interval go to 0 and set $Z_t = C_t dW_t$, then we have the formulation of the continuous setting: $\frac{\nu_t(x)}{\nu(x)} = F_t(x)$

$$dF_t(x) = F_t(x)\langle x - b(\nu_t), C_t dW_t \rangle.$$

2.2 dynamic of $a_t = b(\nu_t)$ and $A_t = Cov(\nu_t)$

[3, 1, 4]Given linear tilt localization process as above, we have

$$\begin{aligned} da_t &= A_t C_t dW_t \\ dA_t &= -A_t C_t^2 A_t dt + \underbrace{\int (x - a_t)(x - a_t)^\top dF_t(x) v(x) dx}_{martingale} \end{aligned}$$

A further calculation gives

$$A_t = \mathbb{E}[A_T|\mathcal{F}_t] + \int_t^T \mathbb{E}[A_s C_s^2 A_s|\mathcal{F}_t] ds$$

3 Sampling in a stochastic localization framework

The framework of stochastic localization [4, 6]

3.1 sampling via the stochastic localization process

The first type of sampling method stems from simulating the localization process [8, 10], and the intuition is that as time t evolves, the mixing time (spectral gap) of distribution ν_t is getting better and better.

[5, 8]We can generate the information process as

$$Y_t = \alpha(t)X + \sigma W_t$$

for $\alpha = t^{0.5}g(t)$ where $g(t)$ strictly increases to infinity as t reaches terminating time. The process Y_t obeys the SDE

$$\begin{aligned} &dY_t = \alpha'(t)u_t(Y_t)dt + \sigma dB_t, \quad u_t(y) = E_{X\sim q_t(\cdot|y)}[X] = \int xq_t(x|y)dx \\ &q_t(x|y) \propto \pi(x)N(y/\alpha(t), \sigma^2/g(t)^2 I_d) \end{aligned}$$

Stochastic Localization via Iterative Posterior Sampling (SLIPS) iteratively samples Y_t starting from a initialization distribution $p_{t_0}(y)$ until the terminating time T, where $Y_T/\alpha(T)$ has the same distribution as X. To estimate score $u_t(Y_t)$ in each time step t, SLIPS uses another Unadjusted Langevin Algorithm (ULA) to sample $\{X_t^j\}_{j=1}^m$ according to $q_t(x|y)$. The same derivation can be implemented in discrete distribution [10].

3.2 Markov chain associated to stochastic localization

The second type of sampling can be viewed as a latent variable method $\nu(x) = \nu_\theta(x)\pi(\theta)$, the underlying essence also indicates the key to the KLS conjecture. We are striving for a balance between the concavity of conditional distribution $\nu_\theta(x)$ for every single θ and the consistency between ν_θ and ν (in some context it is entropy contraction).

Consider the localization scheme $\{\nu_t\}$ (martingale) corresponding to ν. The markov chain associated to $\{\nu_t\}$ at time τ is defined by

$$P_{x\to A} = \mathbb{E}[\frac{\nu_\tau(x)\nu_\tau(A)}{\nu(x)}].$$

It can be interpreted in another way: Consider X, Y are two independent variable drawn from given distribution ν_τ, thus $P(X \in A, Y \in B) = \mathbb{E}[\nu_\tau(A)\nu_\tau(B)]$. Then

$$P_{x\to A} = P(Y \in A|X = x).$$

3.2.1 examples of localization scheme

Glauber Dynamics The Glauber dynamics has the transition kernel

$$P^{GD}_{x\to y}(\nu) = \frac{1}{n}\frac{\nu(y)}{\nu(x)+\nu(y)}$$

3.2.2 spectral gap guarantee

For $P^{(\mathcal{L},\tau)}(\nu)$, which is the markov transition kernel associated to localization process $\mathcal{L}$ from distribution ν at time τ, we can bound its spectral gap (MLSI) by variance (entropy) conservation

$$gap(P^{(\mathcal{L},\tau)}) = \inf \frac{\mathbb{E}[Var_{\nu_\tau}[\phi]]}{Var_\nu[\phi]}$$
$$\rho_{LS}(P^{(\mathcal{L},\tau)}) \geq \inf \frac{\mathbb{E}[Ent_{\nu_\tau}[\phi]]}{Ent_\nu[\phi]}.$$

It is natural to first work on $\frac{\mathbb{E}[Var_{\nu_{t+1}}[\phi]|\nu_t]}{Var_{\nu_t}[\phi]}$. We consider a linear-tilt localization

$$\nu_{t+1}(x) = \nu_t(x)(1+\langle x-b(\nu_t), Z_t\rangle)$$

with $\mathbb{E}[Z_t|\nu_t]=0$, where $b(\nu)$ is the expectation/barycenter of ν. It's easy to verify that $\{\nu_t(x)\}_t$ is a martingale. Then the variance conservation has something to do with the random tilt Z_t:

Lemma 1 *For a test function ϕ, we have*

$$\mathbb{E}[Var_{\nu_{t+1}}(\phi)|\nu_t] - Var_{\nu_t}(\phi) = -\langle v_t, C_t v_t\rangle$$
$$v_t := \int_\Omega (x-b(\nu_t))\phi(x)\nu_t(dx), \quad C_t := Cov(Z_t|\nu_t)$$

And furthermore we have

$$\frac{\mathbb{E}[Var_{\nu_{t+1}}(\phi)|\nu_t]}{Var_{\nu_t}(\phi)} \geq 1 - \|C_t^{1/2}Cov(\nu_t)C_t^{1/2}\|_{op}$$

Question: Is this bound useful in most of the case?
Remark: $C_t = constant \cdot Cov(\nu_t)^{-1}$ preserve the variance best.

In the lemma above, v_t represents exactly the covariance between $\phi(X)$ and X. More precisely, consider $\tilde{v}_t = Cov(\nu_t)^{-1/2}v_t = \int_\Omega Cov(\nu_t)^{-1/2}(x - b(\nu_t))\phi(x)\nu_t(dx) = \int_\Omega y\phi(Cov(\nu_t)^{1/2}y+b)\mu_t(dy) = \int_\Omega y\tilde{\phi}(y)\mu_t(dy)$, where $Y\sim\mu_t$ is the isotropic distribution obtained by rescaling ν_t, then $\langle v_t, C_t v_t\rangle = \langle \tilde{v}_t, \tilde{C}_t\tilde{\nu_t}\rangle$ for $\tilde{C}_t = Cov(\nu_t)^{1/2}C_tCov(\nu_t)^{1/2}$. Furthermore, note that $Var_{\nu_t}(\phi) = Var_{\mu_t}(\tilde{\phi})$.

That means to preserve variance as much as possible, we need to first control the variance of Z in each step and second align C_t with $Cov(\nu_t)^{-1}$.

What we have in practice

- Gaussian channel $Y_t = \alpha(t)X + \sigma B_t$ (stochastic localization via iterative sampling), $C_t = \frac{\alpha(t)'}{\sigma} I$.
- Glauber dynamics, $C_t = diag(Cov(\nu_t)^{-1})$, spectral gap $\Pi(1-\frac{\|diag(Cov(\nu_t))Cov(\nu_t)^{-1}\|}{n-t}$
- Coordinate Gibbs, spectral gap $\frac{1}{n}(1 - \|diag(Cov(\nu_t))Cov(\nu_t)^{-1}\|)$

3.2.3 annealing sampling

Interesting thing about log-concave measure We see in many cases that some assumptions are imposed on the target distribution μ with dimension d especially when it's not log-concave. The most prominent one is that we assume its tail distribution outside a ball with radius R behaves as a Gaussian with covariance $\tau^2 I$. The standard scaling of three factors should be

$$R \sim \sqrt{d}t.$$

And with that, we can also say something more about its Poincarè/log-Sobolev constant.Suppose the measure μ has a C^2 density$d\mu(x) = e^{-V(x)}\, dx$, and that the potential satisfies the uniform convexity condition $\nabla^2 V(x) \succeq t^{-2} I$ when $\|x\| \geq R$. Applying the Bakry–Émery criterion together with the Holley–Stroock bounded-perturbation lemma, one obtains

$$C_{LS}(\mu) \;\leq\; t^2\, e^{\operatorname{osc}_{B_R} V}, \qquad C_P(\mu) \;\leq\; C_{LS}(\mu), \tag{2.3}$$

where

$$\operatorname{osc}_{B_R} V \;:=\; \sup_{\|x\| \leq R} V(x) \;-\; \inf_{\|x\| \leq R} V(x).$$

Because $V(x) \approx \|x\|^2/(2t^2)$ in the vicinity of $\|x\| = R$, the lob-sobolev constant is bounded above by $t^2 e^{\frac{R^2}{2t^2}}$.

4 Ising model

4.1 introduction

Given graph (V, E) and space $\Omega = \{-1, 1\}^{|V|}$, the Hanmiltonian of Ising model is defined as $H(\sigma) = -\sum_E J_{i,j}\sigma_i\sigma_j - \sum_i h_i\sigma_i$. The Gibbs weight of state σ is $\mu(\sigma) \propto e^{-\beta H(\sigma)}$.

- $J_{i,j} \geq 0$ ferromagnetic
- $J_{i,j} \leq 0$ antiferromagnetic
- otherwise spin glass.

4.2 k-Glauber Dynamics

As the external field h grows to infinity, the spectral gap converges to $O(1)$.[10]

4.3 Swendsen Wang method

4.3.1 Algorithm

For a typical no external field case

1. Sampling $\omega_{i,j}$ for each $(i,j) \in E$ and $\sigma_i = \sigma_j$, $\mathbb{P}[\omega_{i,j} = 1] = p_{i,j} = 1 - e^{-2\beta J_{i,j}}$.

2. Sampling each cluster with $p = 0.5$ to get next σ.

When there is an external field Add a fake particle.

Antiferromagnetic Change the probability $p_{i,j}$

spin-glass model Other SW-based model

4.4 Spectral gap results

Trying to recover:

- For a complete graph G, assume $\beta J_{i,j} \sim \frac{\beta}{n}$, $\lambda_{SW} = O(1)$ when β small (high temp), $\lambda_{SW} = n^{-0.25}$ when a critical β [9] and $\lambda_{SW} = c(\beta)\log(n)$ when β is large [2].

4.5 Stochastic localization of SW algorithm in ferromagnetic model

4.5.1 Construction of localization scheme

Consider a ferromagnetic model with graph (V,E), $J_{i,j} \geq 0$ and $h = 0$, suppose that σ is sampled from $\mu(\sigma) \propto e^{-\beta \sum_{(i,j)} J_{i,j}\sigma_i\sigma_j}$. For each edge $e = e_{i,j} \in E$ of σ with $\sigma_i = \sigma_j$, we assign a exponential clock with intensity $\lambda_e = -\log(1-p_{i,j}) = 2\beta J_{i,j}$, and when it rings we connect σ_i and σ_j. We define the state of edge $e_{i,j}$ at time t as $w^t_{i,j}$;$w^t_{i,j} = 1$ if σ_i and σ_j are connected at time t and 0 otherwise. The stochastic localization can then be framed as $\mu_t(\cdot) = \mathbb{E}[1_{\{\sigma\in\cdot\}}|\mathcal{F}_t]$, where $\mathcal{F}_t$ is spanned by $w^s_{i,j}$ for $s \leq t$ and $(i,j) \in E$. Define $C_t = \{e = (i,j)|w^t_{i,j} = 1\}$ as the connected edges at time t.

Proposition 2 *$\{\mu_t\}_t$ is a stochastic localization and*

$$\begin{aligned}\mu_t(d\sigma) &\propto \mu(d\sigma)(\Pi_{e=(i,j)\in C_t} 1_{\{\sigma_i=\sigma_j\}})e^{-\sum_{e=(i,j)\notin C_t}\lambda_e t 1_{\{\sigma_i=\sigma_j\}}}\\ &\propto \mu(d\sigma)\Pi_{(i,j)}(1_{\{w^t_{i,j}=1\}}1_{\{\sigma_i=\sigma_j\}} + 1_{\{w^t_{i,j}=0\}}e^{-\lambda_e t 1_{\{\sigma_i=\sigma_j\}}})\\ &= \frac{\mu(d\sigma)W_t(\sigma,w^t)\kappa_t(w^t)}{\kappa_t(w^t)\sum_{\sigma'}\mu(\sigma')W_t(\sigma',w^t)}\end{aligned}$$

where $W_t(\sigma, w^t) = \Pi_{(i,j)}(1_{\{w^t_{i,j}=1\}}1_{\{\sigma_i=\sigma_j\}} + 1_{\{w^t_{i,j}=0\}}e^{-\lambda_e t 1_{\{\sigma_i=\sigma_j\}}})$, $\kappa_t(w^t) = \Pi_e(1-e^{-\lambda_e t})^{1_{\{w^t_e=1\}}}$ and $\sum_w \kappa_t(w)W_t(\sigma,w) = 1$ for any σ.

proof. By Bayes formula we have $\mu_t(\sigma|\mathcal{F}_t) = \mu_t(\sigma|w^t_e, \forall e\in E) \propto \mu(\sigma)(\Pi_{\{e|w^t_e=1\}}\mathbb{P}(w^t_e = 1|\sigma))(\Pi_{\{e|w^t_e=0,\sigma_i=\sigma_j\}}e^{-\lambda_e t})$. Since $\mathbb{P}(w^t_e=1|\sigma) = (1-e^{-\lambda_e t})1_{\{\sigma_i=\sigma_j\}}$. Combining the two equations gives

$$\mu_t \propto \mu(d\sigma)(\Pi_{e=(i,j)\in C_t}1_{\{\sigma_i=\sigma_j\}})e^{-\sum_{e=(i,j)\notin C_t}\lambda_e t 1_{\{\sigma_i=\sigma_j\}}}.$$

Note that for any $e=(i,j)$,

$$\sum_{w^t_{i,j}\in\{0,1\}}(1-e^{-\lambda_e t})^{1_{\{w^t_e=1\}}}(1_{\{w^t_{i,j}=1\}}1_{\{\sigma_i=\sigma_j\}} + 1_{\{w^t_{i,j}=0\}}e^{-\lambda_e t 1_{\{\sigma_i=\sigma_j\}}}) = 1.$$

Thus $\kappa_t(w)W_t(\sigma,w)$ is a density function of w regardless of σ.

Proposition 3 *Denote $\frac{d\mu_t}{d\mu} = F_t(\sigma)$. Then the process $\{F_t\}_t$ is driven by*

$$\partial_t F_t(\sigma) = -F_t(\sigma)\sum_{e\notin C_t}\lambda_e(I_e(\sigma)-\mu_t(I_e))dt + \sum_{e\notin C_t}F_t(\sigma)(\frac{I_e(\sigma)}{\mu_t(I_e)}-1)dw^t_e$$

where $I_e(\sigma) = 1_{\{\sigma_i=\sigma_j\}}$.

proof.

Lemma 2 *For a test function f, one has*

$$dVar_{\mu_t}(f) = -\sum_{e\notin C_t}\lambda_e\frac{Cov_{\mu_t}(f,I_e)^2}{\mu_t(I_e)}dt$$

proof. Given a small $h>0$, one has

$$\begin{aligned}&Var_{\mu_{t+h}}(f) - Var_{\mu_t}(f)\\ =&\mu_{t+h}(f^2) - \mu_t(f^2) - 2\mu_t(f)(\mu_{t+h}(f)-\mu_t(f)) - (\mu_{t+h}(f)-\mu_t(f))^2.\end{aligned}$$

Note that $\mathbb{E}[\mu_{t+h}(f)|\mu_t] - \mu_t(f) = o(h)$, then $\mathbb{E}[Var_{\mu_{t+h}}(f)|\mu_t] - Var_{\mu_t}(f) = \mathbb{E}[(\mu_{t+h}(f)-\mu_t(f))^2|\mu_t]$. Since $\mu_{t+h}(\cdot) = \mathbb{E}[1_{\{\sigma\in\cdot\}}|\mathcal{F}_{t+h}]$, we have $\mu_{t+h}(\cdot) =$

$\mu_t(\cdot|I_e(\cdot) = 1)$ if e is connected during $(t, t+h)$. Note that the edge e is connected with probability $\lambda_e \mu_t(I_e)h$, thus we have

$$\begin{aligned}\partial_t Var_{\mu_t}(f) &= \lim_{h\to 0} \frac{\mathbb{E}[Var_{\mu_{t+h}}(f)|\mu_t] - Var_{\mu_t}(f)}{h} \\ &= -\sum_{e\notin C_t} \lambda_e \mu_t(I_e)(\mu_t(f|I_e = 1) - \mu_t(f))^2 \\ &= -\sum_{e\notin C_t} \lambda_e \frac{Cov_{\mu_t}(f, I_e)^2}{\mu_t(I_e)}\end{aligned}$$

Lemma 3 *Under the same setting of $G = (V, E), J, \mu_t$, the following holds:*

- *The probability of edge e being chosen during time $(t, t+h)$ is $\lambda_e \mu_t(I_e)h + o(h)$.*
- $\mathbb{E}[\mu_{t+h}(f)|\mu_t] - \mu_t(f) = o(h)$.

proof.

Lemma 4 *A lower bound of sepctral gap λ_{SW}.*

proof. Define $\phi_e = \sqrt{\frac{\lambda_e}{\mu_t(I_e)}}(I_e - \mu_t(I_e))$, then

$$-\sum_{e\notin C_t} \lambda_e \frac{Cov_{\mu_t}(f, I_e)^2}{\mu_t(I_e)} = -\sum_{e\notin C_t} \langle \phi_e, f - \mu_t(f)\rangle^2 = -\langle f - \mu_t(f), K_t(f - \mu_t(f))\rangle$$

where $K_t = \sum_{e\notin C_t} \phi_e \phi_e^\top$. Thus we have

$$dVar_{\mu_t}(f) \geq -\|K_t\| \|f - \mu_t(f)\|^2_{\mu_t} = -\|K_t\| \cdot Var_{\mu_t}(f)$$

Since $\|K_t\| \leq (|E| - |C_t|)\lambda_{max}$ ($\|K_t\| \leq \sum_{e\notin C_t} \lambda_e$), we have

References

[1] N. Anari, F. Koehler, and T.-D. Vuong. Trickle-down in localization schemes and applications. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1094–1105, 2024.

[2] A. Blanca and Z. Song. Cutoff for the swendsen-wang dynamics on the complete graph. *arXiv preprint arXiv:2507.20482*, 2025.

[3] Y. Chen. An almost constant lower bound of the isoperimetric coefficient in the kls conjecture. *Geometric and Functional Analysis*, 31(1):34–61, 2021.

[4] Y. Chen and R. Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 110–122. IEEE, 2022.

[5] A. El Alaoui and A. Montanari. An information-theoretic view of stochastic localization. *IEEE Transactions on Information Theory*, 68(11):7423–7426, 2022.

[6] A. El Alaoui, A. Montanari, and M. Sellke. Sampling from the sherrington-kirkpatrick gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 323–334. IEEE, 2022.

[7] R. Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.

[8] L. Grenioux, M. Noble, M. Gabrié, and A. O. Durmus. Stochastic localization via iterative posterior sampling. *arXiv preprint arXiv:2402.10758*, 2024.

[9] Y. Long, A. Nachmias, W. Ning, and Y. Peres. *A power law of order 1/4 for critical mean field Swendsen-Wang dynamics*, volume 232. American Mathematical Society, 2014.

[10] C. Wang, K. Cui, W. Zhao, and T. Yu. Sampling from binary quadratic distributions via stochastic localization. *arXiv preprint arXiv:2505.19438*, 2025.