Sampling: stochastic localization, parallel sampling, auxiliary method

Ruihan Xu

July 22, 2025

1 Introduction

1.1 stochastic localization for KLS conjecture

The stochastic Localization scheme was initially developed by Eldan [6] 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 $f(x) = E[f_t(x)]$ 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 almostv 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 $\int_{\mathbb{R}^n} \langle x, \theta\rangle^2 d\mu(x) \le \Theta K^2$.*

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 $\mu(E) \ge \frac{1}{2}$ one has*

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

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

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

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 $E[Z_t|\nu_t] = 0$ 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 $A_t = Cov(\nu_t)$

[2, 1, 3]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

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

3 Sampling in a stochastic localization framework

The framework of stochastic localization [3, 5]

3.1 sampling via the stochastic localization process

The first type of sampling method stems from simulating the localization process [7, 8], and the intuition is that as time t evolves, the mixing time (spectral gap) of distribution ν_t is getting better and better.

[4, 7]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

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

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 [8].

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

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

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

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

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

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

with $E[Z_t|\nu_t] = 0$, 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*

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

And furthermore we have

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

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 $\tilde{C}_t = Cov(\nu_t)^{1/2} C_t Cov(\nu_t)^{1/2}$. 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 $t^2 e^{\frac{R^2}{2t^2}}$.

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

[3] 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.

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

[5] 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.

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

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

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