Sampling from spherical spin glasses: diffusions and simulated annealing

Brice Huang (Stanford → Yale)

Workshop on High-Dimensional Learning Dynamics

Banff International Research Station | February 2, 2026

Stanford | Science Fellows

NSF

Thanks to wonderful collaborators

Andrea Montanari

Huy Tuan Pham

Sidhanth Mohanty

Amit Rajaraman

David X. Wu

Outline of talk:

Introduction and motivating problems

Results: sampling from spin glasses

Algorithmic stochastic localization

Sampling guarantee for simulated annealing

Conclusion

High-dimensional sampling

Motivating problem

Sample from a high-dimensional probability measure

$$\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}$$

with a polynomial-time algorithm.

High-dimensional sampling

Motivating problem

Sample from a high-dimensional probability measure

$$\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}$$

with a polynomial-time algorithm.

Useful in: inference, integration, optimization, ...

High-dimensional sampling

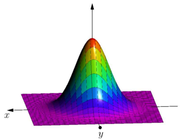
Motivating problem

Sample from a high-dimensional probability measure

$$\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}$$

with a polynomial-time algorithm.

Useful in: inference, integration, optimization, ...



μ strongly log-concave:
$\nabla^2 V \preceq -\alpha I$
Sampling is easy

High-dimensional sampling

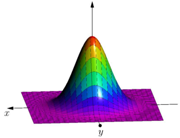
Motivating problem

Sample from a high-dimensional probability measure

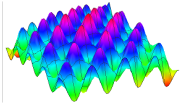
$$\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}$$

with a polynomial-time algorithm.

Useful in: inference, integration, optimization, ...



μ strongly log-concave:
$\nabla^2 V \preceq -\alpha I$
Sampling is easy



μ non-log-concave, highly multimodal
Setting of this talk

Model: mean field spin glass

E.g. random cubic polynomial on spherical domain $S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$:

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} \, x_{i_1} x_{i_2} x_{i_3} \qquad g_{i_1,i_2,i_3} \stackrel{\text{IID}}{\sim} \mathcal{N}(0,1)$$

Model: mean field spin glass

E.g. random cubic polynomial on spherical domain $S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$:

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} x_{i_1} x_{i_2} x_{i_3} \qquad g_{i_1,i_2,i_3} \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$$

Gibbs measure: $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$ $\qquad \beta \geqslant 0$ inverse temperature

Model: mean field spin glass

E.g. random cubic polynomial on spherical domain $S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$:

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} x_{i_1} x_{i_2} x_{i_3} \qquad g_{i_1,i_2,i_3} \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$$

Gibbs measure: $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$ $\quad \beta \geqslant 0$ inverse temperature

- $\beta = 0$: $\mu_\beta = \mathrm{unif}(S_N)$. Sampling is easy.

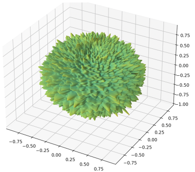
Model: mean field spin glass

E.g. random cubic polynomial on spherical domain $S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$:

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1, i_2, i_3 = 1}^{N} g_{i_1, i_2, i_3} x_{i_1} x_{i_2} x_{i_3} \qquad g_{i_1, i_2, i_3} \overset{\text{IID}}{\sim} \mathcal{N}(0, 1)$$

Gibbs measure: $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$ $\qquad \beta \geqslant 0$ inverse temperature

- $\beta = 0$: $\mu_\beta = \mathrm{unif}(S_N)$. Sampling is easy.
- β large: very non-log-concave, $e^{\Theta(N)}$ maxima.



Model: mean field spin glass

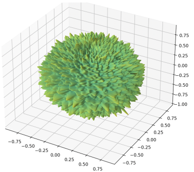
E.g. random cubic polynomial on spherical domain $S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$:

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} x_{i_1} x_{i_2} x_{i_3} \qquad g_{i_1,i_2,i_3} \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$$

Gibbs measure: $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$ $\qquad \beta \geqslant 0$ inverse temperature

- $\beta = 0$: $\mu_\beta = \mathrm{unif}(S_N)$. Sampling is easy.
- β large: very non-log-concave, $e^{\Theta(N)}$ maxima.

Sampling seems hard; known NP-hard in worst case.



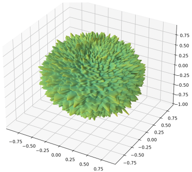
Model: mean field spin glass

E.g. random cubic polynomial on spherical domain $S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$:

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} x_{i_1} x_{i_2} x_{i_3} \qquad g_{i_1,i_2,i_3} \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$$

Gibbs measure: $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$ $\qquad$ $\beta \geqslant 0$ inverse temperature

- $\beta = 0$: $\mu_\beta = \mathrm{unif}(S_N)$. Sampling is easy.
- β large: very non-log-concave, $e^{\Theta(N)}$ maxima.
Sampling seems hard; known NP-hard in worst case.



More generally:

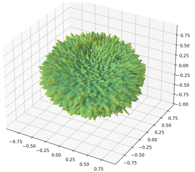
- linear combination of degrees (**mixed p-spin model**)

Model: mean field spin glass

E.g. random cubic polynomial on spherical domain $S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$:

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1,i_2,i_3=1}^{N} g_{i_1,i_2,i_3} x_{i_1} x_{i_2} x_{i_3} \qquad g_{i_1,i_2,i_3} \overset{\text{IID}}{\sim} \mathcal{N}(0,1)$$

Gibbs measure: $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$ $\quad \beta \geqslant 0$ inverse temperature



- $\beta = 0$: $\mu_\beta = \mathrm{unif}(S_N)$. Sampling is easy.
- β large: very non-log-concave, $e^{\Theta(N)}$ maxima.
Sampling seems hard; known NP-hard in worst case.

More generally:

- linear combination of degrees (**mixed p-spin model**)
- models on cube $\Sigma_N = \{\pm 1\}^N$, e.g. **Sherrington–Kirkpatrick model** (1975) is degree 2

Applications of spin glasses

Prototype of disordered probability measures / objective functions:

Applications of spin glasses

Prototype of disordered probability measures / objective functions:

- Random **constraint satisfaction** (MaxCut, MaxSAT, q-coloring):

$$\frac{1}{N}\,\mathrm{MaxCut}\Big(\ G(N, d/N)\ \Big) \ = \ d/4 + C_\star\sqrt{d} + o(\sqrt{d})$$

$C_\star$ defined by a spin glass. (Dembo Montanari Sen 17, Panchenko 18)

Applications of spin glasses

Prototype of disordered probability measures / objective functions:

- Random **constraint satisfaction** (MaxCut, MaxSAT, q-coloring):

$$\frac{1}{N}\,\mathrm{MaxCut}\Big(\;G(N,d/N)\;\Big) \;=\; d/4 + C_\star\sqrt{d} + o(\sqrt{d})$$

$C_\star$ defined by a spin glass. (Dembo Montanari Sen 17, Panchenko 18)

- **Neural network** loss landscapes (Choromanska Henaff Mathieu Ben Arous LeCun 15)

Applications of spin glasses

Prototype of disordered probability measures / objective functions:

- Random **constraint satisfaction** (MaxCut, MaxSAT, q-coloring):

$$\frac{1}{N}\,\text{MaxCut}\Big(\;G(N,d/N)\;\Big)\;=\;d/4+C_\star\sqrt{d}+o(\sqrt{d})$$

$C_\star$ defined by a spin glass. (Dembo Montanari Sen 17, Panchenko 18)

- **Neural network** loss landscapes (Choromanska Henaff Mathieu Ben Arous LeCun 15)

- Posteriors in **high-dimensional Bayesian inference**, e.g. tensor PCA, generalized linear models (Ben Arous Mei Montanari Nica 17, Barbier Krzakala Macris Miolane Zdeborová 18, Fan Mei Montanari 21)

Outline of talk:

Introduction and motivating problems

Results: sampling from spin glasses

Algorithmic stochastic localization

Sampling guarantee for simulated annealing

Conclusion

Problem restatement

We are given a (random instance of) spin glass Hamiltonian, e.g.

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1, i_2, i_3 = 1}^{N} g_{i_1, i_2, i_3} \, x_{i_1} x_{i_2} x_{i_3} \qquad\qquad g_{i_1, i_2, i_3} \overset{\text{IID}}{\sim} \mathcal{N}(0, 1)$$

More generally, linear combination of different degrees. **Gibbs measure**:

$$\mu_\beta(\mathrm{d}\boldsymbol{x}) = \frac{1}{Z_\beta} e^{\beta H(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x} \qquad\qquad \boldsymbol{x} \in S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$$

Problem restatement

We are given a (random instance of) spin glass Hamiltonian, e.g.

$$H(\boldsymbol{x}) = \frac{1}{N} \sum_{i_1, i_2, i_3 = 1}^{N} g_{i_1, i_2, i_3} \, x_{i_1} x_{i_2} x_{i_3} \qquad\qquad g_{i_1, i_2, i_3} \overset{\text{IID}}{\sim} \mathcal{N}(0, 1)$$

More generally, linear combination of different degrees. **Gibbs measure**:

$$\mu_\beta(\mathrm{d}\boldsymbol{x}) = \frac{1}{Z_\beta} e^{\beta H(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x} \qquad\qquad \boldsymbol{x} \in S_N = \sqrt{N} \cdot \mathbb{S}^{N-1}$$

Goal: devise randomized poly-time algorithm outputting $\boldsymbol{x}_{\mathrm{alg}}$ such that

$$\mathrm{Law}(\boldsymbol{x}_{\mathrm{alg}}) \approx \mu_\beta \qquad\qquad \text{whp over } H$$

For which β is this possible?

Langevin dynamics

Langevin dynamics

Canonical Markov chain for continuous distributions:

$$\mathrm{d}\boldsymbol{X}_t = \beta \nabla H(\boldsymbol{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\boldsymbol{B}_t \qquad \boldsymbol{B}_t \text{ std } \mathbb{R}^N\text{-Brownian motion}$$

Stationary distr $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$.

Langevin dynamics

Langevin dynamics

Canonical Markov chain for continuous distributions:

$$\mathrm{d}\boldsymbol{X}_t = \beta \nabla H(\boldsymbol{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\boldsymbol{B}_t \qquad \boldsymbol{B}_t \text{ std } \mathbb{R}^N\text{-Brownian motion}$$

Stationary distr $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$. (Actually need spherical version)

Langevin dynamics

Langevin dynamics

Canonical Markov chain for continuous distributions:

$$\mathrm{d}\boldsymbol{X}_t = \beta \nabla H(\boldsymbol{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\boldsymbol{B}_t \qquad \boldsymbol{B}_t \text{ std } \mathbb{R}^N\text{-Brownian motion}$$

Stationary distr $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$. (Actually need spherical version)

Yields poly-time sampling algorithm if **mixes rapidly**:

$$d_{\mathrm{TV}}(\mathrm{Law}(\boldsymbol{X}_T), \mu_\beta) = o(1) \qquad \text{for } T = \mathrm{poly}(N)$$

Langevin dynamics

Langevin dynamics

Canonical Markov chain for continuous distributions:

$$\mathrm{d}\boldsymbol{X}_t = \beta \nabla H(\boldsymbol{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\boldsymbol{B}_t \qquad \boldsymbol{B}_t \text{ std } \mathbb{R}^N\text{-Brownian motion}$$

Stationary distr $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$. (Actually need spherical version)

Yields poly-time sampling algorithm if **mixes rapidly**:

$$d_{\mathrm{TV}}(\mathrm{Law}(\boldsymbol{X}_T), \mu_\beta) = o(1) \qquad \text{for } T = \mathrm{poly}(N)$$

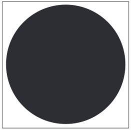
Implement by discretizing time (Chewi Erdoğlu Li Shen Zhang 24, . . .) but we won't worry about this.

Worst-case vs random initialization

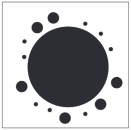
Ideal guarantee: rapid mixing from **worst-case** init. Not always possible!

Worst-case vs random initialization

Ideal guarantee: rapid mixing from **worst-case** init. Not always possible!



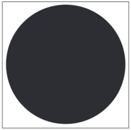
μ well-connected



μ well-connected except $o(1)$ mass in **metastable states**

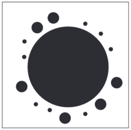
Worst-case vs random initialization

Ideal guarantee: rapid mixing from **worst-case** init. Not always possible!



μ well-connected

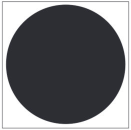
Mixes from worst-case init



μ well-connected except $o(1)$ mass in **metastable states**

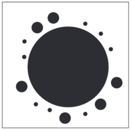
Worst-case vs random initialization

Ideal guarantee: rapid mixing from **worst-case** init. Not always possible!



μ well-connected

Mixes from worst-case init

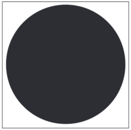


μ well-connected except $o(1)$ mass in **metastable states**

Fails to mix from worst-case init

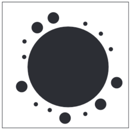
Worst-case vs random initialization

Ideal guarantee: rapid mixing from **worst-case** init. Not always possible!



μ well-connected

Mixes from worst-case init



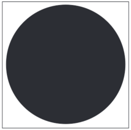
μ well-connected except $o(1)$ mass in **metastable states**

Fails to mix from worst-case init

Does this mix from random init?

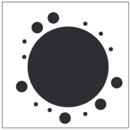
Worst-case vs random initialization

Ideal guarantee: rapid mixing from **worst-case** init. Not always possible!



μ well-connected

Mixes from worst-case init



μ well-connected except $o(1)$ mass in **metastable states**

Fails to mix from worst-case init

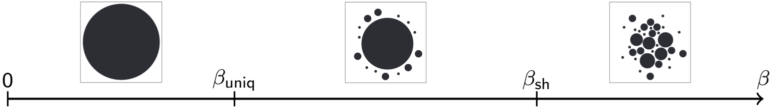
Does this mix from random init?

μ **shattered**

Fails to mix from any init

Results (informal)

Picture conjectured by physics:



Results (informal)

Picture conjectured by physics:

Results (informal)

Picture conjectured by physics:

Results (informal)

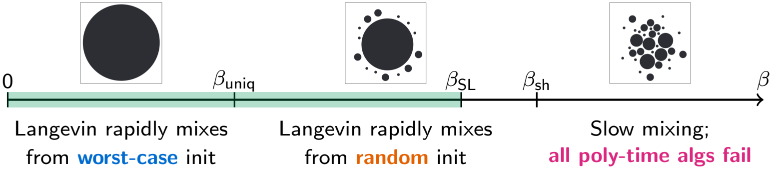
Picture conjectured by physics:

Key physics conjecture

Poly-time algorithms can sample for $\beta < \beta_{\text{sh}}$, and no further. Both sides are open.

Results (informal)

Picture conjectured by physics:



Key physics conjecture

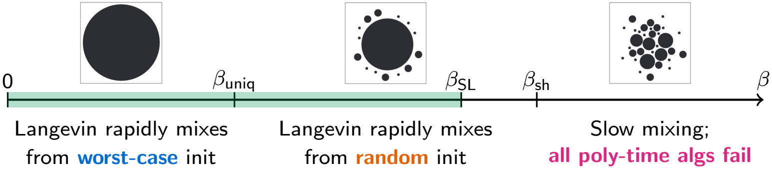
Poly-time algorithms can sample for $\beta < \beta_{\text{sh}}$, and no further. Both sides are open.

Main results

Two algorithms succeed up to intermediate β_{SL}:
simulated annealing and **algorithmic stochastic localization**.

Results (informal)

Picture conjectured by physics:



Key physics conjecture

Poly-time algorithms can sample for $\beta < \beta_{sh}$, and no further. Both sides are open.

Main results

Two algorithms succeed up to intermediate β_{SL}:
simulated annealing and **algorithmic stochastic localization**.

For **pure** p-spin model, $\frac{\beta_{SL}}{\beta_{sh}} = \frac{\sqrt{e}}{2} + O(\frac{1}{p}) \approx 0.824$.

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $\ell - 1$ output

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $\ell - 1$ output

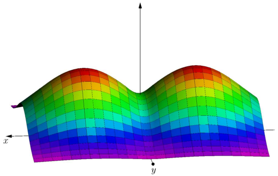


Figure: $\ell / L = 0.1$

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $\ell - 1$ output

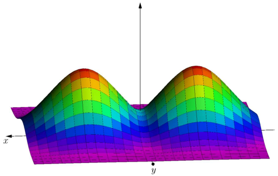


Figure: $\ell/L = 0.2$

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \dots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $\ell - 1$ output

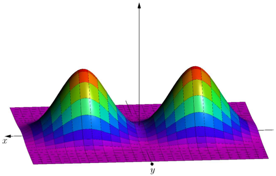


Figure: $\ell / L = 0.4$

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $\ell - 1$ output

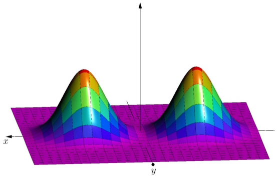


Figure: $\ell / L = 0.6$

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell/L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $\ell - 1$ output

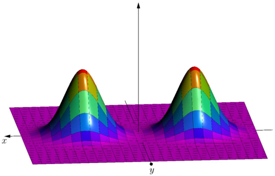


Figure: $\ell/L = 0.8$

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $\ell - 1$ output

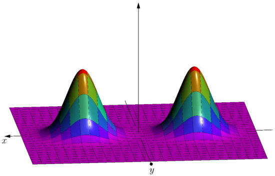


Figure: $\ell / L = 1.0$

Simulated annealing

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $\ell - 1$ output

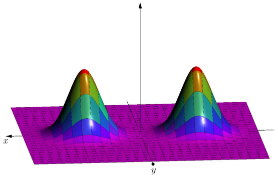


Figure: $\ell/L = 1.0$

Natural and long-studied algorithm for sampling / optimization (Kirkpatrick Gelatt Vecchi 83, . . .) but rigorous guarantees scarce.

Stochastic localization

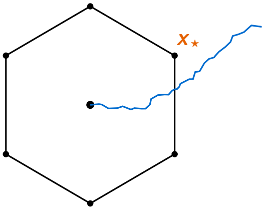
Given target measure $\mu \in \mathcal{P}(\mathbb{R}^N)$. Consider Bayesian inference setup:

Stochastic localization

Given target measure $\mu \in \mathcal{P}(\mathbb{R}^N)$. Consider Bayesian inference setup:

Nature generates secret $\boldsymbol{x}_\star \sim \mu$. We observe the process:

$$\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t \qquad \boldsymbol{B}_t \text{ std } \mathbb{R}^N\text{-Brownian motion}$$

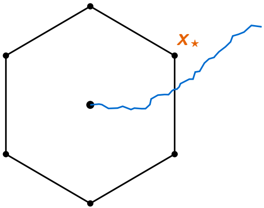


Stochastic localization

Given target measure $\mu \in \mathcal{P}(\mathbb{R}^N)$. Consider Bayesian inference setup:

Nature generates secret $\boldsymbol{x}_\star \sim \mu$. We observe the process:

$$\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t \qquad \boldsymbol{B}_t \text{ std } \mathbb{R}^N\text{-Brownian motion}$$



We form μ^t as posterior on $\boldsymbol{x}_\star$:

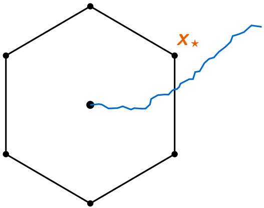
$$\mu^t = \mathrm{Law}(\boldsymbol{x}_\star | (\boldsymbol{y}_s)_{s \leqslant t})$$

Stochastic localization

Given target measure $\mu \in \mathcal{P}(\mathbb{R}^N)$. Consider Bayesian inference setup:

Nature generates secret $\boldsymbol{x}_\star \sim \mu$. We observe the process:

$$\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t \qquad \boldsymbol{B}_t \text{ std } \mathbb{R}^N\text{-Brownian motion}$$



We form μ^t as posterior on $\boldsymbol{x}_\star$:

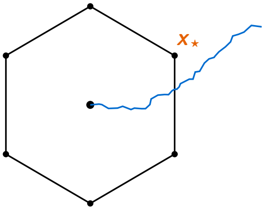
$$\begin{aligned} \mu^t &= \mathrm{Law}(\boldsymbol{x}_\star | (\boldsymbol{y}_s)_{s \leqslant t}) \\ &= \mathrm{Law}(\boldsymbol{x}_\star | \boldsymbol{y}_t) \end{aligned}$$

Stochastic localization

Given target measure $\mu \in \mathcal{P}(\mathbb{R}^N)$. Consider Bayesian inference setup:

Nature generates secret $\boldsymbol{x}_\star \sim \mu$. We observe the process:

$$\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t \qquad \boldsymbol{B}_t \text{ std } \mathbb{R}^N\text{-Brownian motion}$$



We form μ^t as posterior on $\boldsymbol{x}_\star$:

$$\begin{aligned}\mu^t &= \mathrm{Law}(\boldsymbol{x}_\star | (\boldsymbol{y}_s)_{s \leqslant t}) \\ &= \mathrm{Law}(\boldsymbol{x}_\star | \boldsymbol{y}_t)\end{aligned}$$

SL process is the **path of measures** $(\mu^t)_{t \geqslant 0}$. Localizes to $\boldsymbol{x}_\star$ as $t \to \infty$.

Two usages of stochastic localization

SL as a sampling algorithm

Simulate SL process $(\mu^t)_{t \geqslant 0}$ in distribution (without knowing $\boldsymbol{x}_\star$!) until $\mu^\infty = \delta_{\boldsymbol{x}}$ to obtain sample $\boldsymbol{x} \sim \mu$.

Introduced in El Alaoui Montanari Sellke 22+23 to sample from Sherrington–Kirkpatrick model. Equivalent to denoising diffusions.

Two usages of stochastic localization

SL as a sampling algorithm

Simulate SL process $(\mu^t)_{t \geqslant 0}$ in distribution (without knowing $\boldsymbol{x}_\star$!) until $\mu^\infty = \delta_{\boldsymbol{x}}$ to obtain sample $\boldsymbol{x} \sim \mu$.

Introduced in El Alaoui Montanari Sellke 22+23 to sample from Sherrington–Kirkpatrick model. Equivalent to denoising diffusions.

SL as a proof technique

Control of SL process $(\mu^t)_{t \geqslant 0} \Rightarrow$ functional ineqs for $\mu \Rightarrow$ rapid mixing of Langevin dynamics on μ.

Originated in convex geometry, breakthrough applications to KLS conjecture (Eldan 13, Lee Vempala 17, Chen 21, Klartag Lehec 22+25)

Results: sampling from spin glass $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$

Results: sampling from spin glass $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$

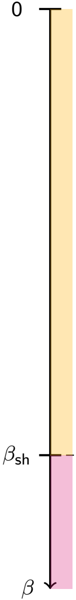
0

β_{sh}

Gibbs measure shatters; sampling predicted hard.
Rigorous hardness for **stable** samplers.
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23,
Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

β

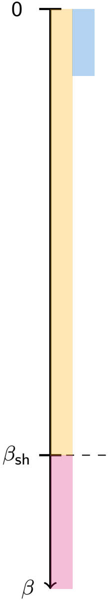
Results: sampling from spin glass $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$



Prediction: Langevin succeeds from random init

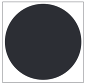
Gibbs measure shatters; sampling predicted hard.
Rigorous hardness for **stable** samplers.
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Results: sampling from spin glass $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$



Langevin mixes rapidly from worst-case init

(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



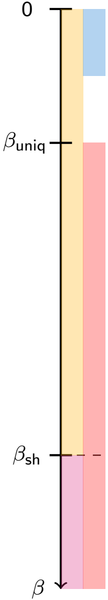
Prediction: Langevin succeeds from random init

Gibbs measure shatters; sampling predicted hard.

Rigorous hardness for **stable** samplers.

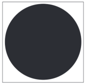
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Results: sampling from spin glass $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$



Langevin mixes rapidly from worst-case init

(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



Langevin mixes slowly from worst-case init

(Ben Arous Jagannath 24)

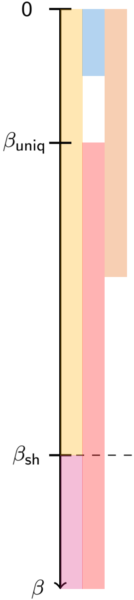
Prediction: Langevin succeeds from random init

Gibbs measure shatters; sampling predicted hard.

Rigorous hardness for **stable** samplers.

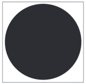
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Results: sampling from spin glass $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$



Langevin mixes rapidly from worst-case init

(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



Algorithmic SL succeeds in Wasserstein error

(El Alaoui Montanari Sellke 22+23)

Langevin mixes slowly from worst-case init

(Ben Arous Jagannath 24)

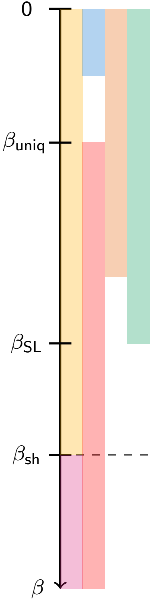
Prediction: Langevin succeeds from random init

Gibbs measure shatters; sampling predicted hard.

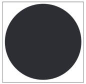
Rigorous hardness for **stable** samplers.

(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Results: sampling from spin glass $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$



Langevin mixes rapidly from worst-case init
(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



Algorithmic SL succeeds in Wasserstein error
(El Alaoui Montanari Sellke 22+23)

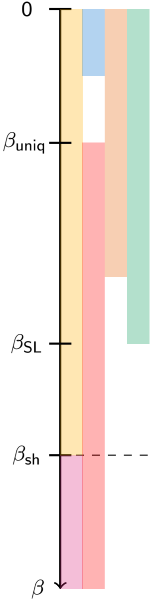
Result 1: Algorithmic SL succeeds in **total variation** error
+ matching hardness result (H Montanari Pham 24)

Langevin mixes slowly from worst-case init
(Ben Arous Jagannath 24)

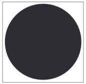
Prediction: Langevin succeeds from random init

Gibbs measure shatters; sampling predicted hard.
Rigorous hardness for **stable** samplers.
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Results: sampling from spin glass $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$



Langevin mixes rapidly from worst-case init
(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



Algorithmic SL succeeds in Wasserstein error
(El Alaoui Montanari Sellke 22+23)

Result 1: Algorithmic SL succeeds in **total variation** error
+ matching hardness result (H Montanari Pham 24)

Result 2: Simulated annealing succeeds (in TV error)
(H Mohanty Rajaraman Wu 24)

Langevin mixes slowly from worst-case init
(Ben Arous Jagannath 24)

Prediction: Langevin succeeds from random init

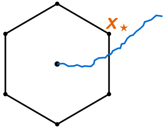
Gibbs measure shatters; sampling predicted hard.
Rigorous hardness for **stable** samplers.
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Outline of talk:

Stochastic localization as an algorithm

Recall: $\boldsymbol{x}_\star \sim \mu$ secret, we observe

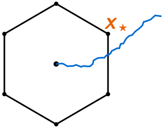
$$\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t \quad \text{and infer} \quad \mu^t = \text{Law}(\boldsymbol{x}_\star | \boldsymbol{y}_t)$$



Stochastic localization as an algorithm

Recall: $\boldsymbol{x}_\star \sim \mu$ secret, we observe

$$\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t \quad \text{and infer} \quad \mu^t = \text{Law}(\boldsymbol{x}_\star | \boldsymbol{y}_t)$$



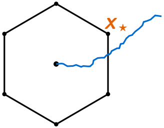
Key idea (El Alaoui Montanari Sellke 22)

We can simulate the SL process, in distribution, without knowing $\boldsymbol{x}_\star$.

Stochastic localization as an algorithm

Recall: $\boldsymbol{x}_\star \sim \mu$ secret, we observe

$$\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t \quad \text{and infer} \quad \mu^t = \mathrm{Law}(\boldsymbol{x}_\star | \boldsymbol{y}_t)$$



Key idea (El Alaoui Montanari Sellke 22)

We can simulate the SL process, in distribution, without knowing $\boldsymbol{x}_\star$.

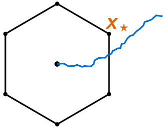
For another Brownian motion $\widetilde{\boldsymbol{B}}_t$ adapted to $\mathcal{F}((\boldsymbol{y}_s)_{s\leqslant t})$,

$$\mathrm{d}\boldsymbol{y}_t = \boldsymbol{m}_t\,\mathrm{d}t + \mathrm{d}\widetilde{\boldsymbol{B}}_t \qquad\qquad \boldsymbol{m}_t = \mathrm{mean}(\mu^t) = \mathbb{E}[\boldsymbol{x}_\star | \boldsymbol{y}_t]$$

Stochastic localization as an algorithm

Recall: $x_\star \sim \mu$ secret, we observe

$$y_t = t x_\star + B_t \quad \text{and infer} \quad \mu^t = \mathrm{Law}(x_\star | y_t)$$



Key idea (El Alaoui Montanari Sellke 22)

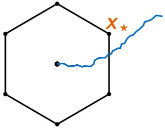
We can simulate the SL process, in distribution, without knowing $x_\star$.

For another Brownian motion $\widetilde{B}_t$ adapted to $\mathcal{F}((y_s)_{s \leqslant t})$,

$$\mathrm{d}y_t = m_t \, \mathrm{d}t + \mathrm{d}\widetilde{B}_t \qquad\qquad m_t = \mathrm{mean}(\mu^t) = \mathbb{E}[x_\star | y_t]$$

💡 This SDE is autonomous. Can simulate without "ground truth" $x_\star$

Stochastic localization as an algorithm



Recall: $x_\star \sim \mu$ secret, we observe

$$y_t = t x_\star + B_t \quad \text{and infer} \quad \mu^t = \mathrm{Law}(x_\star | y_t)$$

Key idea (El Alaoui Montanari Sellke 22)

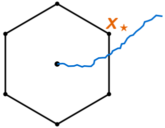
We can simulate the SL process, in distribution, without knowing $x_\star$.

For another Brownian motion $\widetilde{B}_t$ adapted to $\mathcal{F}((y_s)_{s \leqslant t})$,

$$\mathrm{d}y_t = m_t \, \mathrm{d}t + \mathrm{d}\widetilde{B}_t \qquad m_t = \mathrm{mean}(\mu^t) = \mathbb{E}[x_\star | y_t] = f(t, y_t)$$

This SDE is autonomous. Can simulate without "ground truth" $x_\star$

Stochastic localization as an algorithm



Recall: $x_\star \sim \mu$ secret, we observe

$$y_t = t x_\star + B_t \quad \text{and infer} \quad \mu^t = \mathrm{Law}(x_\star | y_t)$$

Key idea (El Alaoui Montanari Sellke 22)

We can simulate the SL process, in distribution, without knowing $x_\star$.

For another Brownian motion $\widetilde{B}_t$ adapted to $\mathcal{F}((y_s)_{s \leqslant t})$,

$$\mathrm{d}y_t = m_t \, \mathrm{d}t + \mathrm{d}\widetilde{B}_t \qquad\qquad m_t = \mathrm{mean}(\mu^t) = \mathbb{E}[x_\star | y_t] = f(t, y_t)$$

This SDE is autonomous. Can simulate without “ground truth” $x_\star$

Upshot: reduces sampling to posterior mean estimation!

$$\text{Estimate } m_t \;\Rightarrow\; \text{simulate } (y_t)_{t \geqslant 0} \;\Rightarrow\; \text{sample from } \mu$$

Form of the posterior μ^t

Recall $\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t$. If we observe $\boldsymbol{y}_t = \boldsymbol{y}$:

$$\mu^t(\mathrm{d}\boldsymbol{x}) = \mathbb{P}\Big(\boldsymbol{x}_\star \in \boldsymbol{x} + \mathrm{d}\boldsymbol{x} \,\Big|\, \boldsymbol{y}_t = \boldsymbol{y}\Big)$$

Form of the posterior μ^t

Recall $\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t$. If we observe $\boldsymbol{y}_t = \boldsymbol{y}$:

$$\mu^t(\mathrm{d}\boldsymbol{x}) = \mathbb{P}\Big(\boldsymbol{x}_\star \in \boldsymbol{x} + \mathrm{d}\boldsymbol{x} \,\Big|\, \boldsymbol{y}_t = \boldsymbol{y}\Big)$$

$$\propto \mathbb{P}\Big(\boldsymbol{y}_t = \boldsymbol{y} \,\Big|\, \boldsymbol{x}_\star = \boldsymbol{x}\Big)\, \mu(\mathrm{d}\boldsymbol{x}) \qquad \text{(Bayes' rule)}$$

Form of the posterior μ^t

Recall $\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t$. If we observe $\boldsymbol{y}_t = \boldsymbol{y}$:

$$\begin{aligned}
\mu^t(\mathrm{d}\boldsymbol{x}) &= \mathbb{P}\Big(\boldsymbol{x}_\star \in \boldsymbol{x} + \mathrm{d}\boldsymbol{x} \,\Big|\, \boldsymbol{y}_t = \boldsymbol{y}\Big) \\
&\propto \mathbb{P}\Big(\boldsymbol{y}_t = \boldsymbol{y} \,\Big|\, \boldsymbol{x}_\star = \boldsymbol{x}\Big)\,\mu(\mathrm{d}\boldsymbol{x}) && \text{(Bayes' rule)} \\
&\propto \exp(\langle \boldsymbol{y}, \boldsymbol{x}\rangle)\,\mu(\mathrm{d}\boldsymbol{x})
\end{aligned}$$

Form of the posterior μ^t

Recall $\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t$. If we observe $\boldsymbol{y}_t = \boldsymbol{y}$:

$$\begin{aligned}\mu^t(\mathrm{d}\boldsymbol{x}) &= \mathbb{P}\Big(\boldsymbol{x}_\star \in \boldsymbol{x} + \mathrm{d}\boldsymbol{x} \,\Big|\, \boldsymbol{y}_t = \boldsymbol{y}\Big) \\ &\propto \mathbb{P}\Big(\boldsymbol{y}_t = \boldsymbol{y} \,\Big|\, \boldsymbol{x}_\star = \boldsymbol{x}\Big)\,\mu(\mathrm{d}\boldsymbol{x}) \qquad \text{(Bayes' rule)} \\ &\propto \exp(\langle \boldsymbol{y}, \boldsymbol{x}\rangle)\,\mu(\mathrm{d}\boldsymbol{x})\end{aligned}$$

That is, μ^t is μ with randomly evolving exponential tilt ($\leftrightarrow$ external field)

$$\mu^t(\mathrm{d}\boldsymbol{x}) \propto \exp\Big(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x}\rangle\Big)$$

Mean estimation

El Alaoui Montanari Sellke 22: estimator $\widetilde{\boldsymbol{m}}_t$ for $\boldsymbol{m}_t = \text{mean}(\mu^t)$ by **approximate message passing** (Bolthausen 14, Bayati Montanari 11)

Mean estimation

El Alaoui Montanari Sellke 22: estimator $\widetilde{\boldsymbol{m}}_t$ for $\boldsymbol{m}_t = \text{mean}(\mu^t)$ by **approximate message passing** (Bolthausen 14, Bayati Montanari 11)

This estimator has accuracy

$$\mathbb{E}[\|\widetilde{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2] = o(N)$$

Mean estimation

El Alaoui Montanari Sellke 22: estimator $\widetilde{\boldsymbol{m}}_t$ for $\boldsymbol{m}_t = \text{mean}(\mu^t)$ by **approximate message passing** (Bolthausen 14, Bayati Montanari 11)

This estimator has accuracy

$$\mathbb{E}[\|\widetilde{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2] = o(N)$$

which yields sampler with **Wasserstein** guarantee:

$$\mathbb{E}[\|\boldsymbol{x}_{\text{alg}} - \boldsymbol{x}\|^2] = o(N) \qquad \text{for coupled } \boldsymbol{x}_{\text{alg}} \sim \mu_{\text{alg}},\ \boldsymbol{x} \sim \mu_\beta$$

Mean estimation

El Alaoui Montanari Sellke 22: estimator $\widetilde{\boldsymbol{m}}_t$ for $\boldsymbol{m}_t = \text{mean}(\mu^t)$ by **approximate message passing** (Bolthausen 14, Bayati Montanari 11)

This estimator has accuracy

$$\mathbb{E}[\|\widetilde{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2] = o(N)$$

which yields sampler with **Wasserstein** guarantee:

$$\mathbb{E}[\|\boldsymbol{x}_{\text{alg}} - \boldsymbol{x}\|^2] = o(N) \qquad \text{for coupled } \boldsymbol{x}_{\text{alg}} \sim \mu_{\text{alg}},\ \boldsymbol{x} \sim \mu_\beta$$

Also “asymptotically accurate” in the sense $d_{\text{KL}}(\mu_{\text{alg}}, \mu_\beta) = o(N)$

Mean estimation

El Alaoui Montanari Sellke 22: estimator $\widetilde{\boldsymbol{m}}_t$ for $\boldsymbol{m}_t = \text{mean}(\mu^t)$ by **approximate message passing** (Bolthausen 14, Bayati Montanari 11)

This estimator has accuracy

$$\mathbb{E}[\|\widetilde{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2] = o(N)$$

which yields sampler with **Wasserstein** guarantee:

$$\mathbb{E}[\|\boldsymbol{x}_{\text{alg}} - \boldsymbol{x}\|^2] = o(N) \qquad \text{for coupled } \boldsymbol{x}_{\text{alg}} \sim \mu_{\text{alg}},\ \boldsymbol{x} \sim \mu_\beta$$

Also "asymptotically accurate" in the sense $d_{\text{KL}}(\mu_{\text{alg}}, \mu_\beta) = o(N)$

Main contribution of (H Montanari Pham 24)

For $\beta < \beta_{\text{SL}}$, we develop improved estimator $\widehat{\boldsymbol{m}}_t$ with accuracy

$$\mathbb{E}[\|\widehat{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2] = o(1)$$

This implies $d_{\text{TV}}(\mu_{\text{alg}}, \mu_\beta) = o(1)$.

Outline of talk:

Introduction and motivating problems

Results: sampling from spin glasses

Algorithmic stochastic localization

Sampling guarantee for simulated annealing

Conclusion

Poincaré inequalities and rapid mixing

For $\mu \in \mathcal{P}(S_N)$, test function $f : S_N \to \mathbb{R}$, define **Dirichlet form**

$$\mathcal{E}_\mu(f,f) = \mathbb{E}_\mu\left[\|\nabla_{\mathsf{sp}} f\|^2\right]$$

Poincaré inequality

μ satisfies a Poincaré inequality with constant $C > 0$ if for all f,

$$C\mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f,f)$$

Poincaré inequalities and rapid mixing

For $\mu \in \mathcal{P}(S_N)$, test function $f : S_N \to \mathbb{R}$, define **Dirichlet form**

$$\mathcal{E}_\mu(f, f) = \mathbb{E}_\mu\left[\|\nabla_{\mathsf{sp}} f\|^2\right]$$

Poincaré inequality

μ satisfies a Poincaré inequality with constant $C > 0$ if for all f,

$$C\mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f, f)$$

Implies Langevin dynamics exponentially contract χ^2 distance to μ:

$$\chi^2(P_t\nu \parallel \mu) \leqslant e^{-Ct}\chi^2(\nu \parallel \mu)$$

Poincaré inequalities and rapid mixing

For $\mu \in \mathcal{P}(S_N)$, test function $f : S_N \to \mathbb{R}$, define **Dirichlet form**

$$\mathcal{E}_\mu(f, f) = \mathbb{E}_\mu\left[\|\nabla_{\mathsf{sp}} f\|^2\right]$$

Poincaré inequality

μ satisfies a Poincaré inequality with constant $C > 0$ if for all f,

$$C\mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f, f)$$

Implies Langevin dynamics exponentially contract χ^2 distance to μ:

$$\chi^2(P_t\nu \parallel \mu) \leqslant e^{-Ct}\chi^2(\nu \parallel \mu)$$

If $C \geqslant 1/\mathrm{poly}(N) \Rightarrow$ rapid mixing from **worst-case** initialization

Poincaré inequalities and rapid mixing

For $\mu \in \mathcal{P}(S_N)$, test function $f : S_N \to \mathbb{R}$, define **Dirichlet form**

$$\mathcal{E}_\mu(f, f) = \mathbb{E}_\mu\left[\|\nabla_{\mathsf{sp}} f\|^2\right]$$

Poincaré inequality

μ satisfies a Poincaré inequality with constant $C > 0$ if for all f,

$$C\mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f, f)$$

Implies Langevin dynamics exponentially contract χ^2 distance to μ:

$$\chi^2(P_t\nu \parallel \mu) \leqslant e^{-Ct}\chi^2(\nu \parallel \mu)$$

If $C \geqslant 1/\mathrm{poly}(N) \Rightarrow$ rapid mixing from **worst-case** initialization

! Great if this works, but too much to hope for if metastable states ...

Weak Poincaré inequalities

Weak Poincaré inequality

μ satisfies a (C, δ)-weak Poincaré inequality if for all test fns f,

$$C\mathsf{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f, f) + \delta\|f\|_\infty^2$$

($\delta\|f\|_\infty^2$ accounts for metastable states. Think $\delta = e^{-N^c}$)

Weak Poincaré inequalities

Weak Poincaré inequality

μ satisfies a (C, δ)-weak Poincaré inequality if for all test fns f,

$$C\mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f, f) + \delta \|f\|_\infty^2$$

($\delta \|f\|_\infty^2$ accounts for metastable states. Think $\delta = e^{-N^c}$)

Implies "almost" exponential contraction of χ^2 to μ:

$$\chi^2(P_t \nu \parallel \mu) \leqslant e^{-Ct} \chi^2(\nu \parallel \mu) + \delta \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right\|_\infty^2$$

Weak Poincaré inequalities

Weak Poincaré inequality

μ satisfies a (C, δ)-weak Poincaré inequality if for all test fns f,

$$C\mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f, f) + \delta\|f\|_\infty^2$$

($\delta\|f\|_\infty^2$ accounts for metastable states. Think $\delta = e^{-N^c}$)

Implies "almost" exponential contraction of χ^2 to μ:

$$\chi^2(P_t\nu \parallel \mu) \leqslant e^{-Ct}\chi^2(\nu \parallel \mu) + \delta\left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_\infty^2$$

Consequence: rapid mixing from **warm starts** ν with $\|\mathrm{d}\nu/\mathrm{d}\mu\|_\infty \ll 1/\sqrt{\delta}$

WPI $\Rightarrow$ simulated annealing succeeds

Consequence of (C, δ)-weak Poincaré inequality

Langevin dynamics rapidly mix to μ from **warm starts** $\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\|_\infty \ll \frac{1}{\sqrt{\delta}}$

WPI $\Rightarrow$ simulated annealing succeeds

Consequence of (C, δ)-weak Poincaré inequality

Langevin dynamics rapidly mix to μ from **warm starts** $\|\frac{d\nu}{d\mu}\|_\infty \ll \frac{1}{\sqrt{\delta}}$

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(d\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $t - 1$ output

WPI $\Rightarrow$ simulated annealing succeeds

Consequence of (C, δ)-weak Poincaré inequality

Langevin dynamics rapidly mix to μ from **warm starts** $\|\frac{d\nu}{d\mu}\|_\infty \ll \frac{1}{\sqrt{\delta}}$

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(d\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $t-1$ output

μ^ℓ is warm start for $\mu^{\ell+1}$. WPI $\Rightarrow$ inductively show each stage mixes

WPI $\Rightarrow$ simulated annealing succeeds

Consequence of (C, δ)-weak Poincaré inequality

Langevin dynamics rapidly mix to μ from **warm starts** $\|\frac{d\nu}{d\mu}\|_\infty \ll \frac{1}{\sqrt{\delta}}$

Simulated annealing

Input: β, H. Goal: sample from $\mu_\beta(d\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})}$

In stage $\ell = 0, \ldots, L$:

- Set $\boldsymbol{\beta}_\ell = \beta \cdot \ell / L$. Interpolates $\boldsymbol{\beta}_0 = 0$ to $\boldsymbol{\beta}_L = \beta$
- Run Langevin on $\mu^\ell(\boldsymbol{x}) \propto e^{\boldsymbol{\beta}_\ell H(\boldsymbol{x})}$, init from stage $t - 1$ output

μ^ℓ is warm start for $\mu^{\ell+1}$. WPI $\Rightarrow$ inductively show each stage mixes

Related "staging warm starts" idea in El Alaoui Eldan Gheissari Piana 23

Proving a weak Poincaré inequality

For $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\boldsymbol{h} \in \mathbb{R}^N$, define **exponential tilt** $\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle} \mu(\mathrm{d}\boldsymbol{x})$

Proving a weak Poincaré inequality

For $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\boldsymbol{h} \in \mathbb{R}^N$, define **exponential tilt** $\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle}\mu(\mathrm{d}\boldsymbol{x})$

Message of localizaton schemes framework (Chen Eldan 22, ...)

μ satisfies a $O(1)$-Poincaré inequality if

$$\|\mathrm{cov}(\mu^{\boldsymbol{h}})\|_{\mathrm{op}} = O(1) \qquad \text{for all } \boldsymbol{h} \in \mathbb{R}^N$$

Proving a weak Poincaré inequality

For $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\boldsymbol{h} \in \mathbb{R}^N$, define **exponential tilt** $\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle}\mu(\mathrm{d}\boldsymbol{x})$

Message of localizaton schemes framework (Chen Eldan 22, . . .)

μ satisfies a $O(1)$-Poincaré inequality if

$$\|\mathrm{cov}(\mu^{\boldsymbol{h}})\|_{\mathrm{op}} = O(1) \qquad \text{for all } \boldsymbol{h} \in \mathbb{R}^N$$

Message of (H Mohanty Rajaraman Wu 24)

μ satisfies a $(O(1), \delta)$-weak Poincaré inequality if

$$\|\mathrm{cov}(\mu^{\boldsymbol{h}})\|_{\mathrm{op}} = O(1) \qquad \text{whp for random } \boldsymbol{h}$$

Proving a weak Poincaré inequality

For $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\boldsymbol{h} \in \mathbb{R}^N$, define **exponential tilt** $\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle}\mu(\mathrm{d}\boldsymbol{x})$

Message of localizaton schemes framework (Chen Eldan 22, ...)

μ satisfies a $O(1)$-Poincaré inequality if

$$\|\mathrm{cov}(\mu^{\boldsymbol{h}})\|_{\mathrm{op}} = O(1) \qquad \text{for all } \boldsymbol{h} \in \mathbb{R}^N$$

Message of (H Mohanty Rajaraman Wu 24)

μ satisfies a $(O(1), \delta)$-weak Poincaré inequality if

$$\|\mathrm{cov}(\mu^{\boldsymbol{h}})\|_{\mathrm{op}} = O(1) \qquad \text{whp for random } \boldsymbol{h}$$

More precisely: $\boldsymbol{h}$ is the SL tilt $\boldsymbol{y}_t$.

Proving a weak Poincaré inequality

For $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\boldsymbol{h} \in \mathbb{R}^N$, define **exponential tilt** $\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle} \mu(\mathrm{d}\boldsymbol{x})$

Message of localizaton schemes framework (Chen Eldan 22, ...)

μ satisfies a $O(1)$-Poincaré inequality if

$$\|\mathrm{cov}(\mu^{\boldsymbol{h}})\|_{\mathrm{op}} = O(1) \qquad \text{for all } \boldsymbol{h} \in \mathbb{R}^N$$

Message of (H Mohanty Rajaraman Wu 24)

μ satisfies a $(O(1), \delta)$-weak Poincaré inequality if

$$\|\mathrm{cov}(\mu^{\boldsymbol{h}})\|_{\mathrm{op}} = O(1) \qquad \text{whp for random } \boldsymbol{h}$$

More precisely: $\boldsymbol{h}$ is the SL tilt $\boldsymbol{y}_t$. That is, if

$$\mathbb{P}\Big(\sup_{t \geqslant 0} \|\mathrm{cov}(\mu^{\boldsymbol{y}_t})\|_{\mathrm{op}} = O(1) \Big) \geqslant 1 - \delta$$

then μ satisfies a $(O(1), \delta)$-weak Poincaré inequality

Summary of proof strategy

Approach of Chen Eldan 22:

Covariance bound for worst-case tilts	$\Rightarrow$	Poincaré inequality	$\Rightarrow$	Rapid mixing from worst-case init

Summary of proof strategy

Approach of Chen Eldan 22:

Covariance bound for worst-case tilts ⇒ Poincaré inequality ⇒ Rapid mixing from worst-case init

Approach of H Mohanty Rajaraman Wu 24:

Covariance bound for random tilts whp ⇒ Weak Poincaré inequality ⇒ Rapid mixing from warm start

⇓

Simulated annealing succeeds

Summary of proof strategy

Approach of Chen Eldan 22:

Covariance bound for worst-case tilts ⇒ Poincaré inequality ⇒ Rapid mixing from worst-case init

Approach of H Mohanty Rajaraman Wu 24:

Covariance bound for random tilts whp ⇒ Weak Poincaré inequality ⇒ Rapid mixing from warm start

⇓

Simulated annealing succeeds

We now discuss the proofs of these steps.
SL process will enter here.

Worst-case tilts covariance bound $\Rightarrow$ Poincaré inequality

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be target measure,

want to show (for all f)

$$O(1) \cdot \mathrm{Var}_\mu(f) \qquad \leqslant \qquad \mathcal{E}_\mu(f, f)$$

Worst-case tilts covariance bound $\Rightarrow$ Poincaré inequality

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be target measure, $T = O(1)$ large

want to show (for all f)

$$O(1) \cdot \mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f, f)$$

Worst-case tilts covariance bound $\Rightarrow$ Poincaré inequality

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be target measure, $T = O(1)$ large

want to show (for all f)

$$O(1) \cdot \mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f,f)$$

$\rotatebox{90}{$\geqslant$}$ Supermartingality of $\mathcal{E}_{\mu^t}$

$$O(1) \cdot \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] \leqslant \mathbb{E}_{\mu^T}[\,\mathcal{E}_{\mu^T}(f,f)\,]$$

localized measures have PI

Worst-case tilts covariance bound $\Rightarrow$ Poincaré inequality

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be target measure, $T = O(1)$ large

want to show (for all f)

$$O(1) \cdot \mathrm{Var}_\mu(f) \quad \leqslant \quad \mathcal{E}_\mu(f, f)$$

key difficulty $\quad /\!\!\!\wedge \quad\quad\quad \mathbin{\rotatebox{90}{$\geqslant$}}$ Supermartingality of $\mathcal{E}_{\mu^t}$

$$O(1) \cdot \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] \leqslant \mathbb{E}_{\mu^T}[\,\mathcal{E}_{\mu^T}(f, f)\,]$$

localized measures have PI

Worst-case tilts covariance bound $\Rightarrow$ Poincaré inequality

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be target measure, $T = O(1)$ large

want to show (for all f)

$$O(1) \cdot \mathrm{Var}_\mu(f) \leqslant \mathcal{E}_\mu(f,f)$$

key difficulty $\not\wedge$ $\qquad$ $\vee$ Supermartingality of $\mathcal{E}_{\mu^t}$

$$O(1) \cdot \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] \leqslant \mathbb{E}_{\mu^T}[\,\mathcal{E}_{\mu^T}(f,f)\,]$$

localized measures have PI

Main idea of (Chen Eldan 22)

Stochastically differentiating $\mathrm{Var}_{\mu^t}(f)$ yields

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\,\mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

Worst-case tilts covariance bound $\Rightarrow$ Poincaré inequality

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be target measure, $T = O(1)$ large

want to show (for all f)

$$O(1) \cdot \mathrm{Var}_\mu(f) \qquad \leqslant \qquad \mathcal{E}_\mu(f, f)$$

key difficulty $\not\wedge$ $\qquad$ $\vee$ Supermartingality of $\mathcal{E}_{\mu^t}$

$$O(1) \cdot \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] \;\leqslant\; \mathbb{E}_{\mu^T}[\,\mathcal{E}_{\mu^T}(f, f)\,]$$

localized measures have PI

Main idea of (Chen Eldan 22)

Stochastically differentiating $\mathrm{Var}_{\mu^t}(f)$ yields

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\, \mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

Recall μ^t is tilt of μ.

Worst-case tilts covariance bound $\Rightarrow$ Poincaré inequality

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be target measure, $T = O(1)$ large

want to show (for all f)

$$O(1) \cdot \mathrm{Var}_\mu(f) \quad \leqslant \quad \mathcal{E}_\mu(f,f)$$

key difficulty $\not\wedge$ $\quad\quad$ $\mathbin{\rotatebox[origin=c]{90}{$\geqslant$}}$ Supermartingality of $\mathcal{E}_{\mu^t}$

$$O(1) \cdot \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] \leqslant \quad \mathbb{E}_{\mu^T}[\,\mathcal{E}_{\mu^T}(f,f)\,]$$

localized measures have PI

Main idea of (Chen Eldan 22)

Stochastically differentiating $\mathrm{Var}_{\mu^t}(f)$ yields

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\,\mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

Recall μ^t is tilt of μ. Since $\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \leqslant K = O(1)$ almost surely,

Worst-case tilts covariance bound $\Rightarrow$ Poincaré inequality

Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ be target measure, $T = O(1)$ large

want to show (for all f)

$$O(1) \cdot \mathrm{Var}_{\mu}(f) \leqslant \mathcal{E}_{\mu}(f, f)$$

key difficulty $\quad /\!\!\!\wedge \qquad\qquad \mathbin{\rotatebox[origin=c]{90}{$\leqslant$}}$ Supermartingality of $\mathcal{E}_{\mu^t}$

$$O(1) \cdot \mathbb{E}_{\mu^T}[\mathrm{Var}_{\mu^T}(f)] \leqslant \mathbb{E}_{\mu^T}[\mathcal{E}_{\mu^T}(f, f)]$$

localized measures have PI

Main idea of (Chen Eldan 22)

Stochastically differentiating $\mathrm{Var}_{\mu^t}(f)$ yields

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\, \mathrm{d}t + \mathrm{d}(\text{martingale})_t$$

Recall μ^t is tilt of μ. Since $\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \leqslant K = O(1)$ almost surely,

$$\mathbb{E}_{\mu^T}[\mathrm{Var}_{\mu^T}(f)] \geqslant e^{-KT} \mathrm{Var}_{\mu}(f)$$

Random tilts covariance bound whp $\Rightarrow$ WPI

Same strategy, but with a stopping time. Let $T, K = O(1)$ large, and

$$\tau = \inf\left\{t \geqslant 0 : \|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \geqslant K\right\} \wedge T\,.$$

Random tilts covariance bound whp $\Rightarrow$ WPI

Same strategy, but with a stopping time. Let $T, K = O(1)$ large, and

$$\tau = \inf \left\{ t \geqslant 0 : \|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \geqslant K \right\} \wedge T \,.$$

As before,

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\, \mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

Random tilts covariance bound whp $\Rightarrow$ WPI

Same strategy, but with a stopping time. Let $T, K = O(1)$ large, and

$$\tau = \inf \left\{ t \geqslant 0 : \|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \geqslant K \right\} \wedge T \,.$$

As before,

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\, \mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

which implies

$$\mathbb{E}_{\mu^\tau}\left[\, \mathrm{Var}_{\mu^\tau}(f) \,\right] \geqslant e^{-KT} \mathrm{Var}_{\mu}(f)$$

Random tilts covariance bound whp $\Rightarrow$ WPI

Same strategy, but with a stopping time. Let $T, K = O(1)$ large, and

$$\tau = \inf\left\{t \geqslant 0 : \|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \geqslant K\right\} \wedge T\,.$$

As before,

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}}\mathrm{Var}_{\mu^t}(f)\,\mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

which implies

$$\mathbb{E}_{\mu^\tau}\left[\,\mathrm{Var}_{\mu^\tau}(f)\,\right] \geqslant e^{-KT}\mathrm{Var}_\mu(f)$$

High-probability covariance bound $\Rightarrow \mathbb{P}(\tau < T) \leqslant \delta$.

Random tilts covariance bound whp $\Rightarrow$ WPI

Same strategy, but with a stopping time. Let $T, K = O(1)$ large, and

$$\tau = \inf\Big\{ t \geqslant 0 : \|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \geqslant K \Big\} \wedge T\,.$$

As before,

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\,\mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

which implies

$$\mathbb{E}_{\mu^\tau}[\,\mathrm{Var}_{\mu^\tau}(f)\,] \geqslant e^{-KT}\mathrm{Var}_\mu(f)$$

High-probability covariance bound $\Rightarrow \mathbb{P}(\tau < T) \leqslant \delta$. So

$$\mathbb{E}_{\mu^\tau}[\,\mathrm{Var}_{\mu^\tau}(f)\,] \leqslant \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] + \mathbb{E}_{\mu^\tau}[\mathbf{1}\{\tau < T\}\,\mathrm{Var}_{\mu^\tau}(f)\,]$$

Random tilts covariance bound whp $\Rightarrow$ WPI

Same strategy, but with a stopping time. Let $T, K = O(1)$ large, and

$$\tau = \inf\Big\{ t \geqslant 0 : \|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \geqslant K \Big\} \wedge T\,.$$

As before,

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\,\mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

which implies

$$\mathbb{E}_{\mu^\tau}[\,\mathrm{Var}_{\mu^\tau}(f)\,] \geqslant e^{-KT}\mathrm{Var}_\mu(f)$$

High-probability covariance bound $\Rightarrow \mathbb{P}(\tau < T) \leqslant \delta$. So

$$\begin{aligned}\mathbb{E}_{\mu^\tau}[\,\mathrm{Var}_{\mu^\tau}(f)\,] &\leqslant \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] + \mathbb{E}_{\mu^\tau}[\mathbf{1}\{\tau < T\}\,\mathrm{Var}_{\mu^\tau}(f)\,] \\ &\leqslant \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] + \delta\|f\|_\infty^2\end{aligned}$$

Random tilts covariance bound whp $\Rightarrow$ WPI

Same strategy, but with a stopping time. Let $T, K = O(1)$ large, and

$$\tau = \inf\left\{t \geqslant 0 : \|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \geqslant K\right\} \wedge T\,.$$

As before,

$$\mathrm{d}\mathrm{Var}_{\mu^t}(f) \geqslant -\|\mathrm{cov}(\mu^t)\|_{\mathrm{op}} \mathrm{Var}_{\mu^t}(f)\,\mathrm{d}t + \mathrm{d}(\mathrm{martingale})_t$$

which implies

$$\mathbb{E}_{\mu^\tau}[\,\mathrm{Var}_{\mu^\tau}(f)\,] \geqslant e^{-KT}\mathrm{Var}_\mu(f)$$

High-probability covariance bound $\Rightarrow \mathbb{P}(\tau < T) \leqslant \delta$. So

$$\begin{aligned}\mathbb{E}_{\mu^\tau}[\,\mathrm{Var}_{\mu^\tau}(f)\,] &\leqslant \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] + \mathbb{E}_{\mu^\tau}[\mathbf{1}\{\tau < T\}\,\mathrm{Var}_{\mu^\tau}(f)\,] \\ &\leqslant \mathbb{E}_{\mu^T}[\,\mathrm{Var}_{\mu^T}(f)\,] + \delta\|f\|_\infty^2\end{aligned}$$

Retracing the same steps then proves the WPI.

Outline of talk:

Introduction and motivating problems

Results: sampling from spin glasses

Algorithmic stochastic localization

Sampling guarantee for simulated annealing

Conclusion

Recap: SL and reduced objectives

Goal: sample from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto \exp(\beta H(\boldsymbol{x}))$

SL: nature generates secret $\boldsymbol{x}_\star \sim \mu$, we observe $\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t$. Posterior:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x} \rangle)$$

Recap: SL and reduced objectives

Goal: sample from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto \exp(\beta H(\boldsymbol{x}))$

SL: nature generates secret $\boldsymbol{x}_\star \sim \mu$, we observe $\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t$. Posterior:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x} \rangle)$$

Reduced objectives

To sample with algorithmic SL with total variation error $o(1)$, need estimator $\widehat{\boldsymbol{m}}_t$ of $\boldsymbol{m}_t = \mathrm{mean}(\mu_\beta^t)$ with accuracy

$$\mathbb{E}[\|\widehat{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2] = o(1) \qquad \text{for all } t$$

Recap: SL and reduced objectives

Goal: sample from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto \exp(\beta H(\boldsymbol{x}))$

SL: nature generates secret $\boldsymbol{x}_\star \sim \mu$, we observe $\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t$. Posterior:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x} \rangle)$$

Reduced objectives

To sample with algorithmic SL with total variation error $o(1)$, need estimator $\widehat{\boldsymbol{m}}_t$ of $\boldsymbol{m}_t = \mathrm{mean}(\mu_\beta^t)$ with accuracy

$$\mathbb{E}[\|\widehat{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2] = o(1) \qquad \text{for all } t$$

To show WPI / simulated annealing succeeds, need to prove

$$\|\mathrm{cov}(\mu_{\beta'}^t)\|_{\mathrm{op}} = O(1) \text{ whp} \qquad \text{for all } \beta' \leqslant \beta, \text{ all } t$$

Recap: SL and reduced objectives

Goal: sample from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto \exp(\beta H(\boldsymbol{x}))$

SL: nature generates secret $\boldsymbol{x}_\star \sim \mu$, we observe $\boldsymbol{y}_t = t\boldsymbol{x}_\star + \boldsymbol{B}_t$. Posterior:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x} \rangle)$$

Reduced objectives

To sample with algorithmic SL with total variation error $o(1)$, need estimator $\widehat{\boldsymbol{m}}_t$ of $\boldsymbol{m}_t = \mathrm{mean}(\mu_\beta^t)$ with accuracy

$$\mathbb{E}[\|\widehat{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2] = o(1) \qquad \text{for all } t$$

To show WPI / simulated annealing succeeds, need to prove

$$\|\mathrm{cov}(\mu_{\beta'}^t)\|_{\mathrm{op}} = O(1) \text{ whp} \qquad \text{for all } \beta' \leqslant \beta, \text{ all } t$$

Most of the technical content of our papers is to establish these inputs.

One proof idea: re-centering measure with external field

Time t measure μ_β^t has external field:

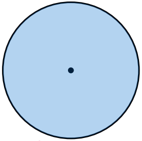
$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp\left(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x}\rangle\right)$$

One proof idea: re-centering measure with external field

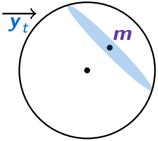
Time t measure μ_β^t has external field:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp\Big(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x}\rangle\Big)$$

Gibbs measure concentrates near random codimension-1 band



no field: μ_β^t has mean $\mathbf{0}$ and spread "in all directions"



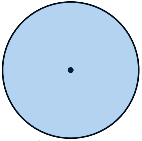
with field: μ_β^t concentrates on band with **random** center

One proof idea: re-centering measure with external field

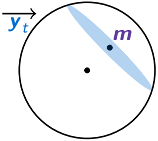
Time t measure μ_β^t has external field:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp\Big(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x}\rangle\Big)$$

Gibbs measure concentrates near random codimension-1 band



no field: μ_β^t has mean **0** and spread "in all directions"



with field: μ_β^t concentrates on band with **random** center

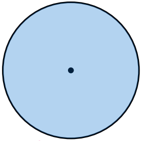
Reduction: restriction to band "looks like" no-field model

One proof idea: re-centering measure with external field

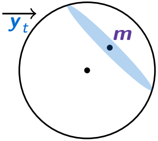
Time t measure μ_β^t has external field:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp\Big(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x}\rangle\Big)$$

Gibbs measure concentrates near random codimension-1 band



no field: μ_β^t has mean $\mathbf{0}$ and spread "in all directions"



with field: μ_β^t concentrates on band with **random** center

Reduction: restriction to band "looks like" no-field model

We develop method to "condition on $\boldsymbol{m}$," which solves **TAP equation**

One proof idea: re-centering measure with external field

Time t measure μ_β^t has external field:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp\Big(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x}\rangle\Big)$$

Gibbs measure concentrates near random codimension-1 band

no field: μ_β^t has mean $\mathbf{0}$ and spread "in all directions"

with field: μ_β^t concentrates on band with **random** center

Reduction: restriction to band "looks like" no-field model

We develop method to "condition on $\boldsymbol{m}$," which solves **TAP equation**

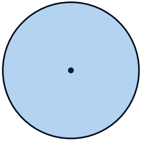
$$\nabla H(\boldsymbol{m}) + \boldsymbol{y}_t = \left(\frac{1}{1-\|\boldsymbol{m}\|^2/N} + (1-\tfrac{\|\boldsymbol{m}\|^2}{N})\xi''(\tfrac{\|\boldsymbol{m}\|^2}{N})\right)\boldsymbol{m}$$

One proof idea: re-centering measure with external field

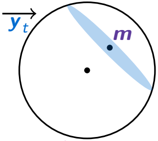
Time t measure μ_β^t has external field:

$$\mu_\beta^t(\mathrm{d}\boldsymbol{x}) \propto \exp\Big(\beta H(\boldsymbol{x}) + \langle \boldsymbol{y}_t, \boldsymbol{x}\rangle\Big)$$

Gibbs measure concentrates near random codimension-1 band



no field: μ_β^t has mean **0** and spread "in all directions"



with field: μ_β^t concentrates on band with **random** center

Reduction: restriction to band "looks like" no-field model

We develop method to "condition on $\boldsymbol{m}$," which solves **TAP equation**

$$\nabla H(\boldsymbol{m}) + \boldsymbol{y}_t = \left(\frac{1}{1-\|\boldsymbol{m}\|^2/N} + \left(1-\tfrac{\|\boldsymbol{m}\|^2}{N}\right)\xi''\left(\tfrac{\|\boldsymbol{m}\|^2}{N}\right)\right)\boldsymbol{m}$$

Mean estimator: find $\boldsymbol{m}$ with AMP + GD, plus nontrivial correction

Hardness results

Consider generalization of SL: observe $(\boldsymbol{x}_\star, \boldsymbol{x}_\star^{\otimes 2}, \boldsymbol{x}_\star^{\otimes 3}, \ldots)$ through independent gaussian channels, with varying noise levels.

Hardness results

Consider generalization of SL: observe $(\boldsymbol{x}_\star, \boldsymbol{x}_\star^{\otimes 2}, \boldsymbol{x}_\star^{\otimes 3}, \ldots)$ through independent gaussian channels, with varying noise levels.

Formally, observe $(\boldsymbol{y}_t^{(1)}, \boldsymbol{y}_t^{(2)}, \boldsymbol{y}_t^{(3)}, \ldots)$ where

$$\boldsymbol{y}_t^{(p)} = \lambda_p(t)\boldsymbol{x}_\star^{\otimes p} + \boldsymbol{B}_{\lambda_p(t)}^{(p)} \qquad \boldsymbol{B}_t^{(p)} \text{ std } \mathbb{R}^{N^p}\text{-Brownian motion}$$

Hardness results

Consider generalization of SL: observe $(\boldsymbol{x}_\star, \boldsymbol{x}_\star^{\otimes 2}, \boldsymbol{x}_\star^{\otimes 3}, \ldots)$ through independent gaussian channels, with varying noise levels.

Formally, observe $(\boldsymbol{y}_t^{(1)}, \boldsymbol{y}_t^{(2)}, \boldsymbol{y}_t^{(3)}, \ldots)$ where

$$\boldsymbol{y}_t^{(p)} = \lambda_p(t)\boldsymbol{x}_\star^{\otimes p} + \boldsymbol{B}_{\lambda_p(t)}^{(p)} \qquad \boldsymbol{B}_t^{(p)} \text{ std } \mathbb{R}^{N^p}\text{-Brownian motion}$$

Theorem (H Montanari Pham 24)

If $\beta > \beta_{\mathsf{SL}}$, even this generalized algorithmic SL fails (because AMP fails to estimate mean(μ^t)).

Hardness results

Consider generalization of SL: observe $(\boldsymbol{x}_\star, \boldsymbol{x}_\star^{\otimes 2}, \boldsymbol{x}_\star^{\otimes 3}, \ldots)$ through independent gaussian channels, with varying noise levels.

Formally, observe $(\boldsymbol{y}_t^{(1)}, \boldsymbol{y}_t^{(2)}, \boldsymbol{y}_t^{(3)}, \ldots)$ where

$$\boldsymbol{y}_t^{(p)} = \lambda_p(t)\boldsymbol{x}_\star^{\otimes p} + \boldsymbol{B}^{(p)}_{\lambda_p(t)} \qquad \boldsymbol{B}_t^{(p)} \text{ std } \mathbb{R}^{N^p}\text{-Brownian motion}$$

Theorem (H Montanari Pham 24)

If $\beta > \beta_{\mathsf{SL}}$, even this generalized algorithmic SL fails (because AMP fails to estimate mean(μ^t)).

Mechanism: over SL process $(\mu_\beta^t)_{t \geqslant 0}$, mean$(\mu_\beta^t)$ can move violently, which AMP can't track

Hardness results

Consider generalization of SL: observe $(\boldsymbol{x}_\star, \boldsymbol{x}_\star^{\otimes 2}, \boldsymbol{x}_\star^{\otimes 3}, \dots)$ through independent gaussian channels, with varying noise levels.

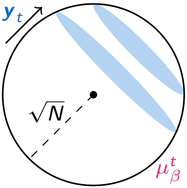
Formally, observe $(\boldsymbol{y}_t^{(1)}, \boldsymbol{y}_t^{(2)}, \boldsymbol{y}_t^{(3)}, \dots)$ where

$$\boldsymbol{y}_t^{(p)} = \lambda_p(t)\boldsymbol{x}_\star^{\otimes p} + \boldsymbol{B}_{\lambda_p(t)}^{(p)} \qquad \boldsymbol{B}_t^{(p)} \text{ std } \mathbb{R}^{N^p}\text{-Brownian motion}$$

Theorem (H Montanari Pham 24)

If $\beta > \beta_{\mathsf{SL}}$, even this generalized algorithmic SL fails (because AMP fails to estimate mean(μ^t)).

Mechanism: over SL process $(\mu_\beta^t)_{t \geqslant 0}$, mean($\mu_\beta^t$) can move violently, which AMP can't track



Hardness results

Consider generalization of SL: observe $(\boldsymbol{x}_\star, \boldsymbol{x}_\star^{\otimes 2}, \boldsymbol{x}_\star^{\otimes 3}, \dots)$ through independent gaussian channels, with varying noise levels.

Formally, observe $(\boldsymbol{y}_t^{(1)}, \boldsymbol{y}_t^{(2)}, \boldsymbol{y}_t^{(3)}, \dots)$ where

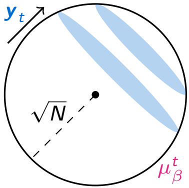
$$\boldsymbol{y}_t^{(p)} = \lambda_p(t)\boldsymbol{x}_\star^{\otimes p} + \boldsymbol{B}_{\lambda_p(t)}^{(p)} \qquad \boldsymbol{B}_t^{(p)} \text{ std } \mathbb{R}^{N^p}\text{-Brownian motion}$$

Theorem (H Montanari Pham 24)

If $\beta > \beta_{\mathsf{SL}}$, even this generalized algorithmic SL fails (because AMP fails to estimate mean(μ^t)).

Mechanism: over SL process $(\mu_\beta^t)_{t\geqslant 0}$, mean$(\mu_\beta^t)$ can move violently, which AMP can't track

This also breaks WPI proof: $\|\mathrm{cov}(\mu_\beta^t)\|_{\mathrm{op}} \gg 1$ when μ_β^t straddles two bands



Conclusion

We show two algorithms sample from a spin glass Gibbs measure μ_β in total variation, for all $\beta < \beta_{\mathsf{SL}}$

Conclusion

We show two algorithms sample from a spin glass Gibbs measure μ_β in total variation, for all $\beta < \beta_{\mathsf{SL}}$

- Alg SL: simulate SL process $(\mu^t)_{t \geqslant 0}$ with improved mean estimator

Conclusion

We show two algorithms sample from a spin glass Gibbs measure μ_β in total variation, for all $\beta < \beta_{\mathsf{SL}}$

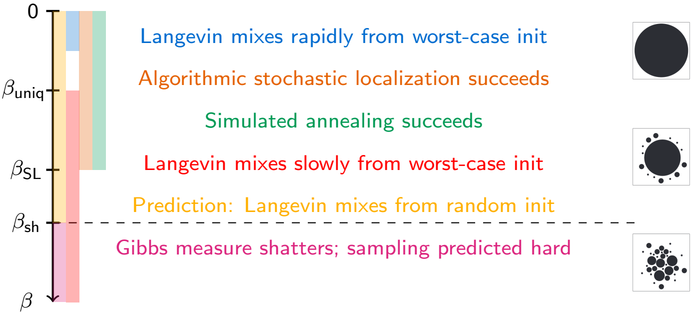
- Alg SL: simulate SL process $(\mu^t)_{t \geqslant 0}$ with improved mean estimator
- Sim annealing: whp covariance bound $\Rightarrow$ weak Poincaré inequality $\Rightarrow$ sampling guarantee

Conclusion

We show two algorithms sample from a spin glass Gibbs measure μ_β in total variation, for all $\beta < \beta_{\mathsf{SL}}$

- Alg SL: simulate SL process $(\mu^t)_{t \geqslant 0}$ with improved mean estimator
- Sim annealing: whp covariance bound $\Rightarrow$ weak Poincaré inequality $\Rightarrow$ sampling guarantee

- Open Q: sample up to shattering threshold β_{sh}

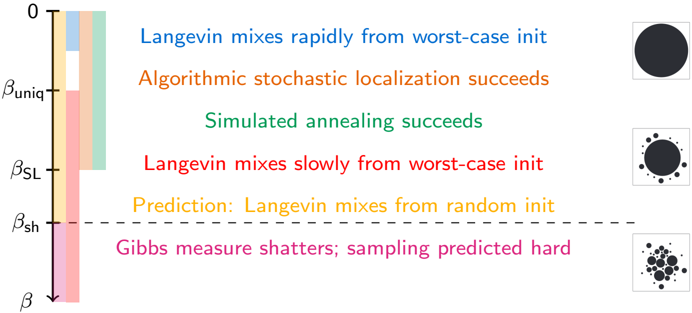


Conclusion

We show two algorithms sample from a spin glass Gibbs measure μ_β in total variation, for all $\beta < \beta_{\mathsf{SL}}$

- Alg SL: simulate SL process $(\mu^t)_{t \geqslant 0}$ with improved mean estimator
- Sim annealing: whp covariance bound $\Rightarrow$ weak Poincaré inequality $\Rightarrow$ sampling guarantee

- Open Q: sample up to shattering threshold β_{sh}



Thank you!