Weak Poincaré inequalities, simulated annealing, and sampling from spherical spin glasses

Brice Huang
(MIT→Stanford→Yale)

Sidhanth Mohanty
(MIT→Northwestern)

Amit Rajaraman
(MIT)

David X. Wu
(UC Berkeley)

INFORMS APS 2025: Mean-field and statistical physics models
in modern high-dimensional statistics

Motivation

Motivating problem

Sample from a high-dimensional distribution $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}$ with an efficient algorithm.

Motivation

Motivating problem

Sample from a high-dimensional distribution $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}\ \mathrm{d}\boldsymbol{x}$ with an efficient algorithm.

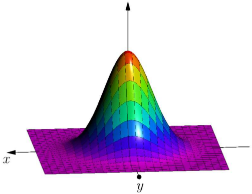
Useful in: inference, integration, optimization, ...

Motivation

Motivating problem

Sample from a high-dimensional distribution $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}$ with an efficient algorithm.

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



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

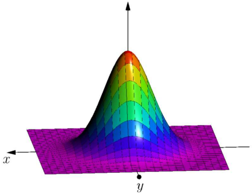
Sampling is easy!

Motivation

Motivating problem

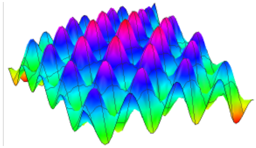
Sample from a high-dimensional distribution $\mu(\mathbf{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}\,\mathbf{d}\boldsymbol{x}$ with an efficient algorithm.

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

Our model: spherical mean-field spin glass

Our model: spherical mean-field spin glass

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

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

Our model: spherical mean-field spin glass

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

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

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

Our model: spherical mean-field spin glass

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

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

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

- $\beta = 0$: $\mu_\beta = \text{uniform}(S_N)$. Easy to sample.

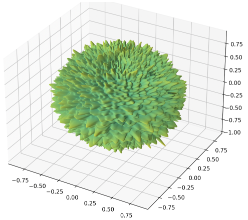
Our model: spherical mean-field spin glass

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

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

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

- $\beta = 0$: $\mu_\beta = \mathrm{uniform}(S_N)$. Easy to sample.
- β large: highly non-log-concave. H has exponentially many local maxima. Sampling seems hard!



Langevin dynamics

Canonical Markov chain for continuous distributions:

$$\mathrm{d}X_t = \nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

Langevin dynamics

Canonical Markov chain for continuous distributions:

$$\mathrm{d}X_t = \nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

$$X_{t+\delta} - X_t = \delta \cdot \nabla V(X_t) + \sqrt{2\delta} \cdot \underset{\substack{\uparrow \\ \mathcal{N}(0, I_N)}}{g_t}$$

Langevin dynamics

Canonical Markov chain for continuous distributions:

$$\mathrm{d}X_t = \nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

$$X_{t+\delta} - X_t = \delta \cdot \nabla V(X_t) + \sqrt{2\delta} \cdot \underset{\substack{\uparrow \\ \mathcal{N}(0, I_N)}}{g_t}$$

Stationary distribution $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.

Langevin dynamics

Canonical Markov chain for continuous distributions:

$$\mathrm{d}X_t = \nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t$$

$$X_{t+\delta} - X_t = \delta \cdot \nabla V(X_t) + \sqrt{2\delta} \cdot \underset{\substack{\uparrow \\ \mathcal{N}(0,I_N)}}{g_t}$$

Stationary distribution $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.

Yields an efficient sampler if Markov chain **mixes rapidly**, i.e.

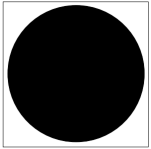
$$d_{\mathsf{TV}}(\mathsf{Law}(X_T), \mu) = o(1) \qquad \text{for } T = \mathrm{poly}(N)\,.$$

Mixing guarantees: worst-case vs. random initialization

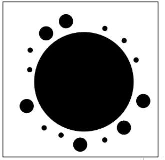
Ideal guarantee: rapid mixing from **worst-case** initialization. But not always possible!

Mixing guarantees: worst-case vs. random initialization

Ideal guarantee: rapid mixing from **worst-case** initialization. But not always possible!



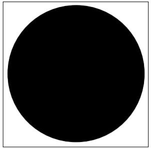
μ well-connected



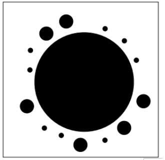
μ well-connected except
$o(1)$ mass in **metastable states**

Mixing guarantees: worst-case vs. random initialization

Ideal guarantee: rapid mixing from **worst-case** initialization. But not always possible!



μ well-connected

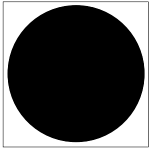


μ well-connected except
$o(1)$ mass in **metastable states**

Mixes from worst-case init

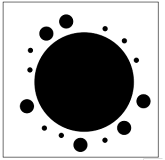
Mixing guarantees: worst-case vs. random initialization

Ideal guarantee: rapid mixing from **worst-case** initialization. But not always possible!



μ well-connected

Mixes from worst-case init

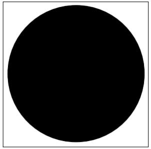


μ well-connected except
$o(1)$ mass in **metastable states**

Fails to mix from worst-case init

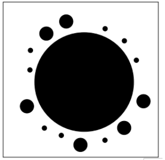
Mixing guarantees: worst-case vs. random initialization

Ideal guarantee: rapid mixing from **worst-case** initialization. But not always possible!



μ well-connected

Mixes from worst-case init

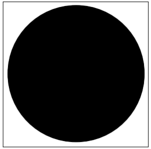


μ well-connected except
$o(1)$ mass in **metastable states**

Fails to mix from worst-case init
Can we show mixing from random init?

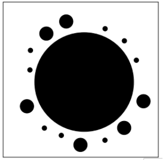
Mixing guarantees: worst-case vs. random initialization

Ideal guarantee: rapid mixing from **worst-case** initialization. But not always possible!



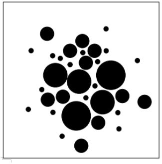
μ well-connected

Mixes from worst-case init



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

Fails to mix from worst-case init
Can we show mixing from random init?



μ **shattered**

Fails to mix from any init

Simulated annealing

Simulated annealing

- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.

Simulated annealing

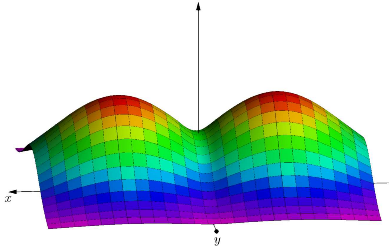
- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$.

Simulated annealing

- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$. $\beta \leftrightarrow$ "artificial" inverse temperature.

Simulated annealing

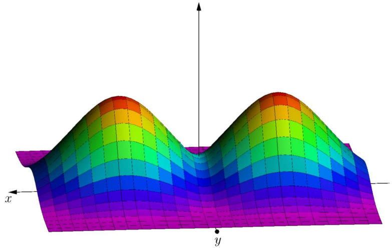
- Target distribution: $\mu(\mathbf{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$. $\beta \leftrightarrow$ "artificial" inverse temperature.
- Algorithm: run Langevin dynamics for μ_β, and slowly increase β from 0 to 1.



$\beta = 0.1$

Simulated annealing

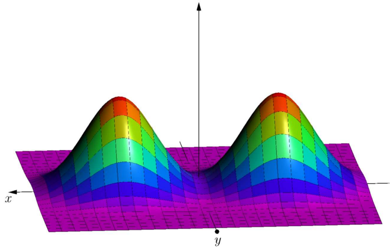
- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$. $\beta \leftrightarrow$ "artificial" inverse temperature.
- Algorithm: run Langevin dynamics for μ_β, and slowly increase β from 0 to 1.



$\beta = 0.2$

Simulated annealing

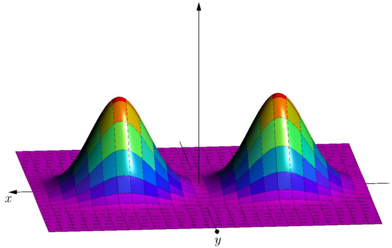
- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$. $\beta \leftrightarrow$ "artificial" inverse temperature.
- Algorithm: run Langevin dynamics for μ_β, and slowly increase β from 0 to 1.



$\beta = 0.4$

Simulated annealing

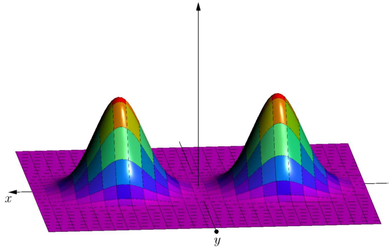
- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$. $\beta \leftrightarrow$ “artificial” inverse temperature.
- Algorithm: run Langevin dynamics for μ_β, and slowly increase β from 0 to 1.



$\beta = 0.6$

Simulated annealing

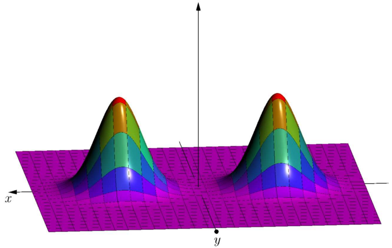
- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$. $\beta \leftrightarrow$ "artificial" inverse temperature.
- Algorithm: run Langevin dynamics for μ_β, and slowly increase β from 0 to 1.



$\beta = 0.8$

Simulated annealing

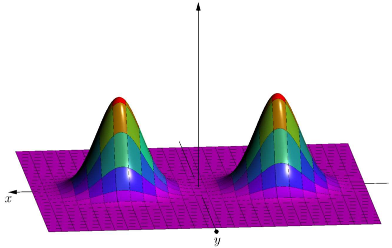
- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$. $\beta \leftrightarrow$ "artificial" inverse temperature.
- Algorithm: run Langevin dynamics for μ_β, and slowly increase β from 0 to 1.



$\beta = 1.0$

Simulated annealing

- Target distribution: $\mu(\mathrm{d}\boldsymbol{x}) \propto e^{V(\boldsymbol{x})}$.
- For $0 \leqslant \beta \leqslant 1$, define intermediate distributions $\mu_\beta(x) \propto e^{\beta V(\boldsymbol{x})}$. $\beta \leftrightarrow$ "artificial" inverse temperature.
- Algorithm: run Langevin dynamics for μ_β, and slowly increase β from 0 to 1.

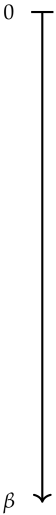


$\beta = 1.0$

- Natural algorithm, but rigorous guarantees scarce.

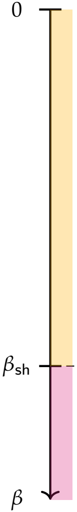
Main result: sampling for spherical mean-field spin glass

For sampling from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})} \equiv e^{V(\boldsymbol{x})}$:



Main result: sampling for spherical mean-field spin glass

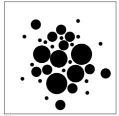
For sampling from Gibbs measure $\mu_\beta(\mathbf{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})} \equiv e^{V(\boldsymbol{x})}$:



Prediction: Langevin mixes from random init

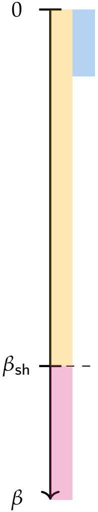
Gibbs measure shatters; sampling predicted hard

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

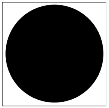


Main result: sampling for spherical mean-field spin glass

For sampling from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})} \equiv e^{V(\boldsymbol{x})}$:

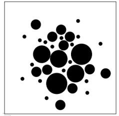


Langevin mixes rapidly from worst-case init
(Gheissari Jagannath 19, Eldan Koehler Zeitouni 22,
Anari Jain Pham Koehler Vuong 24, Anari Koehler Vuong 24)



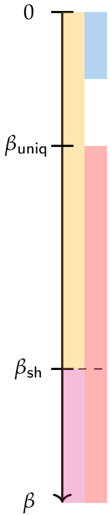
Prediction: Langevin mixes from random init

Gibbs measure shatters; sampling predicted hard
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23,
Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)



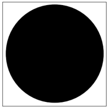
Main result: sampling for spherical mean-field spin glass

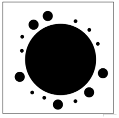
For sampling from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})} \equiv e^{V(\boldsymbol{x})}$:



Langevin mixes rapidly from worst-case init

(Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, Anari Koehler Vuong 24)





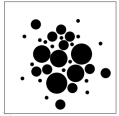
Langevin mixes slowly from worst-case init

(Ben Arous Jagannath 24)

Prediction: Langevin mixes from random init

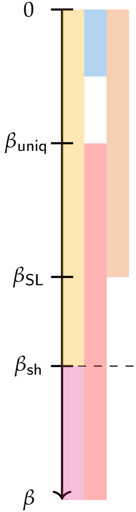
Gibbs measure shatters; sampling predicted hard

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

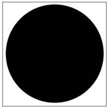


Main result: sampling for spherical mean-field spin glass

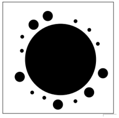
For sampling from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})} \equiv e^{V(\boldsymbol{x})}$:



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



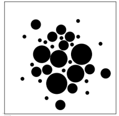
Algorithmic stochastic localization succeeds
(El Alaoui Montanari Sellke 22, **H.** Montanari Pham 24)



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

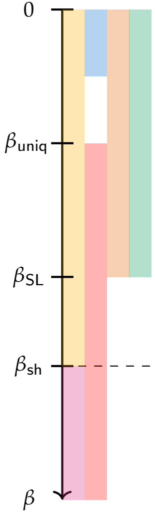
Prediction: Langevin mixes from random init

Gibbs measure shatters; sampling predicted hard
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23,
Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

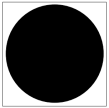


Main result: sampling for spherical mean-field spin glass

For sampling from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{x}) \propto e^{\beta H(\boldsymbol{x})} \equiv e^{V(\boldsymbol{x})}$:

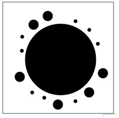


Langevin mixes rapidly from worst-case init
(Gheissari Jagannath 19, Eldan Koehler Zeitouni 22,
Anari Jain Pham Koehler Vuong 24, Anari Koehler Vuong 24)



Algorithmic stochastic localization succeeds
(El Alaoui Montanari Sellke 22, **H.** Montanari Pham 24)

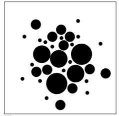
This work: simulated annealing succeeds



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

Prediction: Langevin mixes from random init

Gibbs measure shatters; sampling predicted hard
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23,
Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)



How to prove worst-case rapid mixing

Show the Markov chain P contracts distance to stationarity:

$$\mathbf{distance}(P_t \nu \| \mu) \leqslant e^{-Ct} \cdot \mathbf{distance}(\nu \| \mu) \qquad \text{for all distributions } \nu$$

How to prove worst-case rapid mixing

Show the Markov chain P contracts distance to stationarity:

$$\mathbf{distance}(P_t \nu \| \mu) \leqslant e^{-Ct} \cdot \mathbf{distance}(\nu \| \mu) \qquad \text{for all distributions } \nu$$

where $C \geqslant 1/\operatorname{poly}(N)$.

How to prove worst-case rapid mixing

Show the Markov chain P contracts distance to stationarity:

$$\mathbf{distance}(P_t \nu \| \mu) \leqslant e^{-Ct} \cdot \mathbf{distance}(\nu \| \mu) \qquad \text{for all distributions } \nu$$

where $C \geqslant 1/\operatorname{poly}(N)$.

There are by now many ways to show this:

- Coupling (Bubley Dyer Jerrum 96)
- Path coupling (Jerrum 95, Bubley Dyer 97)
- Canonical paths (Jerrum Sinclair 89)
- Curvature / Bakry–Émery theory (Bakry Émery 06, Villani 09)
- Zero-freeness (Lee Yang 52, Barvinok 16, Chen Liu Vigoda 24)
- Correlation decay (Dyer Sinclair Vigoda Weitz 04, Weitz 06)
- Spectral independence (Anari Liu Oveis Gharan Vinzant 19)
- Entropic independence (Anari Jain Pham Koehler Vuong 21)
- Stochastic localization (Eldan 13, Eldan Koehler Zeitouni 22)

How to prove worst-case rapid mixing

Show the Markov chain P contracts distance to stationarity:

$$\mathrm{distance}(P_t\nu\|\mu) \leqslant e^{-Ct} \cdot \mathrm{distance}(\nu\|\mu) \qquad \text{for all distributions } \nu$$

where $C \geqslant 1/\operatorname{poly}(N)$.

There are by now many ways to show this:

- Coupling (Bubley Dyer Jerrum 96)
- Path coupling (Jerrum 95, Bubley Dyer 97)
- Canonical paths (Jerrum Sinclair 89)
- Curvature / Bakry–Émery theory (Bakry Émery 06, Villani 09)
- Zero-freeness (Lee Yang 52, Barvinok 16, Chen Liu Vigoda 24)
- Correlation decay (Dyer Sinclair Vigoda Weitz 04, Weitz 06)
- Spectral independence (Anari Liu Oveis Gharan Vinzant 19)
- Entropic independence (Anari Jain Pham Koehler Vuong 21)
- Stochastic localization (Eldan 13, Eldan Koehler Zeitouni 22)

Localization schemes

Worst-case mixing via Poincaré inequalities

If μ satisfies a **Poincaré inequality**, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\chi^2(P_t\nu\|\mu) \leqslant -C\chi^2(P_t\nu\|\mu) \qquad \text{for all distributions } \nu.$$

Worst-case mixing via Poincaré inequalities

If μ satisfies a **Poincaré inequality**, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\chi^2(P_t\nu\|\mu) \leqslant -C\chi^2(P_t\nu\|\mu) \qquad \text{for all distributions } \nu.$$

Implies exponential contraction of χ^2 divergence:

$$\chi^2(P_t\nu\|\mu) \leqslant e^{-Ct} \cdot \chi^2(\nu\|\mu)\,.$$

Worst-case mixing via Poincaré inequalities

If μ satisfies a **Poincaré inequality**, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\chi^2(P_t\nu\|\mu) \leqslant -C\chi^2(P_t\nu\|\mu) \qquad \text{for all distributions } \nu.$$

Implies exponential contraction of χ^2 divergence:

$$\chi^2(P_t\nu\|\mu) \leqslant e^{-Ct}\cdot\chi^2(\nu\|\mu)\,.$$

Spectral independence / **localization schemes** provide powerful framework for showing a PI holds. (Anari Liu Oveis Gharan 21, Chen Eldan 22)

Weak Poincaré inequalities

A **weak Poincaré inequality** implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\chi^2(P_t\nu\|\mu) \leqslant -C\chi^2(P_t\nu\|\mu) + \delta\left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_\infty^2 \qquad \text{for all distributions } \nu.$$

Error term captures contribution of metastable states.

Weak Poincaré inequalities

A **weak Poincaré inequality** implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\chi^2(P_t\nu\|\mu) \leqslant -C\chi^2(P_t\nu\|\mu) + \delta\left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_\infty^2 \qquad \text{for all distributions } \nu.$$

Error term captures contribution of metastable states.

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

Weak Poincaré inequality → simulated annealing succeeds

Consequence of weak Poincaré inequality

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

Simulated annealing

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

Weak Poincaré inequality $\to$ simulated annealing succeeds

Consequence of weak Poincaré inequality

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

Simulated annealing

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

In stage $t = 0, \ldots, T$:

Weak Poincaré inequality → simulated annealing succeeds

Consequence of weak Poincaré inequality

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

Simulated annealing

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

In stage $t = 0, \ldots, T$:

- Set $\beta_t = \beta \cdot t/T$. Interpolates from $\beta_0 = 0$ to $\beta_T = \beta$.

Weak Poincaré inequality → simulated annealing succeeds

Consequence of weak Poincaré inequality

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

Simulated annealing

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

In stage $t = 0, \ldots, T$:

- Set $\beta_t = \beta \cdot t/T$. Interpolates from $\beta_0 = 0$ to $\beta_T = \beta$.
- Run Langevin dynamics for $\mu_t(\mathrm{d}\boldsymbol{x}) \propto e^{\beta_t H(\boldsymbol{x})}$, initialized at output of previous stage.

Weak Poincaré inequality → simulated annealing succeeds

Consequence of weak Poincaré inequality

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

Simulated annealing

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

In stage $t = 0, \dots, T$:

- Set $\beta_t = \beta \cdot t/T$. Interpolates from $\beta_0 = 0$ to $\beta_T = \beta$.
- Run Langevin dynamics for $\mu_t(\mathrm{d}\boldsymbol{x}) \propto e^{\beta_t H(\boldsymbol{x})}$, initialized at output of previous stage.

μ_t is a warm start for μ_{t+1}! WPI lets us inductively argue each stage mixes.

Weak Poincaré inequality → simulated annealing succeeds

Consequence of weak Poincaré inequality

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

Simulated annealing

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

In stage $t = 0, \ldots, T$:

- Set $\beta_t = \beta \cdot t/T$. Interpolates from $\beta_0 = 0$ to $\beta_T = \beta$.
- Run Langevin dynamics for $\mu_t(\mathbf{d}\boldsymbol{x}) \propto e^{\beta_t H(\boldsymbol{x})}$, initialized at output of previous stage.

💡 μ_t is a warm start for μ_{t+1}! WPI lets us inductively argue each stage mixes.
See El Alaoui Eldan Gheissari Piana 23 for related “staging warm starts” idea.

How to show a weak Poincaré inequality

For $\boldsymbol{h} \in \mathbb{R}^N$, define tilted measure

$$\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle} \, \mu(\mathrm{d}\boldsymbol{x}) \, .$$

How to show a weak Poincaré inequality

For $\boldsymbol{h} \in \mathbb{R}^N$, define tilted measure

$$\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle} \mu(\mathrm{d}\boldsymbol{x}) \, .$$

Key message of spectral independence + localization schemes (Anari Liu Oveis Gharan 21, ...)

μ satisfies a Poincaré inequality if

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

How to show a weak Poincaré inequality

For $\boldsymbol{h} \in \mathbb{R}^N$, define tilted measure

$$\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle} \mu(\mathrm{d}\boldsymbol{x}) .$$

Key message of spectral independence + localization schemes (Anari Liu Oveis Gharan 21, ...)

μ satisfies a Poincaré inequality if

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

Message of our work

μ satisfies a **weak** Poincaré inequality if

$$\|\mathsf{cov}(\mu^{\boldsymbol{h}})\|_{\mathsf{op}} = O(1) \qquad \text{whp for } \textbf{random } \boldsymbol{h} \in \mathbb{R}^N .$$

How to show a weak Poincaré inequality

For $\boldsymbol{h} \in \mathbb{R}^N$, define tilted measure

$$\mu^{\boldsymbol{h}}(\mathrm{d}\boldsymbol{x}) \propto e^{\langle \boldsymbol{h}, \boldsymbol{x} \rangle} \mu(\mathrm{d}\boldsymbol{x}) .$$

Key message of spectral independence + localization schemes (Anari Liu Oveis Gharan 21, ...)

μ satisfies a Poincaré inequality if

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

Message of our work

μ satisfies a **weak** Poincaré inequality if

$$\|\mathsf{cov}(\mu^{\boldsymbol{h}})\|_{\mathsf{op}} = O(1) \qquad \text{whp for } \textbf{random } \boldsymbol{h} \in \mathbb{R}^N .$$

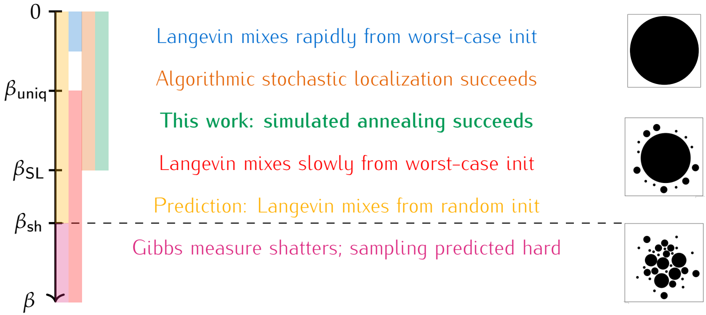
We show this covariance bound when $\mu =$ spherical spin glass Gibbs measure.

Conclusion

- Weak Poinaré inequalities imply sampling guarantees for simulated annealing.
- We introduce method to establish WPI + apply to spherical spin glass Gibbs measures.

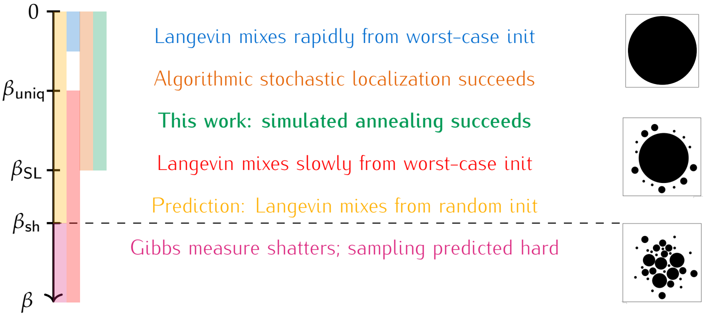
Conclusion

- Weak Poinaré inequalities imply sampling guarantees for simulated annealing.
- We introduce method to establish WPI + apply to spherical spin glass Gibbs measures.
- **Open Qs**: sample up to shattering threshold β_{sh}? Simulated annealing for models on $\{\pm 1\}^N$?



Conclusion

- Weak Poinaré inequalities imply sampling guarantees for simulated annealing.
- We introduce method to establish WPI + apply to spherical spin glass Gibbs measures.
- **Open Qs**: sample up to shattering threshold β_{sh}? Simulated annealing for models on $\{\pm 1\}^N$?



Thank you!