Sampling from spherical spin glasses in total variation via algorithmic stochastic localization

Brice Huang (MIT)

JSM: Advances in the theory of modern sampling algorithms
August 7, 2024

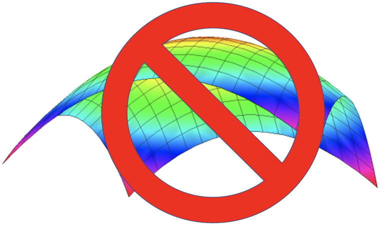
Joint work with Andrea Montanari and Huy Tuan Pham (Stanford)

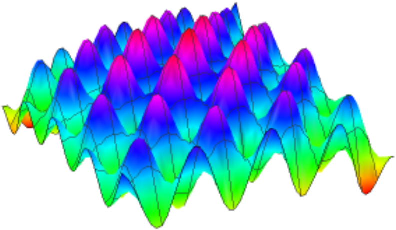
Goal of this talk

Sample from a **high-dimensional** measure

$$\boldsymbol{\sigma} \sim \mu(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{\beta H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}, \qquad \mu \in \mathcal{P}(\mathbb{R}^N)$$

where H_N **non-concave** and **highly multimodal**





Mean-field spin glasses

- Random polynomial over $\{\pm 1\}^N$ or $S_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N : \|\boldsymbol{\sigma}\| = \sqrt{N}\}$
- Sherrington-Kirkpatrick model:

$$H_N(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{i,j}\sigma_i\sigma_j \qquad g_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

Mean-field spin glasses

- Random polynomial over $\{\pm 1\}^N$ or $S_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N : \|\boldsymbol{\sigma}\| = \sqrt{N}\}$
- Sherrington-Kirkpatrick model:

$$H_N(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{i,j} \sigma_i \sigma_j \qquad g_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

- p-spin model: (SK: $p = 2$)

$$H_N(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{\underline{i} \in [N]^p} g_{\underline{i}} \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_p} \qquad g_{\underline{i}} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

Mean-field spin glasses

- Random polynomial over $\{\pm 1\}^N$ or $S_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N : \|\boldsymbol{\sigma}\| = \sqrt{N}\}$
- Sherrington-Kirkpatrick model:

$$H_N(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{i,j} \sigma_i \sigma_j \qquad g_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

- p-spin model: (SK: $p = 2$)

$$H_N(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{\underline{i} \in [N]^p} g_{\underline{i}} \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_p} \qquad g_{\underline{i}} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

- Mixed p-spin model:

$$H_N(\boldsymbol{\sigma}) = \sum_{p \ge 2} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{\underline{i} \in [N]^p} g_{\underline{i}} \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_p}$$

Mean-field spin glasses

- Random polynomial over $\{\pm 1\}^N$ or $S_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N : \|\boldsymbol{\sigma}\| = \sqrt{N}\}$
- Sherrington-Kirkpatrick model:

$$H_N(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{i,j} \sigma_i \sigma_j \qquad g_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

- p-spin model: (SK: $p = 2$)

$$H_N(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{\underline{i} \in [N]^p} g_{\underline{i}} \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_p} \qquad g_{\underline{i}} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

- Mixed p-spin model:

$$H_N(\boldsymbol{\sigma}) = \sum_{p \geq 2} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{\underline{i} \in [N]^p} g_{\underline{i}} \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_p}$$

- $\mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{\beta H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$ **Gibbs measure** at **inverse temperature** β

Connections

Spin glasses are prototypes for disordered, random probability measures:

- Posteriors in high-dimensional Bayesian inference
- Community detection, error-correcting codes, compressed sensing

Connections

Spin glasses are prototypes for disordered, random probability measures:

- Posteriors in high-dimensional Bayesian inference
- Community detection, error-correcting codes, compressed sensing

E.g. $\mathbb{Z}_2$-synchronization (Fan–Mei–Montanari 21, Montanari–Wu 23):

$$\text{Estimate } \boldsymbol{x}_0 \sim \text{unif}(\{\pm 1\}^N) \text{ from noisy observation } \boldsymbol{A} = \lambda \boldsymbol{x}_0^{\otimes 2} + \boldsymbol{W}$$

where $W_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$

Connections

Spin glasses are prototypes for disordered, random probability measures:

- Posteriors in high-dimensional Bayesian inference
- Community detection, error-correcting codes, compressed sensing

E.g. $\mathbb{Z}_2$-synchronization (Fan–Mei–Montanari 21, Montanari–Wu 23):

$$\text{Estimate } \boldsymbol{x}_0 \sim \mathrm{unif}(\{\pm 1\}^N) \text{ from noisy observation } \boldsymbol{A} = \lambda \boldsymbol{x}_0^{\otimes 2} + \boldsymbol{W}$$

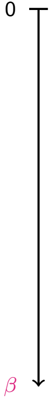
where $W_{i,j} \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$

Posterior is precisely

$$p(\boldsymbol{x}|\boldsymbol{A}) = \frac{1}{Z} \exp\left(\lambda \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle\right)$$

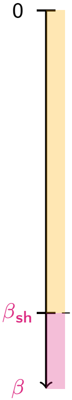
Background and informal result

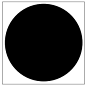
For sampling from Gibbs measure $\mu_\beta(d\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{\beta H_N(\boldsymbol{\sigma})} d\boldsymbol{\sigma}$:

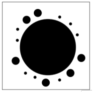


Background and informal result

For sampling from Gibbs measure $\mu_\beta(d\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{\beta H_N(\boldsymbol{\sigma})} d\boldsymbol{\sigma}$:



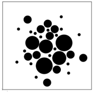




Prediction: Glauber/Langevin dynamics succeed

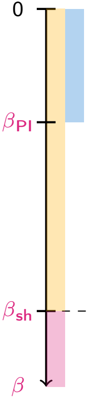
Prediction: Gibbs measure shatters, sampling hard

(Crisanti–Horner–Sommers 93, El Alaoui–Montanari–Sellke 23)



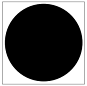
Background and informal result

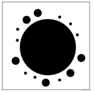
For sampling from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{\beta H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$:



Glauber/Langevin dynamics mix rapidly

(Gheissari–Jagannath 19, Bauerschmidt–Bodineau 20, Eldan–Koehler–Zeitouni 22, Anari–Jain–Koehler–Pham–Vuong 21)

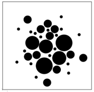




Prediction: Glauber/Langevin dynamics succeed

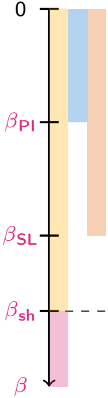
Prediction: Gibbs measure shatters, sampling hard

(Crisanti–Horner–Sommers 93, El Alaoui–Montanari–Sellke 23)



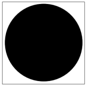
Background and informal result

For sampling from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{\beta H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$:



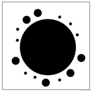
Glauber/Langevin dynamics mix rapidly

(Gheissari–Jagannath 19, Bauerschmidt–Bodineau 20, Eldan–Koehler–Zeitouni 22, Anari–Jain–Koehler–Pham–Vuong 21)



Algorithmic stochastic localization samples with small **Wasserstein** error

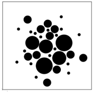
(El Alaoui–Montanari–Sellke 22, Celentano 24)



Prediction: Glauber/Langevin dynamics succeed

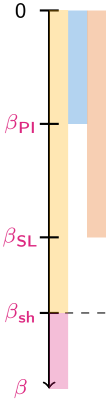
Prediction: Gibbs measure shatters, sampling hard

(Crisanti–Horner–Sommers 93, El Alaoui–Montanari–Sellke 23)



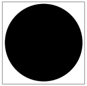
Background and informal result

For sampling from Gibbs measure $\mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{\beta H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$:



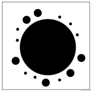
Glauber/Langevin dynamics mix rapidly

(Gheissari–Jagannath 19, Bauerschmidt–Bodineau 20, Eldan–Koehler–Zeitouni 22, Anari–Jain–Koehler–Pham–Vuong 21)



Algorithmic stochastic localization samples with small **Wasserstein** error

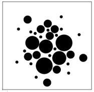
(El Alaoui–Montanari–Sellke 22, Celentano 24)



Prediction: Glauber/Langevin dynamics succeed

Prediction: Gibbs measure shatters, sampling hard

(Crisanti–Horner–Sommers 93, El Alaoui–Montanari–Sellke 23)



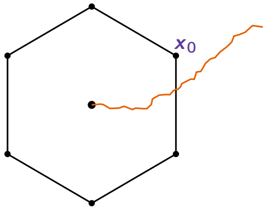
This work: in **spherical models**, $\beta < \beta_{\mathsf{SL}}$, alg-SL samples in **total variation**

Stochastic localization

Given $\mu \in \mathcal{P}(\mathbb{R}^N)$, consider Brownian motion with **unknown drift**

$$\boldsymbol{y}_t = t\boldsymbol{x}_0 + \boldsymbol{B}_t \qquad \sim \mathcal{N}(t\boldsymbol{x}_0, tI_N)$$

Here $\boldsymbol{x}_0 \sim \mu$ independent of $\boldsymbol{B}_t$, and **only $\boldsymbol{y}_t$ observed**

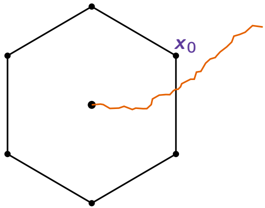


Stochastic localization

Given $\mu \in \mathcal{P}(\mathbb{R}^N)$, consider Brownian motion with **unknown drift**

$$\boldsymbol{y}_t = t\boldsymbol{x}_0 + \boldsymbol{B}_t \qquad \sim \mathcal{N}(t\boldsymbol{x}_0, tI_N)$$

Here $\boldsymbol{x}_0 \sim \mu$ independent of $\boldsymbol{B}_t$, and **only $\boldsymbol{y}_t$ observed**



The sampling strategy:

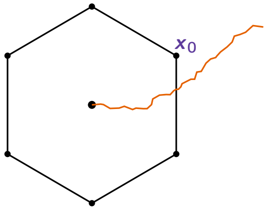
- **Simulate $\boldsymbol{y}_t$** for a long time $t \in [0, T]$ **without knowing $\boldsymbol{x}_0$**
- **Read off $\boldsymbol{x}_0 \approx \boldsymbol{y}_T / T$**

Stochastic localization

Given $\mu \in \mathcal{P}(\mathbb{R}^N)$, consider Brownian motion with **unknown drift**

$$\boldsymbol{y}_t = t\boldsymbol{x}_0 + \boldsymbol{B}_t \qquad \sim \mathcal{N}(t\boldsymbol{x}_0, tI_N)$$

Here $\boldsymbol{x}_0 \sim \mu$ independent of $\boldsymbol{B}_t$, and **only $\boldsymbol{y}_t$ observed**



The sampling strategy:

- **Simulate $\boldsymbol{y}_t$** for a long time $t \in [0, T]$ **without knowing $\boldsymbol{x}_0$**
- **Read off $\boldsymbol{x}_0 \approx \boldsymbol{y}_T / T$**

Reparametrization of "backward process" in denoising diffusions

Stochastic localization

Our goal is to simulate, **on average** over unobserved $\boldsymbol{x}_0 \sim \mu$,

$$\boldsymbol{y}_t = t\boldsymbol{x}_0 + \boldsymbol{B}_t$$

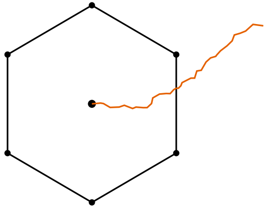
Stochastic localization

Our goal is to simulate, **on average** over unobserved $\boldsymbol{x}_0 \sim \mu$,

$$\boldsymbol{y}_t = t\boldsymbol{x}_0 + \boldsymbol{B}_t$$

Equivalent: Markovian SDE, with drift **current conditional expectation** of $\boldsymbol{x}_0$

$$\mathrm{d}\boldsymbol{y}_t = \boldsymbol{m}_t \mathrm{d}t + \mathrm{d}\boldsymbol{W}_t \qquad \boldsymbol{m}_t = \mathbb{E}[\boldsymbol{x}_0 | \boldsymbol{y}_t]$$



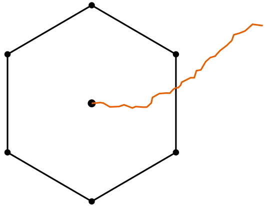
Stochastic localization

Our goal is to simulate, **on average** over unobserved $\boldsymbol{x}_0 \sim \mu$,

$$\boldsymbol{y}_t = t\boldsymbol{x}_0 + \boldsymbol{B}_t$$

Equivalent: Markovian SDE, with drift **current conditional expectation** of $\boldsymbol{x}_0$

$$\mathrm{d}\boldsymbol{y}_t = \boldsymbol{m}_t \mathrm{d}t + \mathrm{d}\boldsymbol{W}_t \qquad \boldsymbol{m}_t = \mathbb{E}[\boldsymbol{x}_0 | \boldsymbol{y}_t]$$



Upshot: estimating $\boldsymbol{m}_t$ lets us simulate $\boldsymbol{y}_t$, thus sample from μ

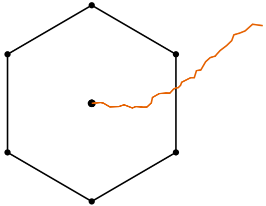
Stochastic localization

Our goal is to simulate, **on average** over unobserved $\boldsymbol{x}_0 \sim \mu$,

$$\boldsymbol{y}_t = t\boldsymbol{x}_0 + \boldsymbol{B}_t$$

Equivalent: Markovian SDE, with drift **current conditional expectation** of $\boldsymbol{x}_0$

$$\mathrm{d}\boldsymbol{y}_t = \boldsymbol{m}_t \mathrm{d}t + \mathrm{d}\boldsymbol{W}_t \qquad \boldsymbol{m}_t = \mathbb{E}[\boldsymbol{x}_0 | \boldsymbol{y}_t]$$



Upshot: estimating $\boldsymbol{m}_t$ lets us simulate $\boldsymbol{y}_t$, thus sample from μ

AMS22: estimate $\boldsymbol{m}_t$ by approximate message passing (AMP) iteration

Results

Recall $H_N(\boldsymbol{\sigma}) = \sum_p \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle$. **Mixture function**: $\xi(q) = \sum_{p \ge 2} \gamma_p^2 q^p$

Results

Recall $H_N(\boldsymbol{\sigma}) = \sum_p \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle$. **Mixture function**: $\xi(q) = \sum_{p \geq 2} \gamma_p^2 q^p$

Theorem (H.–Montanari–Pham 24)

Consider **spherical** model $\mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{\beta H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma}$.

Results

Recall $H_N(\boldsymbol{\sigma}) = \sum_p \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle$. **Mixture function**: $\xi(q) = \sum_{p \ge 2} \gamma_p^2 q^p$

Theorem (H.–Montanari–Pham 24)

Consider **spherical** model $\mu(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma}$.

Results

Recall $H_N(\boldsymbol{\sigma}) = \sum_p \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle$. **Mixture function**: $\xi(q) = \sum_{p \ge 2} \gamma_p^2 q^p$

Theorem (H.–Montanari–Pham 24)

Consider **spherical** model $\mu(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma}$.

1. If

$$\xi''(q) < (1-q)^{-2} \qquad \forall q \in [0,1)$$

a poly-time algorithm samples from μ with **TV** error $o_N(1)$.

Results

Recall $H_N(\boldsymbol{\sigma}) = \sum_p \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle$. **Mixture function**: $\xi(q) = \sum_{p \geq 2} \gamma_p^2 q^p$

Theorem (H.–Montanari–Pham 24)

Consider **spherical** model $\mu(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma}$.

1. If
$$\xi''(q) < (1-q)^{-2} \qquad \forall q \in [0,1)$$
a poly-time algorithm samples from μ with **TV** error $o_N(1)$.
2. Otherwise*, a generalized class of SL algorithms fails. (*in RS phase)

Results

Recall $H_N(\boldsymbol{\sigma}) = \sum_p \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle$. **Mixture function**: $\xi(q) = \sum_{p \geq 2} \gamma_p^2 q^p$

Theorem (H.–Montanari–Pham 24)

Consider **spherical** model $\mu(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma}$.

1. If
$$\xi''(q) < (1-q)^{-2} \qquad \forall q \in [0,1)$$
a poly-time algorithm samples from μ with **TV** error $o_N(1)$.
2. Otherwise*, a generalized class of SL algorithms fails. (*in RS phase)

For pure p-spin model $\xi(q) = \beta^2 q^p$,

$$\frac{\beta_{\mathsf{SL}}}{\beta_{\mathsf{sh}}} \in \left[\frac{\sqrt{e}}{2} \approx 0.824, 1 \right]$$

Improving Wasserstein to total variation

1. Simulate SL process $\boldsymbol{y}_t$ with **improved mean estimator**

Improving Wasserstein to total variation

1. Simulate SL process $\boldsymbol{y}_t$ with **improved mean estimator**

AMS22: estimator $\widetilde{\boldsymbol{m}}_t$ with $\mathbb{E}\|\widetilde{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2 = o(N) \Rightarrow$ Wasserstein sampling

$$\mathop{\mathbb{E}}_{(\widetilde{\boldsymbol{\sigma}},\boldsymbol{\sigma})\sim\text{Coupling}(\mu^{\mathsf{alg}},\mu)} \|\widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|^2 = o(N)$$

Improving Wasserstein to total variation

1. Simulate SL process $\boldsymbol{y}_t$ with **improved mean estimator**

AMS22: estimator $\widetilde{\boldsymbol{m}}_t$ with $\mathbb{E}\|\widetilde{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2 = o(N) \Rightarrow$ Wasserstein sampling

$$\mathop{\mathbb{E}}_{(\widetilde{\boldsymbol{\sigma}},\boldsymbol{\sigma})\sim \mathrm{Coupling}(\mu^{\mathsf{alg}},\mu)} \|\widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|^2 = o(N)$$

This work: improved $\widehat{\boldsymbol{m}}_t$ with $\mathbb{E}\|\widehat{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2 = o(1) \Rightarrow$ can show

$$\mathrm{KL}(\widehat{\boldsymbol{y}}_T, \boldsymbol{y}_T) \le \int_0^T \mathbb{E}\|\widehat{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2 \, \mathrm{d}t = o(1)$$

(and $\mathrm{TV} \le \sqrt{\mathrm{KL}} = o(1)$)

Improving Wasserstein to total variation

1. Simulate SL process $\boldsymbol{y}_t$ with **improved mean estimator**

AMS22: estimator $\widetilde{\boldsymbol{m}}_t$ with $\mathbb{E}\|\widetilde{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2 = o(N) \Rightarrow$ Wasserstein sampling

$$\mathbb{E}_{(\widetilde{\boldsymbol{\sigma}},\boldsymbol{\sigma})\sim \text{Coupling}(\mu^{\text{alg}},\mu)} \|\widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|^2 = o(N)$$

This work: improved $\widehat{\boldsymbol{m}}_t$ with $\mathbb{E}\|\widehat{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2 = o(1) \Rightarrow$ can show

$$\mathrm{KL}(\widehat{\boldsymbol{y}}_T, \boldsymbol{y}_T) \le \int_0^T \mathbb{E}\|\widehat{\boldsymbol{m}}_t - \boldsymbol{m}_t\|^2 \, \mathrm{d}t = o(1)$$

(and $\mathrm{TV} \le \sqrt{\mathrm{KL}} = o(1)$)

2. Sample from $\mu_T = \mathcal{L}(\boldsymbol{x}_0|\boldsymbol{y}_T)$, which is **log-concave** for large $T = O(1)$

Improved mean estimation

$\mu_t = \mathcal{L}(\boldsymbol{x}_0 | \boldsymbol{y}_t)$ is Gibbs measure of **tilted** model: $\mu_t(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_{N,t}(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$ for

$$H_{N,t}(\boldsymbol{\sigma}) = H_N(\boldsymbol{\sigma}) + (\boldsymbol{y}_t, \boldsymbol{\sigma})$$

Improved mean estimation

$\mu_t = \mathcal{L}(\boldsymbol{x}_0 | \boldsymbol{y}_t)$ is Gibbs measure of **tilted** model: $\mu_t(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_{N,t}(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$ for

$$H_{N,t}(\boldsymbol{\sigma}) = H_N(\boldsymbol{\sigma}) + (\boldsymbol{y}_t, \boldsymbol{\sigma})$$

Estimator of AMS22: solution $\widetilde{\boldsymbol{m}}_t = \boldsymbol{m}^{\mathrm{TAP}}$ to **TAP equation**

$$\boldsymbol{m} = a(\|\boldsymbol{m}\|)\nabla H_{N,t}(\boldsymbol{m}) - b(\|\boldsymbol{m}\|)\boldsymbol{m}$$

Improved mean estimation

$\mu_t = \mathcal{L}(\boldsymbol{x}_0 | \boldsymbol{y}_t)$ is Gibbs measure of **tilted** model: $\mu_t(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_{N,t}(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$ for

$$H_{N,t}(\boldsymbol{\sigma}) = H_N(\boldsymbol{\sigma}) + (\boldsymbol{y}_t, \boldsymbol{\sigma})$$

Estimator of AMS22: solution $\widetilde{\boldsymbol{m}}_t = \boldsymbol{m}^{\mathrm{TAP}}$ to **TAP equation**

$$\boldsymbol{m} = a(\|\boldsymbol{m}\|) \nabla H_{N,t}(\boldsymbol{m}) - b(\|\boldsymbol{m}\|) \boldsymbol{m}$$

found via **AMP iteration**

$$\boldsymbol{m}^{k+1} = a_k \nabla H_{N,t}(\boldsymbol{m}^k) - b_k \boldsymbol{m}^{k-1}$$

Improved mean estimation

$\mu_t = \mathcal{L}(\boldsymbol{x}_0 | \boldsymbol{y}_t)$ is Gibbs measure of **tilted** model: $\mu_t(\mathrm{d}\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_{N,t}(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$ for

$$H_{N,t}(\boldsymbol{\sigma}) = H_N(\boldsymbol{\sigma}) + (\boldsymbol{y}_t, \boldsymbol{\sigma})$$

Estimator of AMS22: solution $\widetilde{\boldsymbol{m}}_t = \boldsymbol{m}^{\mathrm{TAP}}$ to **TAP equation**

$$\boldsymbol{m} = a(\|\boldsymbol{m}\|) \nabla H_{N,t}(\boldsymbol{m}) - b(\|\boldsymbol{m}\|) \boldsymbol{m}$$

found via **AMP iteration**

$$\boldsymbol{m}^{k+1} = a_k \nabla H_{N,t}(\boldsymbol{m}^k) - b_k \boldsymbol{m}^{k-1}$$

Analyze by AMP state evolution $\Rightarrow \mathbb{E}\|\boldsymbol{m}^{\mathrm{TAP}} - \mathrm{mean}(\mu_t)\|^2 = o(N)$

Improved mean estimation

$\mu_t = \mathcal{L}(\boldsymbol{x}_0 | \boldsymbol{y}_t)$ is Gibbs measure of **tilted** model: $\mu_t(d\boldsymbol{\sigma}) \equiv \frac{1}{Z} e^{H_{N,t}(\boldsymbol{\sigma})} d\boldsymbol{\sigma}$ for

$$H_{N,t}(\boldsymbol{\sigma}) = H_N(\boldsymbol{\sigma}) + (\boldsymbol{y}_t, \boldsymbol{\sigma})$$

Estimator of AMS22: solution $\widetilde{\boldsymbol{m}}_t = \boldsymbol{m}^{\mathrm{TAP}}$ to **TAP equation**

$$\boldsymbol{m} = a(\|\boldsymbol{m}\|) \nabla H_{N,t}(\boldsymbol{m}) - b(\|\boldsymbol{m}\|) \boldsymbol{m}$$

found via **AMP iteration**

$$\boldsymbol{m}^{k+1} = a_k \nabla H_{N,t}(\boldsymbol{m}^k) - b_k \boldsymbol{m}^{k-1}$$

Analyze by AMP state evolution $\Rightarrow \mathbb{E}\|\boldsymbol{m}^{\mathrm{TAP}} - \mathrm{mean}(\mu_t)\|^2 = o(N)$

Our analysis:

- Actually, $\mathbb{E}\|\boldsymbol{m}^{\mathrm{TAP}} - \mathrm{mean}(\mu_t)\|^2 = O(1)$
- Compute correction $\boldsymbol{\Delta}$, so $\widehat{\boldsymbol{m}}_t = \boldsymbol{m}^{\mathrm{TAP}} + \boldsymbol{\Delta}$ has error $o(1)$

Conclusion

- In spherical models, algorithmic SL samples in total variation for $\beta < \beta_{\mathsf{SL}}$
- Generalized class of SL algorithms fails for $\beta \geq \beta_{\mathsf{SL}}$

Conclusion

- In spherical models, algorithmic SL samples in total variation for $\beta < \beta_{\mathsf{SL}}$
- Generalized class of SL algorithms fails for $\beta \geq \beta_{\mathsf{SL}}$

- Main step: more accurate mean estimator for tilted measures

Conclusion

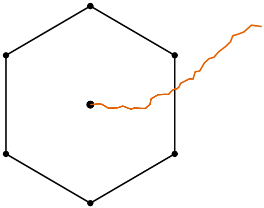
- In spherical models, algorithmic SL samples in total variation for $\beta < \beta_{\mathsf{SL}}$
- Generalized class of SL algorithms fails for $\beta \geq \beta_{\mathsf{SL}}$

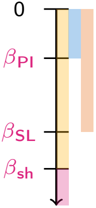
- Main step: more accurate mean estimator for tilted measures
- **Q**: sampling between β_{SL} and β_{sh}?

Conclusion

- In spherical models, algorithmic SL samples in total variation for $\beta < \beta_{\mathsf{SL}}$
- Generalized class of SL algorithms fails for $\beta \geq \beta_{\mathsf{SL}}$

- Main step: more accurate mean estimator for tilted measures
- **Q**: sampling between β_{SL} and β_{sh}?





Glauber/Langevin dynamics mix rapidly

Algorithmic SL samples in **total variation**

Prediction: Glauber/Langevin succeeds

Prediction: shattering; sampling hard

Thanks!