# Diffusion-Based Sampling for Spin Glasses

Mark Sellke (Harvard Statistics)

Harvard Theory of Computing Seminar, Nov 15, 2023

arXiv:2203.05093 (FOCS 2022); 2307.04659; 2310.08912

# Collaborators

Ahmed El Alaoui (Cornell)

Andrea Montanari (Stanford)

# Plan for this Talk

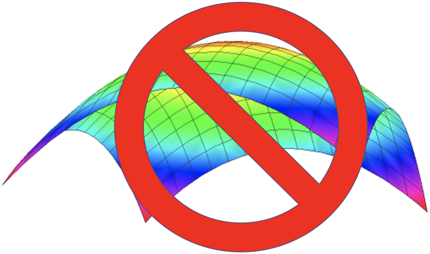
1. Background on High-Dimensional Sampling
    - Sequential sampling
    - Stochastic localization
    - Connection to diffusion models
    - Goal for today: the Sherrington–Kirkpatrick model
2. Main Results
    - Algorithm: approximate message passing and more
    - Stability of the algorithm; hardness from chaos.
    - $p$-spin generalizations; another source of chaos.

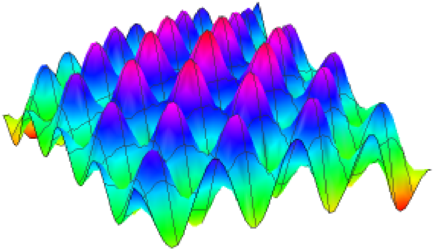
# Sampling

Goal: generate

$$x^* \sim \mu(dx) \quad \text{given} \quad \mu \in \mathcal{P}(\mathbb{R}^n).$$

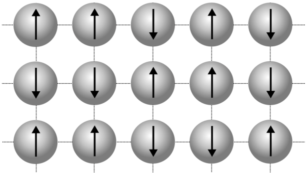
For $\mu$ high-dimensional and NOT log-concave.

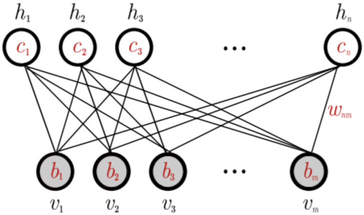




In this talk, focus on Ising models:

$$\mu_{A,\beta}(x) = \frac{1}{Z(\beta)} e^{\beta\langle x, Ax\rangle/2}, \qquad x \in \{-1,+1\}^n.$$

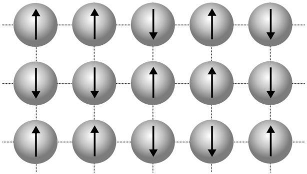


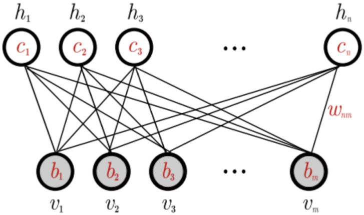


# Sampling

In this talk, focus on Ising models:

$$\mu_{A,\beta}(x) = \frac{1}{Z(\beta)} e^{\beta\langle x, Ax\rangle/2}, \qquad x \in \{-1,+1\}^n.$$






Glauber dynamics

- Repeatedly choose $i \in [n]$ and resample $x_i$ given other coordinates.
- Mixes rapidly if $\beta A$ is small. In general, mixing can be very slow.

# Sequential Sampling

Given a distribution $\mu \in \mathcal{P}(\{-1,+1\}^n)$, suppose we have a conditional expectation **oracle** to evaluate

$$\mathrm{m}^t = \mathbb{E}^{\mathrm{x}\sim\mu}[\mathrm{x} \mid (x_1 = x_1^*, \dots, x_t = x_t^*)], \quad t \in \{0, 1, \dots, n-1\}.$$

Given a distribution $\mu \in \mathcal{P}(\{-1,+1\}^n)$, suppose we have a conditional expectation **oracle** to evaluate

$$\mathrm{m}^t = \mathbb{E}^{\mathrm{x}\sim\mu}[\mathrm{x} \mid (x_1 = x_1^*, \ldots, x_t = x_t^*)], \quad t \in \{0, 1, \ldots, n-1\}.$$

Then we can directly sample x, one coordinate at a time. Namely,

$$\mathbb{P}^t[x_{t+1} = 1 \mid x_1, \ldots, x_t] = \frac{\mathrm{m}_{t+1}^t + 1}{2}.$$

This is the foundation for equivalence between counting and sampling.

# Sequential Sampling

Directly implementing sequential sampling may be too much to hope for.

- Requires a strong oracle, especially for continuous variables.
- Maybe estimating $\mathrm{m}^t$ is no easier than sampling.
- Unclear how to choose a good order for the coordinates.

# Sequential Sampling

Directly implementing sequential sampling may be too much to hope for.

- Requires a strong oracle, especially for continuous variables.
- Maybe estimating $\mathrm{m}^t$ is no easier than sampling.
- Unclear how to choose a good order for the coordinates.

The high-level idea is to reveal $\mathrm{x}^*$ **gradually**. This is fundamentally different from a Markov chain!

And, information can be gradually revealed in other ways.

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

$$y_t = t x^* + B_t \quad \sim \mathcal{N}(t x^*, t I_n).$$

$x^* \sim \mu$ is independent of Brownian motion $B_t$ and **only $y_t$ is observed**.

# Sampling via Stochastic Localization

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

$$y_t = t x^* + B_t \quad \sim \mathcal{N}(t x^*, t I_n).$$

$x^* \sim \mu$ is independent of Brownian motion $B_t$ and **only $y_t$ is observed**.

Our sampling algorithm takes the following form:

1. **Simulate** $y_t$ for a long time $t \in [0, T]$ without knowing $x^*$.
2. **Read off**

$$x^* \approx \frac{y_T}{T}.$$

Equivalently, increments $y_{(k+1)\delta} - y_{k\delta}$ are IID noisy observations of $x^*$.

# Sampling via Stochastic Localization

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

$$y_t = t x^* + B_t \quad \sim \mathcal{N}(t x^*, t I_n).$$

$x^* \sim \mu$ is independent of Brownian motion $B_t$ and **only $y_t$ is observed**.

Our sampling algorithm takes the following form:

1. **Simulate** $y_t$ for a long time $t \in [0, T]$ without knowing $x^*$.
2. **Read off**

$$x^* \approx \frac{y_T}{T}.$$

Equivalently, increments $y_{(k+1)\delta} - y_{k\delta}$ are IID noisy observations of $x^*$.

This process has been popularized in high-dimensional convex geometry [Eldan 13, Lee-Vempala 17, Chen 21, Klartag-Lehec 22, Jambulapati-Lee-Vempala 22].

# Simulating $y_t$

Our goal is to sample a random path

$$y_t = tx^* + B_t \quad \sim \mathcal{N}(tx^*, tI_n)$$

from its distribution **averaged** over the unknown $x^*$.

# Simulating $\mathsf{y}_t$

Our goal is to sample a random path

$$\mathsf{y}_t = t\mathsf{x}^* + \mathsf{B}_t \quad \sim \mathcal{N}(t\mathsf{x}^*, t\mathsf{I}_n)$$

from its distribution **averaged** over the unknown $\mathsf{x}^*$.

This law on paths is actually Markovian. The instantaneous drift is the **current conditional expectation** of the unknown drift:

$$\begin{aligned} d\mathsf{y}_t &= \mathsf{m}_t dt + d\mathsf{W}_t; \\ \mathsf{m}_t &= \mathbb{E}[\mathsf{x}^* \mid \mathcal{F}_t] = \mathbb{E}[\mathsf{x}^* \mid \mathsf{y}_t] \end{aligned}$$

for $W_t$ a (separate) Brownian motion.

Markov property: $y_t$ is a <u>sufficient statistic</u> for $y_{[0,t]}$.

# Parallels with Pólya's Urn

Pólya's urn gives an indirect way to sample $p \sim Unif([0,1])$. Stochastic localization sampling is a continuous-time parallel.

| Goal | Pólya's Urn | Stoch. Loc. |
|---|---|---|
| Want to sample | $p \sim Unif([0,1])$ | $x^* \sim \mu \in \mathcal{P}(\mathbb{R}^n)$ |
| Observation process | $a_1, a_2, \dots \overset{IID}{\sim} Ber(p)$ | $y_t = tx^* + B_t$ |
| Process w/o sample | $a_t \sim Ber(\mathbb{E}^t[p])$ | $dy_t = \mathbb{E}^t[x^*] + dB_t$ |
| Process $\to$ sample | $p \approx (a_1 + \cdots + a_T)/T$ | $x^* \approx y_T/T$ |

In this case, $\mathbb{E}^t[p] = \frac{N_t(1)+1}{N_t(0)+N_t(1)+2}$ by Laplace's rule of succession.

# Parallels with Pólya's Urn

Pólya's urn gives an indirect way to sample $p \sim Unif([0,1])$. Stochastic localization sampling is a continuous-time parallel.

| Goal | Pólya's Urn | Stoch. Loc. |
|---|---|---|
| Want to sample | $p \sim Unif([0,1])$ | $x^* \sim \mu \in \mathcal{P}(\mathbb{R}^n)$ |
| Observation process | $a_1, a_2, \dots \overset{IID}{\sim} Ber(p)$ | $\mathrm{y}_t = tx^* + \mathrm{B}_t$ |
| Process w/o sample | $a_t \sim Ber(\mathbb{E}^t[p])$ | $\mathrm{dy}_t = \mathbb{E}^t[x^*] + \mathrm{d}B_t$ |
| Process $\to$ sample | $p \approx (a_1 + \cdots + a_T)/T$ | $x^* \approx \mathrm{y}_T/T$ |

In this case, $\mathbb{E}^t[p] = \frac{N_t(1)+1}{N_t(0)+N_t(1)+2}$ by Laplace's rule of succession.

$S_N = a_1 + \cdots + a_T$ plays the role of $\mathrm{y}_T$.

- Given $p$: $S_T$ is a discrete walk with drift $p$.
- Given $x^*$: $\mathrm{y}_t$ is a continuous walk with drift $x^*$.
- Increments $\mathrm{y}_{(j+1)\delta} - \mathrm{y}_{j\delta}$ play the role of $a_j$.

# The Resulting Algorithm

$$dy_t = m_t dt + dW_t,$$

A continuous-time stochastic process is not really an algorithm.

Of course, we should discretize time.

# The Resulting Algorithm

$$\mathrm{d}\mathrm{y}_t = \mathrm{m}_t \mathrm{d}t + \mathrm{d}\mathrm{W}_t,$$

---

**Input:** Data: Probability measure $\mu$
**Input:** Result: Sample $\mathrm{x}^* \sim \mu$
**for** $t \in [0, \delta, \dots, T - \delta]$ **do**
  Sample $\mathrm{g}_t \sim \mathcal{N}(0, I_n)$
  Set $\widehat{\mathrm{y}}_{t+\delta} = \widehat{\mathrm{y}}_t + \delta \widehat{\mathrm{m}}_t(\mathrm{y}_t) + \sqrt{\delta}\mathrm{g}_t$
**end**
Set $\widehat{\mathrm{x}}^* = Round(\widehat{\mathrm{y}}_T / T) \in \{-1, +1\}^n$
**return** $\widehat{\mathrm{x}}^*$

---

# The Resulting Algorithm

$$dy_t = m_t dt + dW_t,$$

**Input:** Data: Probability measure $\mu$
**Input:** Result: Sample $x^* \sim \mu$
**for** $t \in [0, \delta, \ldots, T - \delta]$ **do**
  Sample $g_t \sim \mathcal{N}(0, I_n)$
  Set $\widehat{y}_{t+\delta} = \widehat{y}_t + \delta \widehat{m}_t(y_t) + \sqrt{\delta} g_t$
**end**
Set $\widehat{x}^* = Round(\widehat{y}_T / T) \in \{-1, +1\}^n$
**return** $\widehat{x}^*$

Main requirement: a good approximation

$$\widehat{m}_t(\widehat{y}_t) \approx m_t(\widehat{y}_t) \equiv \mathbb{E}[x^* \mid \widehat{y}_t].$$

# Where Do We Stand?

So far:

- General sampling procedure.
- Requires repeatedly estimating $\mathsf{m}_t(\widehat{\mathsf{y}}_t) \approx \mathbb{E}[x^* \mid \widehat{\mathsf{y}}_t]$.

We have replaced the need for one oracle with another...is it any better?

# Where Do We Stand?

So far:

- General sampling procedure.
- Requires repeatedly estimating $\mathrm{m}_t(\widehat{\mathrm{y}}_t) \approx \mathbb{E}[x^* \mid \widehat{\mathrm{y}}_t]$.

We have replaced the need for one oracle with another...is it any better?

Main result for this talk: example where the answer is **yes**.

- SK model: coupling matrix A is GOE.
- Computing $\mathrm{m}_t(\mathrm{y}_t)$ falls into the wheelhouse of high-dimensional statistics/optimization.
- But, provable hardness for "stable" sampling at large $\beta$.

# Connection to Diffusion Models

# Connection to Diffusion Models

Modern diffusion-based sampling has two main processes:

- Forward: turn **sample** $X_0 \sim \mu$ into **noise** via OU flow

$$\mathrm{d}X_T = -X_T \, \mathrm{d}T + \sqrt{2}\mathrm{d}W_T.$$

- Backward: time-reverse the forward process, i.e. **noise** $\to$ **sample**.
- The backward process gains a $\mu$-dependent drift.

# Connection to Diffusion Models

Modern diffusion-based sampling has two main processes:

- Forward: turn **sample** $X_0 \sim \mu$ into **noise** via OU flow

$$dX_T = -X_T \, dT + \sqrt{2} dW_T.$$

- Backward: time-reverse the forward process, i.e. **noise** $\rightarrow$ **sample**.
- The backward process gains a $\mu$-dependent drift.

Given $X_S$ for $S < T$: $\quad e^T X_T \overset{d}{=} e^S X_S + \sqrt{e^{2T} - e^{2S}} \mathcal{N}(0, I_n).$

Given $y_s$ for $s > t$: $\quad y_t/t \overset{d}{=} y_s/s + \sqrt{\frac{s-t}{t}} \mathcal{N}(0, I_n).$

# Connection to Diffusion Models

Modern diffusion-based sampling has two main processes:

- Forward: turn **sample** $X_0 \sim \mu$ into **noise** via OU flow

$$\mathrm{d}X_T = -X_T \, \mathrm{d}T + \sqrt{2}\mathrm{d}W_T.$$

- Backward: time-reverse the forward process, i.e. **noise** $\to$ **sample**.
- The backward process gains a $\mu$-dependent drift.

Given $X_S$ for $S < T$: $\quad e^T X_T \stackrel{d}{=} e^S X_S + \sqrt{e^{2T} - e^{2S}}\,\mathcal{N}(0, I_n).$

Given $y_s$ for $s > t$: $\quad y_t/t \stackrel{d}{=} y_s/s + \sqrt{\frac{s-t}{t}}\mathcal{N}(0, I_n).$

Stochastic localization is a reparametrization of the backward process!

- Diffusion models learn SDE coefficients from forward process on samples. Provable guarantees from good estimates (ask Sitan!)
- This talk: no samples, but a formula for $\mu$.

# Sherrington-Kirkpatrick Model

Ising model with random couplings:

$$\mu_{G,\beta}(x) = \frac{1}{Z_n(\beta)} e^{\beta\langle x, Gx\rangle/2}.$$

Random symmetric matrix $G \sim GOE(n)$:

- $G = G^\top$. Entries otherwise independent.
- $G_{i,j} \sim \mathcal{N}(0, 1/n)$ for $i < j$.

# Sherrington-Kirkpatrick Model

Ising model with random couplings:

$$\mu_{G,\beta}(x) = \frac{1}{Z_n(\beta)} e^{\beta \langle x, Gx \rangle / 2}.$$

Random symmetric matrix $G \sim GOE(n)$:

- $G = G^\top$. Entries otherwise independent.
- $G_{i,j} \sim \mathcal{N}(0, 1/n)$ for $i < j$.

Goal: given $G \sim GOE(n)$, generate a sample from $\mu_{G,\beta}$.

Dobrushin's condition for fast mixing of Glauber works if $\beta \leq cn^{-1/2}$.
But we would like $\beta$ to be constant size.

# Brief History of the SK Model

[Ising 1925]: Ising model for ferromagnets.

[Sherrington-Kirkpatrick 1975]: model for **disordered** magnets.

[Parisi 1982]: non-rigorous solution via replica symmetry breaking.

# Brief History of the SK Model

[Ising 1925]: Ising model for ferromagnets.

[Sherrington-Kirkpatrick 1975]: model for **disordered** magnets.

[Parisi 1982]: non-rigorous solution via replica symmetry breaking.

[Talagrand 2006] proves the Parisi formula.

- Huge amount of other important work including [Aizenman-Ruelle-Lebowitz 82, Ruelle 87, Guerra 03, Chatterjee 09, Panchenko 14, Ding-Sly-Sun 15, Auffinger-Chen 17,...].

# Connections

SK model is a prototype for disordered, random probability measures.

- Random MaxCut and $K$-SAT.
- Coloring random graphs.
- Posteriors in high-dimensional statistics.

# Connections

SK model is a prototype for disordered, random probability measures.

- Random MaxCut and $K$-SAT.
- Coloring random graphs.
- Posteriors in high-dimensional statistics.

E.g. optimal MaxCut in a random sparse graph ([Dembo-Montanari-Sen 17]).
For $\mathrm{G} \sim \mathrm{G}\left(n, \frac{\lambda}{n}\right)$:

$$\mathrm{MaxCut(G)} = n\left(\frac{\lambda}{4} + C_*\sqrt{\frac{\lambda}{4}} + o(\sqrt{\lambda})\right) + o(n).$$

# Rigorous Results on Sampling

$$\mu_{G,\beta}(x) = \frac{1}{Z_n(\beta)} e^{\beta\langle x, Gx\rangle/2}.$$

Expect: efficient sampling possible for $\beta < 1$, impossible for $\beta > 1$.

- "Replica symmetric" for $\beta < 1$. For independent $x, x' \sim \mu_{G,\beta}$,

$$\mathbb{E}[|\langle x, x'\rangle|/n] \approx 0.$$

- "Replica symmetry breaking" for $\beta > 1$. Here

$$\mathbb{E}[|\langle x, x'\rangle|/n] \geq c(\beta) > 0.$$

# Rigorous Results on Sampling

$$\mu_{G,\beta}(x) = \frac{1}{Z_n(\beta)} e^{\beta \langle x, Gx \rangle / 2}.$$

Expect: efficient sampling possible for $\beta < 1$, impossible for $\beta > 1$.

- "Replica symmetric" for $\beta < 1$. For independent $x, x' \sim \mu_{G,\beta}$,

$$\mathbb{E}[|\langle x, x' \rangle|/n] \approx 0.$$

- "Replica symmetry breaking" for $\beta > 1$. Here

$$\mathbb{E}[|\langle x, x' \rangle|/n] \geq c(\beta) > 0.$$

Recent progress: Glauber mixes in $O(n \log n)$ steps for $\beta < 1/4$.

[Bodineau-Bauerschmidt 20, Eldan-Koehler-Zeitouni 21, Anari-Jain-Koehler-Pham-Vuong 21].

# Rigorous Results on Sampling

$$\mu_{G,\beta}(x) = \frac{1}{Z_n(\beta)} e^{\beta\langle x, Gx\rangle/2}.$$

Expect: efficient sampling possible for $\beta < 1$, impossible for $\beta > 1$.

- "Replica symmetric" for $\beta < 1$. For independent $x, x' \sim \mu_{G,\beta}$,

$$\mathbb{E}[|\langle x, x'\rangle|/n] \approx 0.$$

- "Replica symmetry breaking" for $\beta > 1$. Here

$$\mathbb{E}[|\langle x, x'\rangle|/n] \geq c(\beta) > 0.$$

Recent progress: Glauber mixes in $O(n \log n)$ steps for $\beta < 1/4$.
[Bodineau-Bauerschmidt 20, Eldan-Koehler-Zeitouni 21, Anari-Jain-Koehler-Pham-Vuong 21].

Our result: stochastic localization succeeds (in a weaker sense) for $\beta < 1$. (Originally $\beta < 1/2$, improvement by [Celentano 22].)

# Normalized Wasserstein Metric

Given $\mu_1, \mu_2 \in \mathcal{P}(\{-1,1\}^n)$, define the normalized Wasserstein metric

$$W_{1,n}(\mu_1,\mu_2) = \inf_{(x_1,x_2)\sim Coupling(\mu_1,\mu_2)} \frac{\mathbb{E}[\|x_1 - x_2\|_{\ell^1}]}{n}.$$

$W_{1,n}(\mu_1,\mu_2) \leq o(1)$ means that $x_1, x_2$ differ by $o(n)$ coordinates under an optimal coupling. We will consider such pairs of points to be close.

# Main Result

**Theorem (Alaoui-Montanari-S 22, Celentano 22)**

*For any* $\beta < 1$ *and* $\varepsilon > 0$*, there exists a randomized algorithm with complexity* $O(n^2)$ *which given* $\mathsf{G}$ *outputs* $\mathsf{x} \sim \mu^{\mathrm{alg}}_{\mathsf{G},\beta}$ *such that*

$$\mathbb{E}[W_{1,n}(\mu^{\mathrm{alg}}_{\mathsf{G},\beta}, \mu_{\mathsf{G},\beta})] \leq \varepsilon.$$

# Estimating the Mean

# Estimating the Mean

To sample for $\beta < 1$, our main goal is to estimate $m_t = \mathbb{E}[x^* \mid y_t]$ for

$$y_t = t x^* + B_t.$$

The solution goes through several ideas in high-dimensional statistics and optimization.

# Estimating the Mean

To sample for $\beta < 1$, our main goal is to estimate $\mathrm{m}_t = \mathbb{E}[\mathrm{x}^* \mid \mathrm{y}_t]$ for

$$\mathrm{y}_t = t\mathrm{x}^* + \mathrm{B}_t.$$

The solution goes through several ideas in high-dimensional statistics and optimization.

Two phase procedure:

- Rough estimate for $\mathrm{m}_t$ using approximate message passing.
- High-accuracy estimate for $\mathrm{m}_t$ using gradient descent on a well-chosen potential.

For now, assume perfect simulation until time $t$. Observe

$$\mathrm{y}_t \sim \mathcal{N}(tx^*, tI_n),$$

estimate $m_t(\mathrm{y}_t)$.

# Step 1: Rough Estimate of $m_t$

Self-consistent "naive mean-field" equation for $m_t = \mathbb{E}[x \mid y_t]$:

$$m_t \approx \tanh\left(\beta G m_t + y_t\right)$$

- Intuitively, $\left(\beta G m_t + y_t\right)_i$ is the effective field on $x_i$.
- $\tanh(\cdot)$ converts from field on $\{-1, +1\}$ to probabilities

# Step 1: Rough Estimate of $m_t$

Self-consistent "naive mean-field" equation for $m_t = \mathbb{E}[x \mid y_t]$:

$$m_t \approx \tanh\left(\beta G m_t + y_t\right)$$

- Intuitively, $\left(\beta G m_t + y_t\right)_i$ is the effective field on $x_i$.
- $\tanh(\cdot)$ converts from field on $\{-1, +1\}$ to probabilities
- Not quite right. It actually should be

$$m_t = \mathbb{E}^t[\tanh\left(\beta G x + y_t\right)].$$

  $\tanh(\cdot)$ is non-linear and although $\mathbb{E}^t[Gx] = G m_t$ there is nontrivial conditional randomness left.

# Step 1: Rough Estimate of $m_t$

Self-consistent "naive mean-field" equation for $m_t = \mathbb{E}[x \mid y_t]$:

$$m_t \approx \tanh\left(\beta G m_t + y_t\right)$$

- Intuitively, $\left(\beta G m_t + y_t\right)_i$ is the effective field on $x_i$.
- $\tanh(\cdot)$ converts from field on $\{-1, +1\}$ to probabilities
- Not quite right. It actually should be

$$m_t = \mathbb{E}^t[\tanh\left(\beta G x + y_t\right)].$$

  $\tanh(\cdot)$ is non-linear and although $\mathbb{E}^t[Gx] = Gm_t$ there is nontrivial conditional randomness left.

Revised Thouless-Anderson-Palmer (TAP) equation:

$$m_t \approx \tanh\left(\beta G m_t + y_t - \beta^2\left(1 - \frac{\|m_t\|_2^2}{n}\right) m_t\right).$$

# Step 1: Rough Estimate of $\mathrm{m}_t$

Turn the TAP equation into a **recursion** and repeat until convergence to an approximate **fixed point**:

$$\widehat{\mathrm{m}}_t^{(k+1)} = \tanh\left(\beta \mathrm{G}\widehat{\mathrm{m}}_t^{(k)} + \mathrm{y}_t - b_k \widehat{\mathrm{m}}_t^{(k-1)}\right),$$

$$b_k = \beta^2 \left(1 - \frac{\|\mathrm{m}_t^{(k)}\|_2^2}{n}\right).$$

# Step 1: Rough Estimate of $\mathrm{m}_t$

Turn the TAP equation into a **recursion** and repeat until convergence to an approximate **fixed point**:

$$\widehat{\mathrm{m}}_t^{(k+1)} = \tanh\left(\beta \mathrm{G}\widehat{\mathrm{m}}_t^{(k)} + \mathrm{y}_t - b_k \widehat{\mathrm{m}}_t^{(k-1)}\right),$$

$$b_k = \beta^2 \left(1 - \frac{\|\mathrm{m}_t^{(k)}\|_2^2}{n}\right).$$

This is an **approximate message passing** algorithm. Generalizes belief propagation to dense matrices G.

- Onsager term $b_k \widehat{\mathrm{m}}_t^{(k-1)}$ cancels "backtracking" paths.

# Step 1: Rough Estimate of $\mathrm{m}_t$

Turn the TAP equation into a **recursion** and repeat until convergence to an approximate **fixed point**:

$$\widehat{\mathrm{m}}_t^{(k+1)} = \tanh\left(\beta \mathrm{G} \widehat{\mathrm{m}}_t^{(k)} + \mathrm{y}_t - b_k \widehat{\mathrm{m}}_t^{(k-1)}\right),$$

$$b_k = \beta^2 \left(1 - \frac{\|\mathrm{m}_t^{(k)}\|_2^2}{n}\right).$$

This is an **approximate message passing** algorithm. Generalizes belief propagation to dense matrices G.

- Onsager term $b_k \widehat{\mathrm{m}}_t^{(k-1)}$ cancels "backtracking" paths.
- By now, a major tool in high-dimensional statistics.

  [Bolthausen 14, Donoho-Maleki-Montanari 09, Bayati-Montanari 11, Javanmard-Montanari 12, Rush-Venkataramanan 18, Chen-Lam 20, Fan 20, Dudeja-Lu-Sen 22]

For large $n$ and $k = O(1)$ iterations, state evolution describes AMP.

# State Evolution for AMP

$$\widehat{\mathrm{m}}_t^{(k+1)} = \tanh\left(\beta G \widehat{\mathrm{m}}_t^{(k)} + \mathrm{y}_t - b_k \widehat{\mathrm{m}}_t^{(k-1)}\right)$$

Idea of AMP: for deterministic $\mathrm{v}, \mathrm{w}$, the vectors

$$G\mathrm{v}, G\mathrm{w}$$

each have i.i.d. Gaussian coordinates. Covariance between $(G\mathrm{v})_i$ and $(G\mathrm{w})_i$ equals $\langle \mathrm{v}, \mathrm{w} \rangle$.

# State Evolution for AMP

$$\widehat{\mathrm{m}}_t^{(k+1)} = \tanh\left(\beta \mathrm{G}\widehat{\mathrm{m}}_t^{(k)} + \mathrm{y}_t - b_k \widehat{\mathrm{m}}_t^{(k-1)}\right)$$

Idea of AMP: for deterministic $\mathrm{v}, \mathrm{w}$, the vectors

$$\mathrm{Gv}, \mathrm{Gw}$$

each have i.i.d. Gaussian coordinates. Covariance between $(\mathrm{Gv})_i$ and $(\mathrm{Gw})_i$ equals $\langle \mathrm{v}, \mathrm{w}\rangle$.

- Onsager term lets us apply this recursively to each $\widehat{\mathrm{m}}_t^{(k+1)}$, despite re-using the same G many times.

State evolution: from **simple initialization** $(\widehat{m}_t^0, \mathrm{y}_t)$, choose uniform $i \in [n]$. Tells us the $n \to \infty$ limiting distribution of

$$\left((\widehat{m}_t^0)_i, (\widehat{m}_t^1)_i, \ldots, (\widehat{m}_t^k)_i\right) \in \mathbb{R}^{k+1}.$$

# State Evolution for AMP

$$\widehat{\mathrm{m}}_t^{(k+1)} = \tanh\left(\beta \mathrm{G}\widehat{\mathrm{m}}_t^{(k)} + \mathrm{y}_t - b_k \widehat{\mathrm{m}}_t^{(k-1)}\right)$$

Idea of AMP: for deterministic $\mathrm{v}, \mathrm{w}$, the vectors

$$\mathrm{Gv}, \mathrm{Gw}$$

each have i.i.d. Gaussian coordinates. Covariance between $(\mathrm{Gv})_i$ and $(\mathrm{Gw})_i$ equals $\langle \mathrm{v}, \mathrm{w}\rangle$.

- Onsager term lets us apply this recursively to each $\widehat{\mathrm{m}}_t^{(k+1)}$, despite re-using the same G many times.

State evolution: from **simple initialization** $(\widehat{m}_t^0, \mathrm{y}_t)$, choose uniform $i \in [n]$. Tells us the $n \to \infty$ limiting distribution of

$$\left((\widehat{m}_t^0)_i, (\widehat{m}_t^1)_i, \dots, (\widehat{m}_t^k)_i\right) \in \mathbb{R}^{k+1}.$$

**Problem**: $x^* \sim \mu_{G,\beta}$ is NOT SIMPLE. So neither is $y_t \sim \mathcal{N}(tx^*, tI_n)$.

# Contiguity with a Simpler Spiked Model

To analyze the AMP recursion, we switch to a **spiked** joint distribution $\mathbb{Q}$ over $(\mathrm{G}, \mathrm{x}^*, \mathrm{y}_t)$. Under $\mathbb{Q}$:

$$\mathrm{x}^* \sim \mathit{Unif}\left(\{-1, 1\}^n\right), \qquad \mathrm{y}_t = t\mathrm{x}^* + B_t,$$
$$\mathrm{G} \sim \mathit{GOE}(n) + \frac{\beta \mathrm{x}^* (\mathrm{x}^*)^\top}{n}.$$

# Contiguity with a Simpler Spiked Model

To analyze the AMP recursion, we switch to a **spiked** joint distribution $\mathbb{Q}$ over $(\mathsf{G}, \mathsf{x}^*, \mathsf{y}_t)$. Under $\mathbb{Q}$:

$$\mathsf{x}^* \sim \mathit{Unif}\left(\{-1,1\}^n\right), \qquad \mathsf{y}_t = t\mathsf{x}^* + B_t,$$

$$\mathsf{G} \sim \mathit{GOE}(n) + \frac{\beta \mathsf{x}^*(\mathsf{x}^*)^\top}{n}.$$

The conditional law $\mathbb{Q}[\mathsf{G} \mid \mathsf{x}^*]$ looks similar to $\mathbb{P}[\mathsf{x}^* \mid \mathsf{G}]$ in SK:

$$\mathbb{Q}[\mathsf{G} \mid \mathsf{x}^*] \propto e^{\beta\langle \mathsf{x}^*, \mathsf{G}\mathsf{x}^*\rangle/2}\, \nu_{GOE(n)}(G).$$

# Contiguity with a Simpler Spiked Model

To analyze the AMP recursion, we switch to a **spiked** joint distribution $\mathbb{Q}$ over $(\mathrm{G}, \mathrm{x}^*, \mathrm{y}_t)$. Under $\mathbb{Q}$:

$$\mathrm{x}^* \sim \mathit{Unif}\left(\{-1, 1\}^n\right), \qquad \mathrm{y}_t = t\mathrm{x}^* + B_t,$$

$$\mathrm{G} \sim \mathit{GOE}(n) + \frac{\beta \mathrm{x}^*(\mathrm{x}^*)^\top}{n}.$$

The conditional law $\mathbb{Q}[\mathrm{G} \mid \mathrm{x}^*]$ looks similar to $\mathbb{P}[\mathrm{x}^* \mid \mathrm{G}]$ in SK:

$$\mathbb{Q}[\mathrm{G} \mid \mathrm{x}^*] \propto e^{\beta \langle \mathrm{x}^*, \mathrm{G}\mathrm{x}^* \rangle / 2} \, \nu_{\mathit{GOE}(n)}(G).$$

Swapping the order distorts probabilities by a partition function factor

$$Z_{SK}(\mathrm{G}) = \sum_{\mathrm{v} \in \{-1, +1\}^n} e^{\beta \langle \mathrm{v}, \mathrm{G}\mathrm{v} \rangle / 2}.$$

- $Z_{SK}(\mathrm{G})$ fluctuates **mildly** for $\beta < 1$ [Aizenman-Ruelle-Lebowitz 82]. Yields **contiguity**; estimating $m_t$ w.h.p. is **equivalent**.

# State Evolution for AMP

State evolution: $i$-th coordinate of $\widehat{\mathrm{m}}_t^{(k)}$ behaves like

$$\tanh\left(a_t^{(k)} x_i + b_t^{(k)} Z\right), \quad Z \sim \mathcal{N}(0,1).$$

Limits $(a_t^{\infty}, b_t^{\infty})$ yield the asymptotic mean-squared error (MSE)

$$E_*(t) = \lim_{k\to\infty} \operatorname*{p-lim}_{n\to\infty} \mathbb{E}\|\widehat{\mathrm{m}}_t^{(k)} - \mathrm{x}\|_2^2.$$

# State Evolution for AMP

State evolution: $i$-th coordinate of $\widehat{\mathrm{m}}_t^{(k)}$ behaves like

$$\tanh\left(a_t^{(k)} x_i + b_t^{(k)} Z\right), \quad Z \sim \mathcal{N}(0,1).$$

Limits $(a_t^{\infty}, b_t^{\infty})$ yield the asymptotic mean-squared error (MSE)

$$E_*(t) = \lim_{k\to\infty} \operatorname*{p-lim}_{n\to\infty} \mathbb{E}\|\widehat{\mathrm{m}}_t^{(k)} - \mathrm{x}\|_2^2.$$

To conclude $\widehat{\mathrm{m}}_t^{(k)} \approx \mathrm{m}_t$, we want:

$$E_*(t) \approx MMSE(t) \equiv \mathbb{E}\|\mathrm{m}_t - \mathrm{x}\|_2^2,$$

# State Evolution for AMP

State evolution: $i$-th coordinate of $\widehat{\mathrm{m}}_t^{(k)}$ behaves like

$$\tanh\left(a_t^{(k)} x_i + b_t^{(k)} Z\right), \quad Z \sim \mathcal{N}(0,1).$$

Limits $(a_t^{\infty}, b_t^{\infty})$ yield the asymptotic mean-squared error (MSE)

$$E_*(t) = \lim_{k\to\infty} \operatorname*{p-lim}_{n\to\infty} \mathbb{E}\|\widehat{\mathrm{m}}_t^{(k)} - \mathrm{x}\|_2^2.$$

To conclude $\widehat{\mathrm{m}}_t^{(k)} \approx \mathrm{m}_t$, we want:

$$E_*(t) \approx MMSE(t) \equiv \mathbb{E}\|\mathrm{m}_t - \mathrm{x}\|_2^2,$$

I-MMSE Area Law [Guo-Shamai-Verdu 04, Deshpande-Abbe-Montanari 15]:

$$\frac{1}{2}\int_0^{\infty} \mathrm{MMSE}(t)\,\mathrm{d}t = \mathrm{Ent}(\mathrm{x}^*).$$

- We verify explicitly that $\int_0^{\infty} E_*(t)\,\mathrm{d}t \approx \mathrm{Ent}(\mathrm{x}^*)$ for large $n$.

# Conclusion of Step 1: Rough Estimate for $\mathrm{m}_t$

$$\widehat{\mathrm{m}}_t^{(k+1)} = \tanh\left(\beta G \widehat{\mathrm{m}}_t^{(k)} + \mathrm{y}_t - b_k \widehat{\mathrm{m}}_t^{(k-1)}\right),$$

**Proposition (Alaoui-Montanari-S 22)**

For $\beta < 1$ and any $\varepsilon, t \geq 0$ there exists $k_0(t,\varepsilon)$ such that for all $k \geq k_0$,

$$\lim_{n\to\infty} \mathbb{P}\left[\|\widehat{\mathrm{m}}_t^{(k)}(y_t) - \mathrm{m}_t(y_t)\| \leq \varepsilon\sqrt{n}\right] = 1.$$

Here $y_t$ is perfect stochastic localization at time $t$. The algorithm can only use an estimate $\widehat{\mathrm{y}}_t$.

We still must bound $\|\widehat{\mathrm{m}}_t^{(k)}(y_t) - \widehat{\mathrm{m}}_t^{(k)}(\widehat{y}_t)\|$ to control **error accumulation across time**.

## Step 2: Refined Estimate of $\mathrm{m}_t$

Surprisingly, finishing the proof is non-obvious.

- Two types of error: SDE $\delta$-discretization and $\widehat{\mathrm{m}}_t^{(k)} \approx \mathrm{m}_t$.
- Sending $(\delta, k) \to (0, \infty)$ does not suffice.
  - Lipschitz constant of $\widehat{m}_t^{(k)}$ diverges with $k$.

# Step 2: Refined Estimate of $\mathrm{m}_t$

Surprisingly, finishing the proof is non-obvious.

- Two types of error: SDE $\delta$-discretization and $\widehat{\mathrm{m}}_t^{(k)} \approx \mathrm{m}_t$.
- Sending $(\delta, k) \to (0, \infty)$ does not suffice.
  - Lipschitz constant of $\widehat{m}_t^{(k)}$ diverges with $k$.

Second step: by construction, $\widehat{\mathrm{m}}_t^{(k)}$ is an approximate stationary point for the "TAP free energy":

$$F_{TAP}(\mathrm{m}, \mathrm{y}_t) = -\frac{\beta}{2}\langle \mathrm{m}, \mathrm{G}\mathrm{m} \rangle - \langle \widehat{\mathrm{y}}_t, \mathrm{m} \rangle - \sum_{i=1}^{n} h(m_i).$$

- With gradient descent, refine $\widehat{\mathrm{m}}_t^{(k)}$ to

$$\widehat{\mathrm{m}}_t^{\infty} = \arg\min_{\mathrm{m}} F_{TAP}(\mathrm{m}, \mathrm{y}_t).$$

# Step 2: Refined Estimate of $\mathrm{m}_t$

Surprisingly, finishing the proof is non-obvious.

- Two types of error: SDE $\delta$-discretization and $\widehat{\mathrm{m}}_t^{(k)} \approx \mathrm{m}_t$.
- Sending $(\delta, k) \to (0, \infty)$ does not suffice.
  - Lipschitz constant of $\widehat{m}_t^{(k)}$ diverges with $k$.

Second step: by construction, $\widehat{\mathrm{m}}_t^{(k)}$ is an approximate stationary point for the "TAP free energy":

$$F_{TAP}(\mathrm{m}, \mathrm{y}_t) = -\frac{\beta}{2}\langle \mathrm{m}, \mathrm{Gm} \rangle - \langle \widehat{\mathrm{y}}_t, \mathrm{m} \rangle - \sum_{i=1}^{n} h(m_i).$$

- With gradient descent, refine $\widehat{\mathrm{m}}_t^{(k)}$ to

$$\widehat{\mathrm{m}}_t^{\infty} = \arg\min_{\mathrm{m}} F_{TAP}(\mathrm{m}, \mathrm{y}_t).$$

- [Celentano 22]: $F_{TAP}$ is **strongly convex** near $\widehat{\mathrm{m}}_t^{\infty}$ for $\beta < 1$. Hence $\widehat{\mathrm{m}}_t^{\infty}$ is $C_\beta$-Lipschitz in $\widehat{\mathrm{y}}_t$. No blow-up with AMP accuracy.

# Algorithmic Stability

Our algorithm is **stable** with respect to $(\mathsf{G}, \beta)$: just uses $O_{\beta,\varepsilon}(1)$ matrix-vector products, and some 1-dimensional non-linearities.

Concretely, from i.i.d. $\mathsf{G} = \mathsf{G}_0$ and $\mathsf{G}_1$, consider perturbation path

$$\mathsf{G}_s = \sqrt{1-s^2}\mathsf{G}_0 + s\mathsf{G}_1.$$

Stability of the algorithm tells us:

$$\lim_{s\to 0}\lim_{n\to\infty}\mathbb{E}[W_{1,n}(\mu^{\mathrm{alg}}_{\mathsf{G}_0,\beta},\mu^{\mathrm{alg}}_{\mathsf{G}_s,\beta})] = 0.$$

# Algorithmic Stability

Our algorithm is **stable** with respect to $(\mathsf{G}, \beta)$: just uses $O_{\beta,\varepsilon}(1)$ matrix-vector products, and some 1-dimensional non-linearities.

Concretely, from i.i.d. $\mathsf{G} = \mathsf{G}_0$ and $\mathsf{G}_1$, consider perturbation path

$$\mathsf{G}_s = \sqrt{1-s^2}\mathsf{G}_0 + s\mathsf{G}_1.$$

Stability of the algorithm tells us:

$$\lim_{s\to 0} \lim_{n\to\infty} \mathbb{E}[W_{1,n}(\mu^{\mathrm{alg}}_{\mathsf{G}_0,\beta}, \mu^{\mathrm{alg}}_{\mathsf{G}_s,\beta})] = 0.$$

A purely structural consequence with an algorithmic proof:

**Theorem (Alaoui-Montanari-S 22; Celentano 22)**

*The* ***true*** *SK Gibbs measures are stable when $\beta < 1$:*

$$\lim_{s\to 0} \lim_{n\to\infty} \mathbb{E}[W_{1,n}(\mu_{\mathsf{G}_0,\beta}, \mu_{\mathsf{G}_s,\beta})] = 0.$$

*Similar stability holds for small pertubations in $\beta$.*

# Hardness via Chaos

The stability property

$$\lim_{s\to 0}\lim_{n\to\infty}\mathbb{E}[W_{1,n}(\mu_{G_0,\beta},\mu_{G_s,\beta})]=0.$$

for the true Gibbs measure is **false** for $\beta>1$. Combination of:

# Hardness via Chaos

The stability property

$$\lim_{s \to 0} \lim_{n \to \infty} \mathbb{E}[W_{1,n}(\mu_{G_0,\beta}, \mu_{G_s,\beta})] = 0.$$

for the true Gibbs measure is **false** for $\beta > 1$. Combination of:

**Theorem (Chatterjee 09; Disorder Chaos)**

*Let* $(x_0, x_s) \sim \mu_{G_0,\beta} \times \mu_{G_s,\beta}$. *For all* $\beta \in \mathbb{R}$ *and* $s > 0$,

$$\lim_{n \to \infty} \mathbb{E}[|\langle x_0, x_s \rangle|/n] = 0.$$

# Hardness via Chaos

The stability property

$$\lim_{s\to 0}\lim_{n\to\infty}\mathbb{E}[W_{1,n}(\mu_{G_0,\beta},\mu_{G_s,\beta})]=0.$$

for the true Gibbs measure is **false** for $\beta > 1$. Combination of:

**Theorem (Chatterjee 09; Disorder Chaos)**

*Let* $(x_0, x_s) \sim \mu_{G_0,\beta} \times \mu_{G_s,\beta}$. *For all* $\beta \in \mathbb{R}$ *and* $s > 0$,

$$\lim_{n\to\infty}\mathbb{E}[|\langle x_0, x_s\rangle|/n]=0.$$

**Theorem (Replica Symmetry Breaking)**

*Let* $x_0, x_0' \sim \mu_{G_0,\beta}$ *be independent. For all* $\beta > 1$,

$$\liminf_{n\to\infty}\mathbb{E}[|\langle x_0, x_0'\rangle|/n]\geq c(\beta)>0.$$

# Hardness via Chaos

The previous results show that $\mu_{G_0,\beta}$ and $\mu_{G_s,\beta}$ must be significantly different. Therefore:

**Theorem (Alaoui-Montanari-S 22)**

*Let $\mu^{\mathrm{alg}}_{G,\beta}$ be the law of $\mathrm{ALG}_n(G,\beta,\omega)$ conditional on G. If $\mathrm{ALG}_n$ is **stable**, then for all $\beta > 1$,*

$$\liminf_{n\to\infty} \mathbb{E}[W_{1,n}(\mu^{\mathrm{alg}}_{G,\beta}, \mu_{G,\beta})] > c(\beta) > 0.$$

Stability holds for gradient-based methods such as Langevin dynamics and AMP, at least on **dimension-independent** time-scales.

# Extension to $p$-spin Models

Instead of a random matrix, start with a Gaussian **tensor**

$$\mathsf{G}^{(p)} \sim N^{-(p-1)/2} \cdot \mathcal{N}(0, I_{n^p}).$$

The pure $p$-spin glass distribution is:

$$\mu_{p,\beta}(\mathsf{x}) = \frac{1}{Z_{p,n}(\beta)} e^{\beta\langle \mathsf{x}^{\otimes p}, \mathsf{G}^{(p)}\rangle}.$$

# Extension to $p$-spin Models

Instead of a random matrix, start with a Gaussian **tensor**

$$\mathrm{G}^{(p)} \sim N^{-(p-1)/2} \cdot \mathcal{N}(0, I_{n^p}).$$

The pure $p$-spin glass distribution is:

$$\mu_{p,\beta}(\mathrm{x}) = \frac{1}{Z_{p,n}(\beta)} e^{\beta \langle \mathrm{x}^{\otimes p}, \mathrm{G}^{(p)} \rangle}.$$

Physics belief: sampleable for $\beta = \beta_{dyn}(p) \approx \sqrt{\frac{1}{p}}$.

[ABXY 22, AJKPV 23]: Glauber mixes fast for $\beta \ll p^{-3/2}$.

[Alaoui-Montanari-**S** 23]: stochastic localization succeeds for $\beta \ll \frac{1}{p}$.

However, replica-symmetric below $\beta_c(p) \approx \sqrt{2 \log 2} \gg \beta_{dyn}(p)$.

# Shattering

$\beta_c(p) \gg \beta_{dyn}(p)$ is expected for large $p$ due to **shattering**.
This means there are disjoint clusters $\mathcal{C}_1, \ldots, \mathcal{C}_M \subseteq \{-1, +1\}^n$ with...

1. Small diameter and probability:

$$\max_{1\le m\le M} diam(\mathcal{C}_m) \le \varepsilon\sqrt{N}, \qquad \max_{1\le m\le M} \mu_\beta(\mathcal{C}_m) \le e^{-cN}.$$

# Shattering

$\beta_c(p) \gg \beta_{dyn}(p)$ is expected for large $p$ due to **shattering**.
This means there are disjoint clusters $\mathcal{C}_1, \dots, \mathcal{C}_M \subseteq \{-1, +1\}^n$ with...

1. Small diameter and probability:

$$\max_{1 \le m \le M} diam(\mathcal{C}_m) \le \varepsilon\sqrt{N}, \qquad \max_{1 \le m \le M} \mu_\beta(\mathcal{C}_m) \le e^{-cN}.$$

2. Uniform separation:

$$\min_{1 \le m_1 < m_2 \le M} dist(\mathcal{C}_{m_1}, \mathcal{C}_{m_2}) \ge 10\varepsilon\sqrt{N}.$$

# Shattering

$\beta_c(p) \gg \beta_{dyn}(p)$ is expected for large $p$ due to **shattering**.
This means there are disjoint clusters $\mathcal{C}_1, \ldots, \mathcal{C}_M \subseteq \{-1, +1\}^n$ with...

1. Small diameter and probability:

$$\max_{1 \le m \le M} diam(\mathcal{C}_m) \le \varepsilon\sqrt{N}, \qquad \max_{1 \le m \le M} \mu_\beta(\mathcal{C}_m) \le e^{-cN}.$$

2. Uniform separation:

$$\min_{1 \le m_1 < m_2 \le M} dist(\mathcal{C}_{m_1}, \mathcal{C}_{m_2}) \ge 10\varepsilon\sqrt{N}.$$

3. Together, the clusters account for nearly all the probability:

$$\mu_\beta\Big(\bigcup_{m=1}^{M} \mathcal{C}_m\Big) \ge 1 - e^{-cN}.$$

# Hardness for $p$-spin Sampling

For $\beta > \beta_c$, we still have "RSB $\implies$ chaos $\implies$ hardness". Below $\beta_c$...

# Hardness for $p$-spin Sampling

For $\beta > \beta_c$, we still have "RSB $\implies$ chaos $\implies$ hardness". Below $\beta_c$...

**Theorem (Gamarnik-Jagannath-Kizildag 23)**

*Pure p-spin glasses shatter for* $0.51\beta_c(p) < \beta < 0.99\beta_c(p)$ *and* $p \geq O(1)$.

**Theorem (Alaoui-Montanari-S 23b)**

*For spin glasses, "shattering $\implies$ chaos $\implies$ hardness".*

- Noising $G^{(p)}$ re-randomizes cluster weight ratios $\mu_\beta(\mathcal{C}_i)/\mu_\beta(\mathcal{C}_j)$.

# Hardness for $p$-spin Sampling

For $\beta > \beta_c$, we still have "RSB $\implies$ chaos $\implies$ hardness". Below $\beta_c$...

**Theorem (Gamarnik-Jagannath-Kizildag 23)**

*Pure p-spin glasses shatter for $0.51\beta_c(p) < \beta < 0.99\beta_c(p)$ and $p \geq O(1)$.*

**Theorem (Alaoui-Montanari-S 23b)**

*For spin glasses, "shattering $\implies$ chaos $\implies$ hardness".*

- Noising $G^{(p)}$ re-randomizes cluster weight ratios $\mu_\beta(\mathcal{C}_i)/\mu_\beta(\mathcal{C}_j)$.

For spherical analogs: $\beta_{dyn}^{sph}(p) \approx \sqrt{e} \ll \beta_c^{sph}(p) \approx \sqrt{\log p}$.

**Theorem (Alaoui-Montanari-S 23b)**

*Spherical p-spin glasses shatter for $\beta \in [O(1), \beta_c^{sph}(p))$.*

Sharp thresholds (algorithmic and mathematical) open beyond SK.

# Summary

Stochastic localization sampling for the SK model

$$\mu_{G,\beta}(x) = \frac{1}{Z_n(\beta)} e^{\beta \langle x, Gx \rangle / 2}.$$

Approach: to obtain $x^* \sim \mu$, simulate $y_t = t x^* + B_t$.

# Summary

Stochastic localization sampling for the SK model

$$\mu_{G,\beta}(x) = \frac{1}{Z_n(\beta)} e^{\beta\langle x, Gx\rangle/2}.$$

Approach: to obtain $x^* \sim \mu$, simulate $y_t = t x^* + B_t$.

Main result: Wasserstein-approximate samples for $\beta < 1$. For $\beta > 1$, disorder chaos is a natural barrier for stable algorithms.

- For general $p$-spin models, sharp thresholds will require an understanding of **shattering**.
- Upgrade to TV sampling?
- What other distributions are stochastic localization sampleable?

# Summary

Stochastic localization sampling for the SK model

$$\mu_{G,\beta}(x) = \frac{1}{Z_n(\beta)} e^{\beta\langle x, Gx\rangle/2}.$$

Approach: to obtain $x^* \sim \mu$, simulate $y_t = tx^* + B_t$.

Main result: Wasserstein-approximate samples for $\beta < 1$. For $\beta > 1$, disorder chaos is a natural barrier for stable algorithms.

- For general $p$-spin models, sharp thresholds will require an understanding of **shattering**.
- Upgrade to TV sampling?
- What other distributions are stochastic localization sampleable?

Other provable implementations of diffusion sampling:

- [Montanari-Wu 23]: posterior sampling for noisy low-rank matrices.
- [AHLVXY 23]: TV sampling for structured $\mu$, e.g. DPPs.