# OPTIMIZATION OF MEAN-FIELD SPIN GLASSES


**AHMED EL ALAOUI**

**STANFORD UNIVERSITY**


Joint work /

&

Andrea Montanari

Mark Sellke

Youth in High Dimensions, ICTP.

July 3, 2020.

# The Ising mixed p-spin Hamiltonian

$(H_N(\sigma))_{\sigma\in\Sigma_N}$ centered Gaussian process with covariance

$$\mathbb{E}\big[H_N(\sigma)H_N(\sigma')\big] = N\xi\big(\langle\sigma,\sigma'\rangle/N\big)$$

# The Ising mixed p-spin Hamiltonian

$(H_N(\sigma))_{\sigma\in\Sigma_N}$ centered Gaussian process with covariance

$$\mathbb{E}\big[H_N(\sigma)H_N(\sigma')\big] = N\xi\big(\langle\sigma,\sigma'\rangle/N\big)$$

**2-spin (SK)**

$$H_N(\sigma) = \frac{1}{\sqrt{N}}\sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j$$

$$\xi(x) = x^2$$

**3-spin**

$$H_N(\sigma) = \frac{1}{N}\sum_{i<j<k} J_{ijk}\sigma_i\sigma_j\sigma_k$$

$$\xi(x) = x^3$$

# The Ising mixed p-spin Hamiltonian

$(H_N(\sigma))_{\sigma \in \Sigma_N}$ centered Gaussian process with covariance

$$\mathbb{E}\big[H_N(\sigma)H_N(\sigma')\big] = N\xi\big(\langle \sigma, \sigma' \rangle / N\big)$$

**2-spin (SK)**

$$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j$$

$$\xi(x) = x^2$$

**3-spin**

$$H_N(\sigma) = \frac{1}{N} \sum_{i<j<k} J_{ijk}\sigma_i\sigma_j\sigma_k$$

$$\xi(x) = x^3$$

$$\text{maximize} \qquad H_N(\sigma) \qquad \text{s.t.} \quad \sigma \in \Sigma_N = \{-1, +1\}^N$$

**(Ground state)**

# Optimization

$$\text{maximize} \qquad H_N(\sigma) \qquad \text{s.t.} \quad \sigma \in \Sigma_N = \{-1,+1\}^N$$

- Can we **efficiently** find an (approximate) ground state?
- **Input:** Gradient oracle for $H_N$. Tolerance $\epsilon > 0$.
- **Output:** $\sigma^{\text{alg}} = \sigma^{\text{alg}}(\text{queries}) \in \{\pm 1\}^N$ such that

$$H_N\left(\sigma^{\text{alg}}\right) \geq (1-\epsilon) \max_{\sigma \in \Sigma_N} H_N(\sigma) \quad \text{w.h.p.}$$

- **Runtime:** $\#\text{queries} = \text{Poly}(N, 1/\epsilon)$.

# Optimization

$$\text{maximize} \qquad H_N(\sigma) \qquad \text{s.t.} \quad \sigma \in \Sigma_N = \{-1,+1\}^N$$

- Can we **efficiently** find an (approximate) ground state?
- **Input:** Gradient oracle for $H_N$. Tolerance $\epsilon > 0$.
- **Output:** $\sigma^{\text{alg}} = \sigma^{\text{alg}}(\text{queries}) \in \{\pm 1\}^N$ such that

$$H_N\left(\sigma^{\text{alg}}\right) \geq (1-\epsilon) \max_{\sigma \in \Sigma_N} H_N(\sigma) \quad \text{w.h.p.}$$

- **Runtime:** $\#\text{queries} = \text{Poly}(N, 1/\epsilon).$

**Worst case:** Already NP-hard for quadratics and for $1-\epsilon \geq 1/(\log N)^c$

[Arora et al. 2005]

# What is known

**Convex relaxations:** (Pure k-spin spherical model)

Level k Sum of Square relaxation: $\mathsf{SOS}_N(k) \gtrsim N^{(k-2)/4} \cdot \mathsf{OPT}_N$

[Bhattiprolu, Guruswami, Lee 2016]

**Langevin/Glauber dynamics:** Rich literature. Slow dynamics. Gets stuck at **threshold energy**.

[Cugliandolo, Kurchan 92] [Bouchaud et al. 98] [Ben Arous, Guionnet 95] [Ben Arous, Dembo, Guionnet 2001]…

**“Dual” algorithms:** Spherical model [Subag 2018]

Ising SK (p=2) model [Montanari 2018]

⟶ Both achieve approximate global optimum under **no overlap gap assumption.**

Gamarnik & coauthors: Impossibility results if **overlap gap**; (Subhabrata’s talk)

# Asymptotic value: Zero temperature Parisi formula

$$\mathcal{U} \equiv \left\{\gamma : [0,1) \to \mathbb{R}_{\geq 0} \text{ non-decreasing}, \int_0^1 \gamma(t)\mathrm{d}t < \infty\right\}$$

$$\partial_t \Phi_\gamma(t,x) + \frac{\xi''(t)}{2}\Big(\gamma(t)\big(\partial_x \Phi_\gamma(t,x)\big)^2 + \partial_x^2 \Phi_\gamma(t,x)\Big) = 0$$

**(Parisi PDE)**

$$(t,x) \in [0,1] \times \mathbb{R} \qquad \Phi_\gamma(1,x) = |x|$$

$$\mathsf{P}(\gamma) \equiv \Phi_\gamma(0,0) - \frac{1}{2}\int_0^1 t\xi''(t)\gamma(t)\mathrm{d}t$$

**(Parisi functional)**

# Asymptotic value: Zero temperature Parisi formula

$$\mathcal{U} \equiv \left\{ \gamma : [0,1) \to \mathbb{R}_{\geq 0} \text{ non-decreasing}, \int_0^1 \gamma(t)\mathrm{d}t < \infty \right\}$$

$$\partial_t \Phi_\gamma(t,x) + \frac{\xi''(t)}{2}\Big(\gamma(t)\big(\partial_x \Phi_\gamma(t,x)\big)^2 + \partial_x^2 \Phi_\gamma(t,x)\Big) = 0$$

**(Parisi PDE)**

$$(t,x) \in [0,1] \times \mathbb{R} \qquad \Phi_\gamma(1,x) = |x|$$

$$\mathsf{P}(\gamma) \equiv \Phi_\gamma(0,0) - \frac{1}{2}\int_0^1 t\xi''(t)\gamma(t)\mathrm{d}t$$

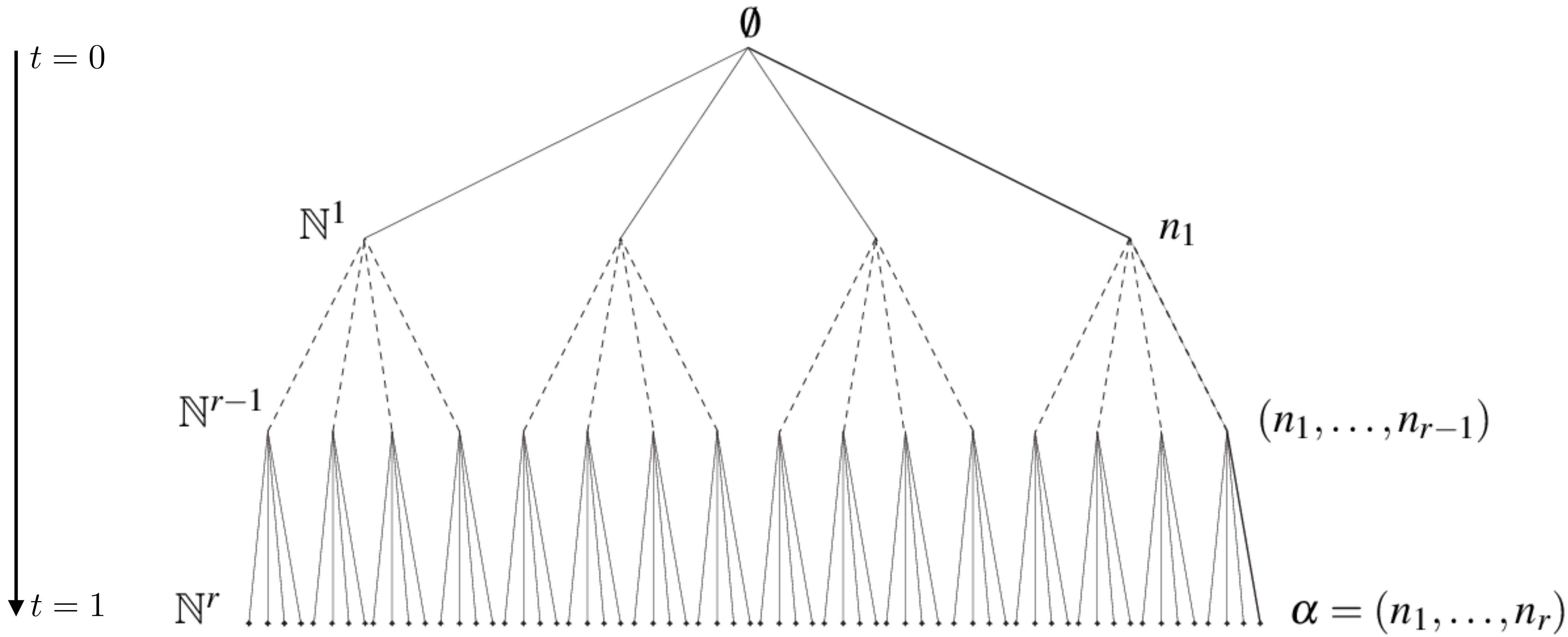
**(Parisi functional)**

**Theorem** [Auffinger, Chen 17]**:**

$$\frac{1}{N} \max_{\sigma \in \Sigma_N} H_N(\sigma) \xrightarrow[N \to \infty]{p} \inf_{\gamma \in \mathcal{U}} \mathsf{P}(\gamma)$$

**(GS Energy)**

# Ultrametric structure




1. Clusters of near-optima in $\{\pm 1\}^N$ are represented by *leaves* of an infinitary tree

2. Internal (ancestor) nodes of the tree are associated to points $m \in [-1, 1]^N$

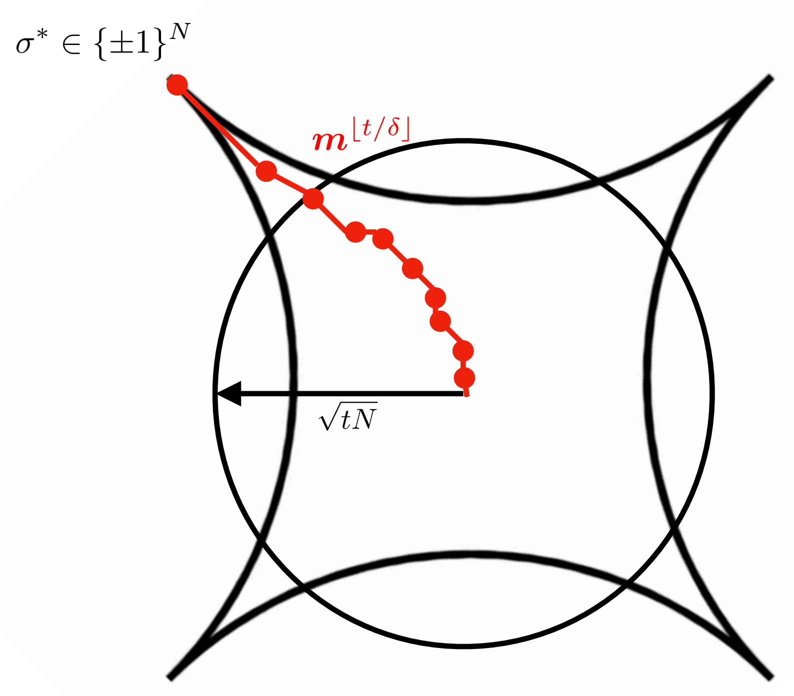
At level t: $\frac{1}{N}\|m\|_2^2 \simeq t$

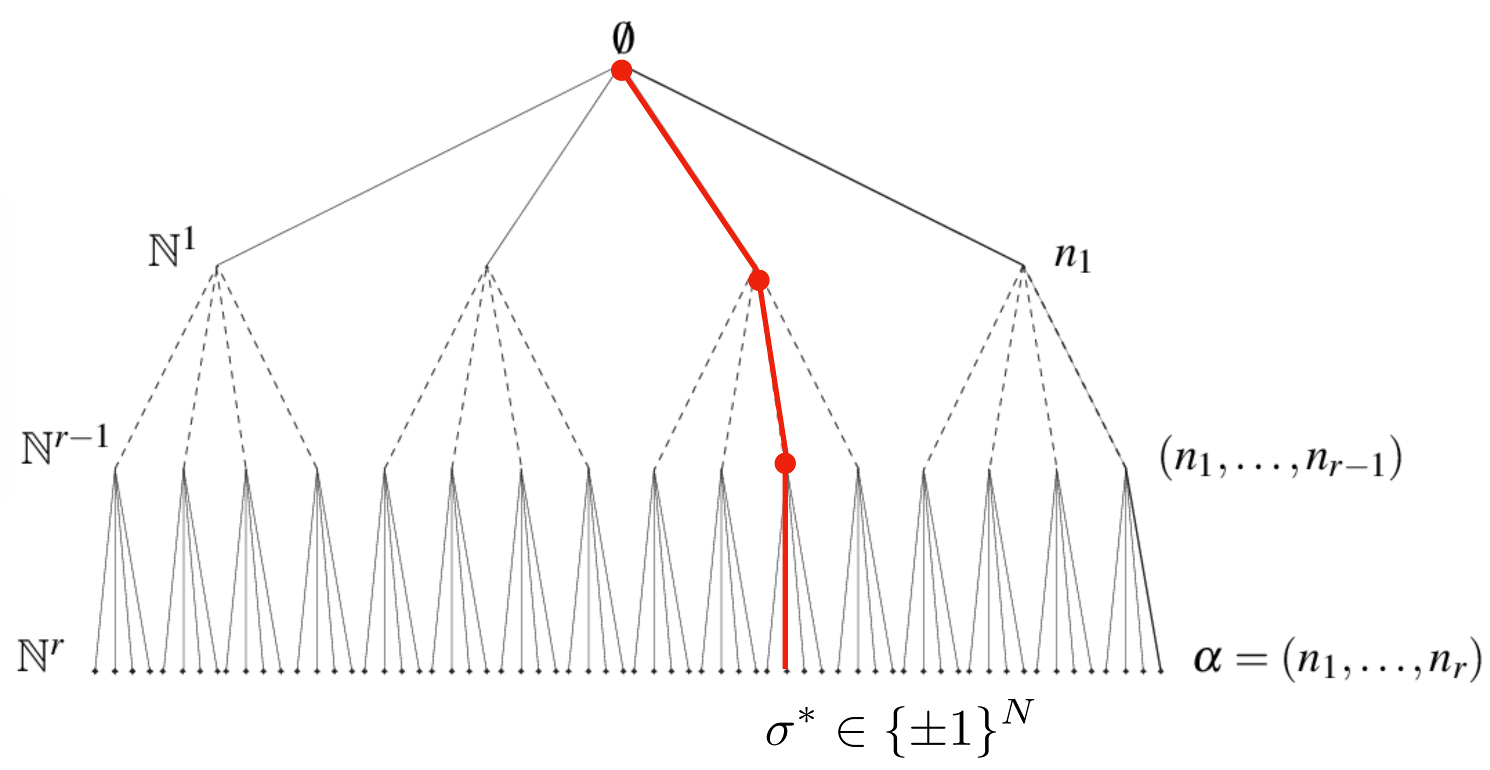
3. Euclidean distance btw clusters is reflected by the tree distance btw nodes

[Mezard, Parisi, Toulouse, Virasoro 84] [Panchenko 13] [Chen, Panchenko, Subag 19]

# Main algorithmic idea:

Start at the root and explore a random path in the tree



# Main result

*Extended* Parisi formula

[**EA,** Montanari, Sellke 2020]

There exist **ALG** which outputs $\sigma^{\text{alg}} \in \{-1,+1\}^N$ in $C(\epsilon)$ iterations such that

$$H_N(\sigma^{\text{alg}}) \geq (1-\epsilon) \inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma)$$

$\mathcal{U} \subset \mathcal{L}$ contains non-monotone functions.

# Main result

*Extended* Parisi formula

[**EA,** Montanari, Sellke 2020]

There exist **ALG** which outputs $\sigma^{\mathrm{alg}} \in \{-1,+1\}^N$ in $C(\epsilon)$ iterations such that

$$H_N(\sigma^{\mathrm{alg}}) \geq (1-\epsilon) \inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma)$$

$\mathcal{U} \subset \mathcal{L}$ contains non-monotone functions.

**Optimal,** i.e. $\inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma) = E_{\mathsf{GS}}$ **if** infimum is achieved at a non-decreasing function

# Main result

*Extended* Parisi formula

[**EA,** Montanari, Sellke 2020]

There exist **ALG** which outputs $\sigma^{\text{alg}} \in \{-1,+1\}^N$ in $C(\epsilon)$ iterations such that

$$H_N(\sigma^{\text{alg}}) \geq (1-\epsilon) \inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma)$$

$\mathcal{U} \subset \mathcal{L}$ contains non-monotone functions.

**Optimal,** i.e. $\inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma) = E_{\mathsf{GS}}$ **if** infimum is achieved at a non-decreasing function

In particular, if $\inf_{\gamma \in \mathcal{U}} \mathsf{P}(\gamma)$ is achieved at a strictly increasing function

→ **No overlap gap, continuous/full replica symmetry breaking**

# Main result

*Extended* Parisi formula

[**EA,** Montanari, Sellke 2020]

There exist **ALG** which outputs $\sigma^{\text{alg}} \in \{-1,+1\}^N$ in $C(\epsilon)$ iterations such that

$$H_N(\sigma^{\text{alg}}) \geq (1-\epsilon) \inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma)$$

$\mathcal{U} \subset \mathcal{L}$ contains non-monotone functions.

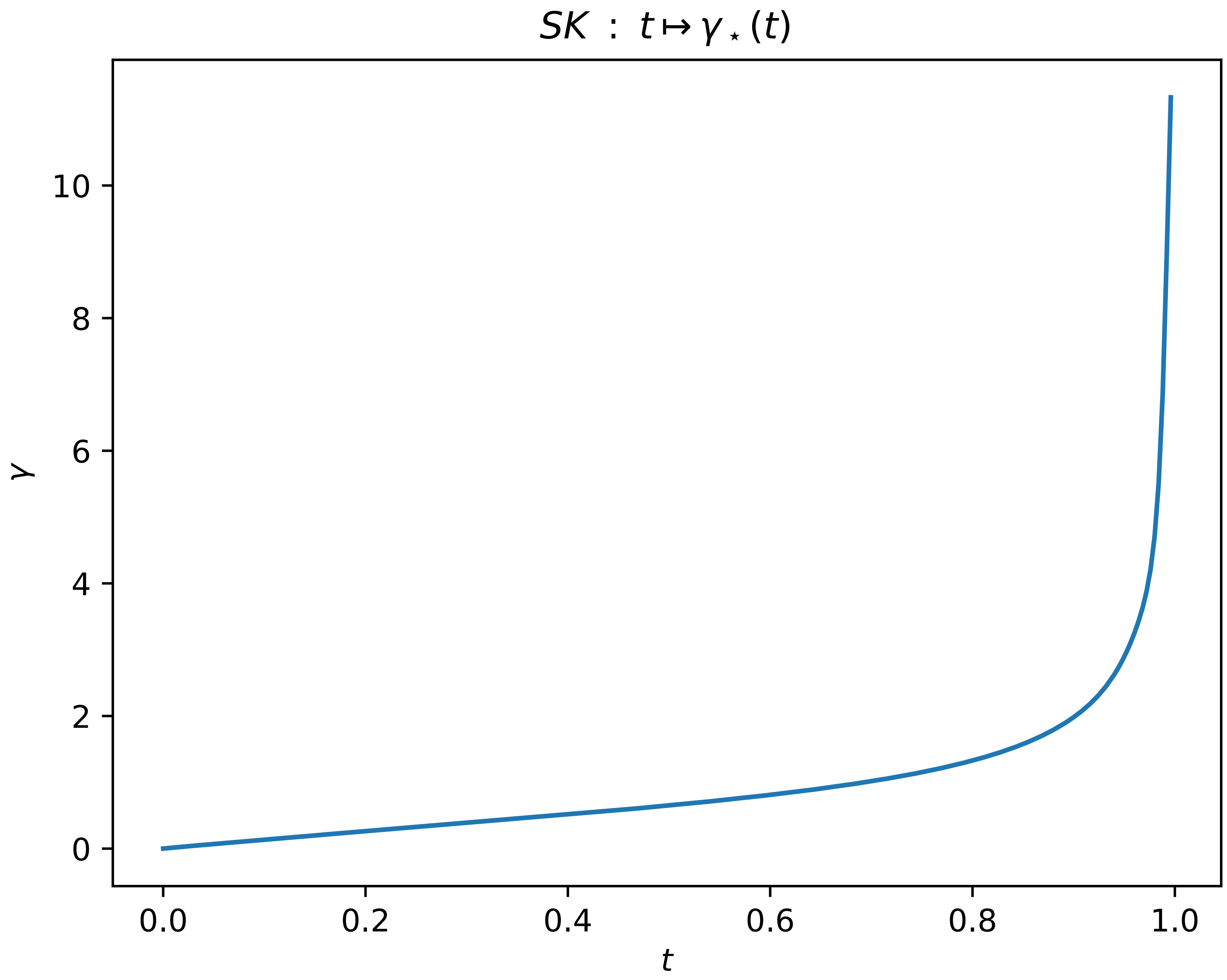
**Best value algorithmically possible:**

NO **AMP** algorithm can achieve a value larger than $\inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma)$

⟶ An **algorithmic threshold** for optimization in spin glasses

# 2-spin/SK

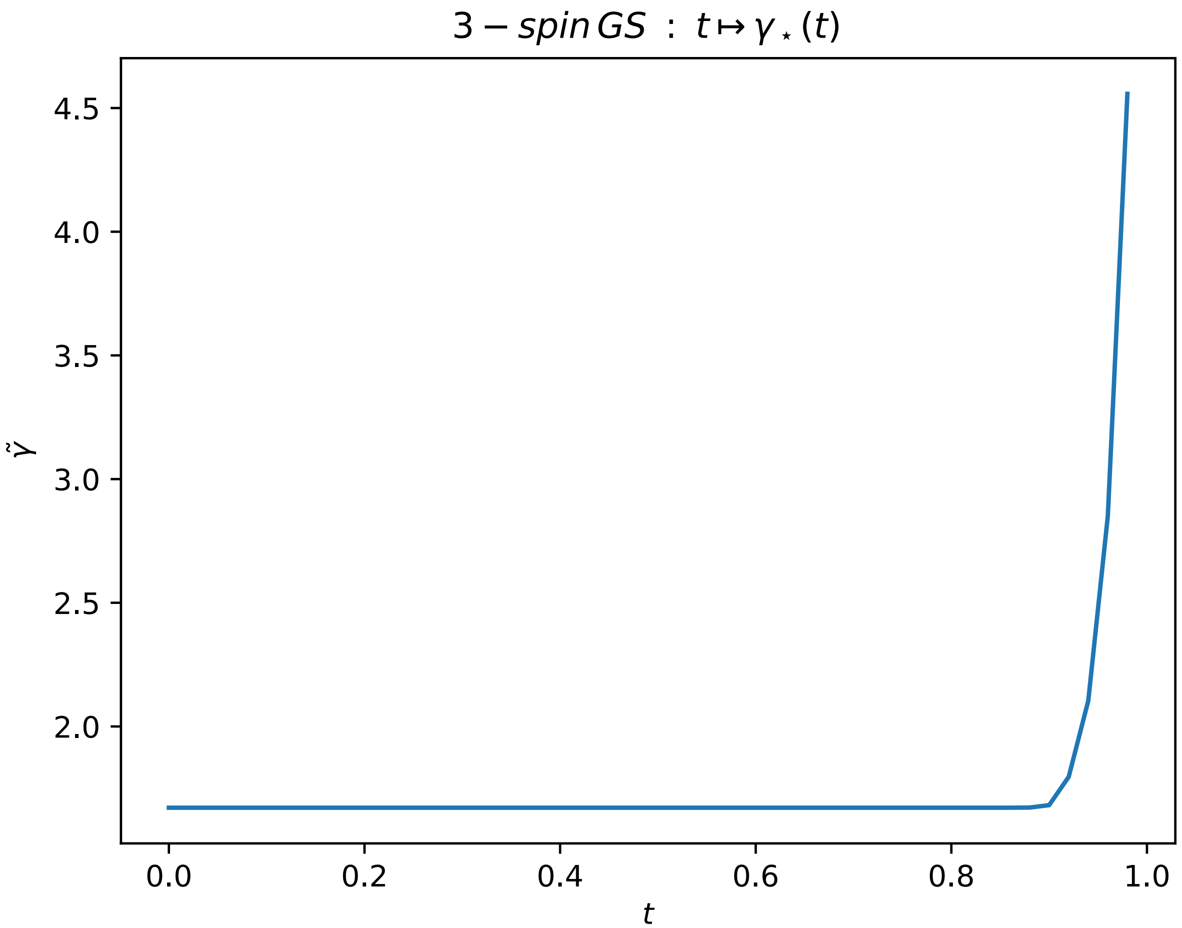
$$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j$$

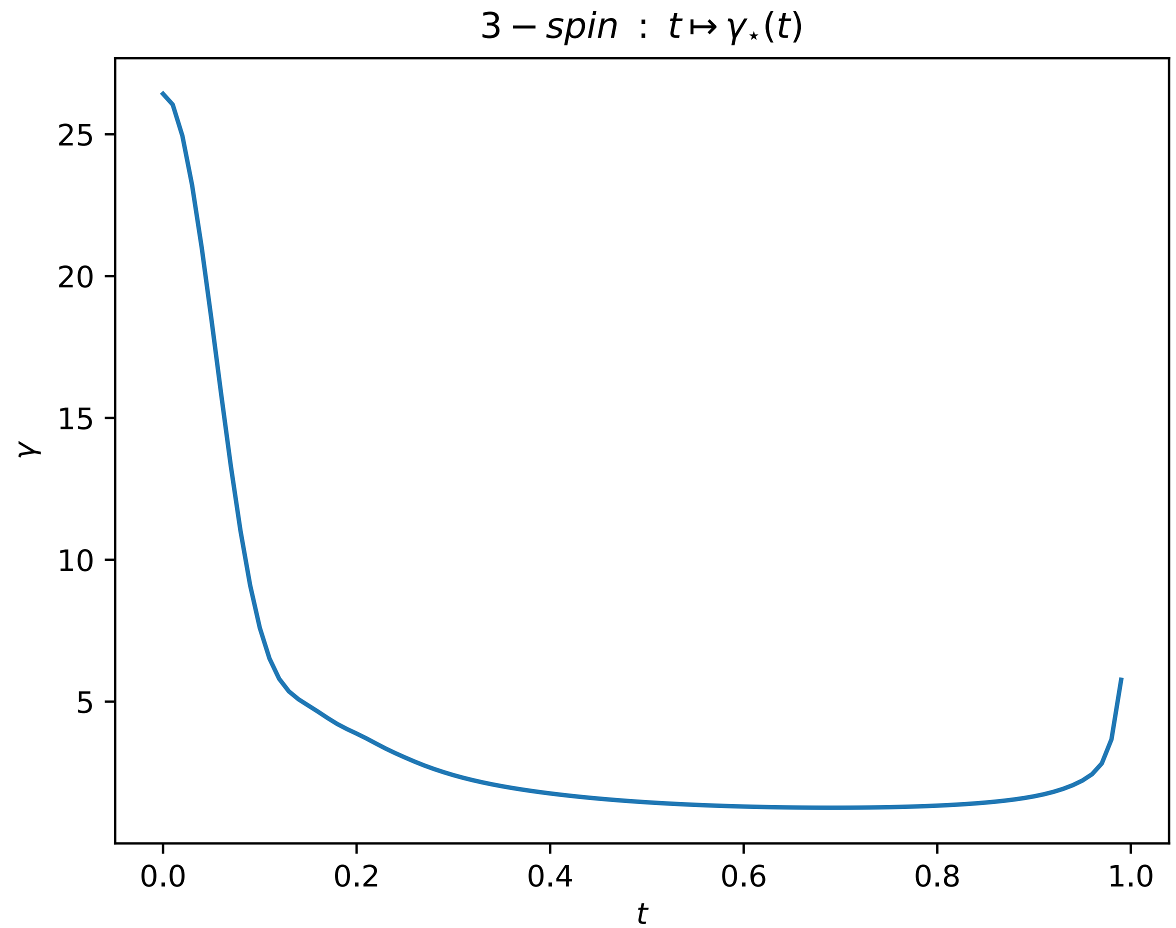



$$E_{\mathsf{GS}} = E_{\mathsf{ALG}} = \mathsf{P}(\gamma_\star) \simeq .76316$$

# 3-spin

$$H_N(\sigma) = \frac{1}{N} \sum_{i<j<k} J_{ijk}\sigma_i\sigma_j\sigma_k$$



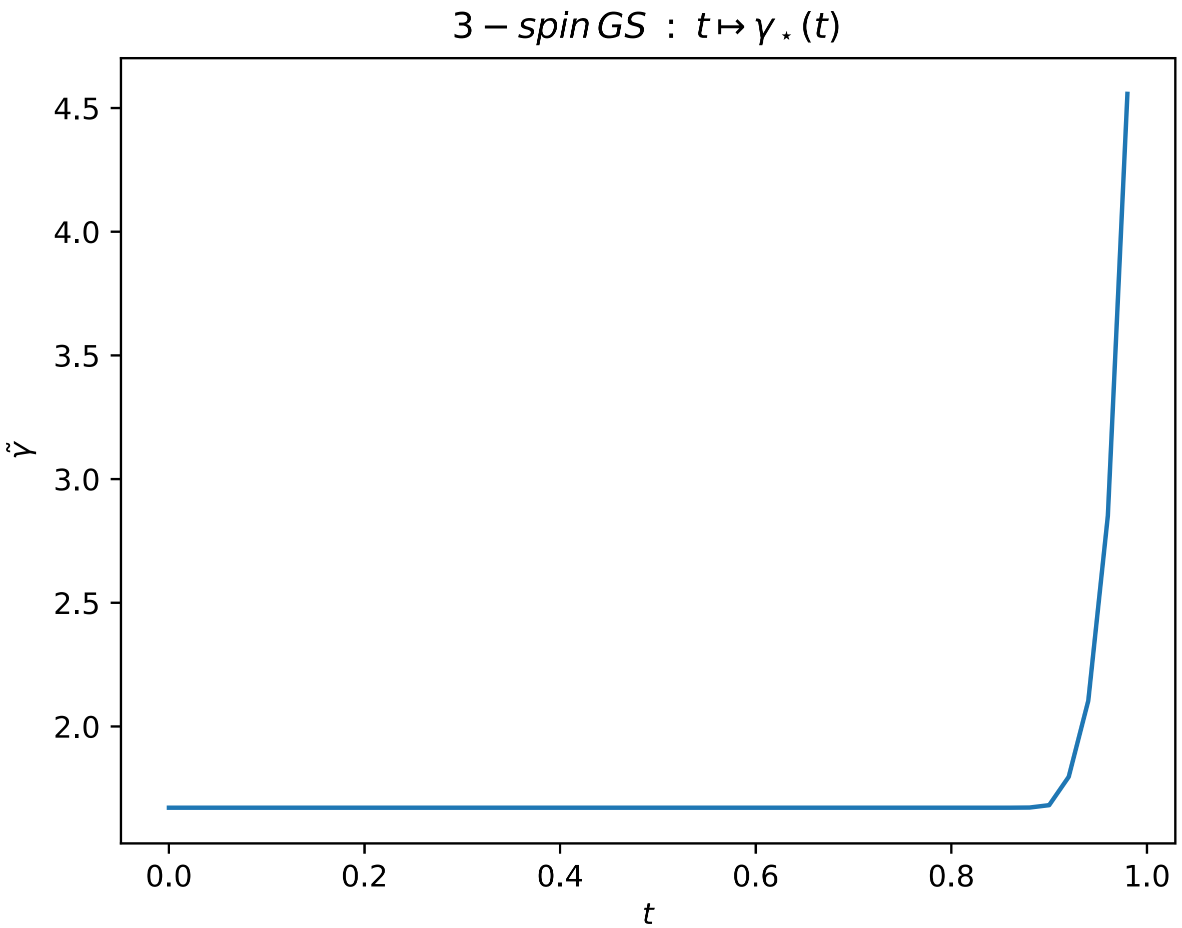

$$E_{\mathsf{GS}} = \inf_{\gamma \in \mathcal{U}} \mathsf{P}(\gamma) \simeq .8134$$

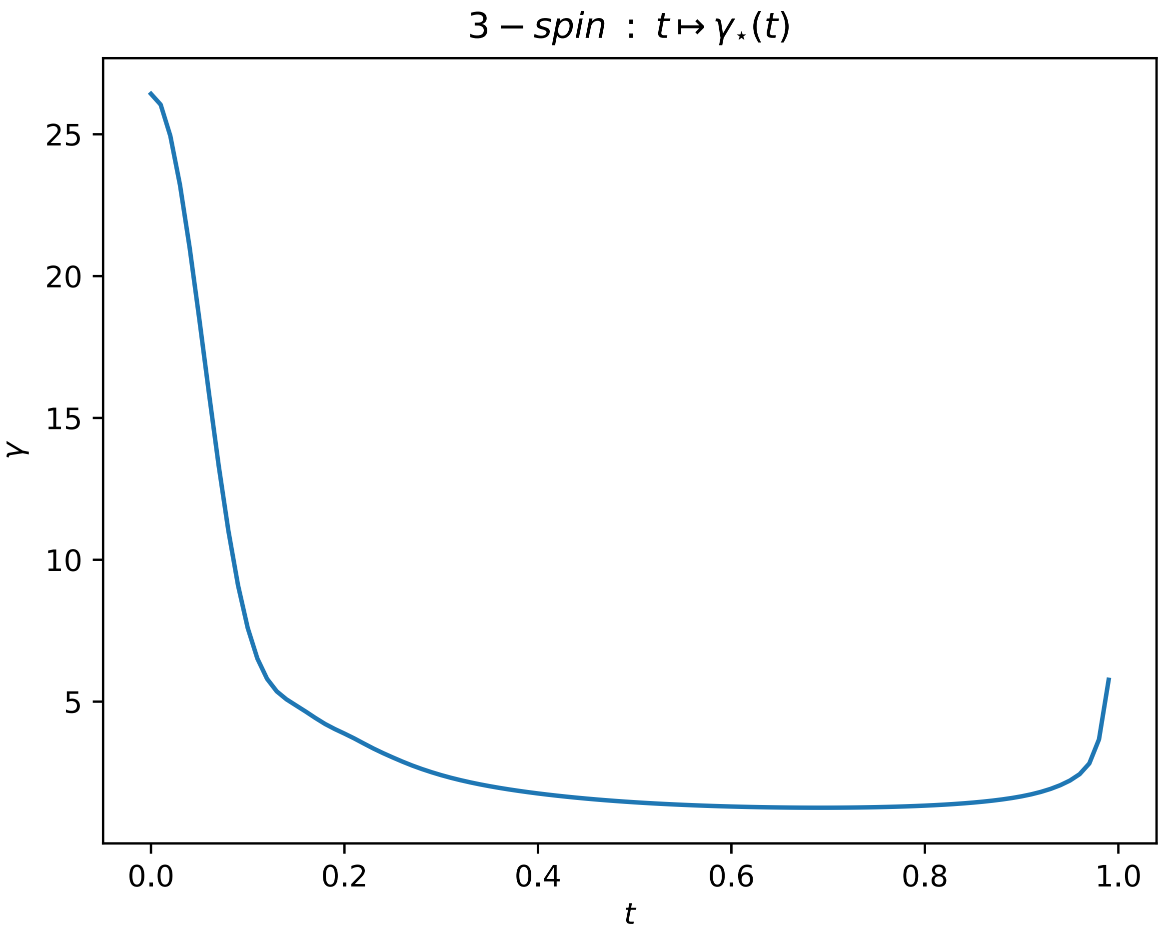
$$E_{\mathsf{ALG}} = \inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma) \simeq .8006$$

$$E_{\mathsf{ALG}} < E_{\mathsf{GS}}$$

# 3-spin

$$H_N(\sigma) = \frac{1}{N} \sum_{i<j<k} J_{ijk}\sigma_i\sigma_j\sigma_k$$



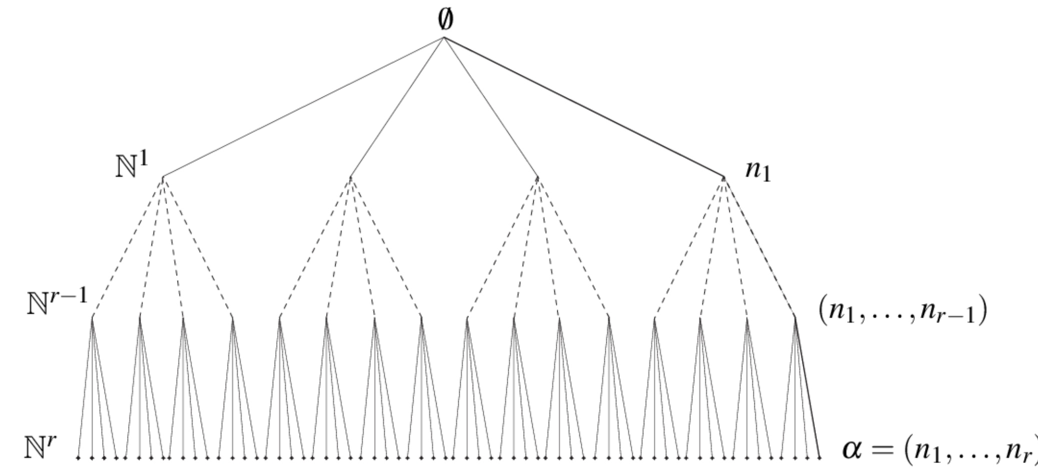

$$E_{\mathsf{GS}} = \inf_{\gamma \in \mathcal{U}} \mathsf{P}(\gamma) \simeq .8134$$

$$E_{\mathsf{ALG}} = \inf_{\gamma \in \mathcal{L}} \mathsf{P}(\gamma) \simeq .8006$$

$$.7882 \simeq E_{\mathsf{thr}} < E_{\mathsf{ALG}} < E_{\mathsf{GS}}$$

[Rizzo 2013]

# Stochastic formulation




(2-spin) $$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j$$

Effective (cavity) field: $$\mathrm{d}X_t = \gamma(t)\partial_x\Phi_\gamma(t, X_t)\mathrm{d}t + \mathrm{d}B_t \qquad X_0 = 0$$

Effective magnetization: $$M_t = \partial_x\Phi_\gamma(t, X_t) = \int_0^t \partial_x^2\Phi_\gamma(s, X_s)\mathrm{d}B_s$$

RS: $X_q \sim N(0, q), \qquad M_q = \tanh(\beta X_q + h)$

Discretize SDE with driving noise coming from $\boldsymbol{J} = (J_{i,j})$ in the “right” way.

**AMP**

# *Incremental* Approximate Message Passing

$t \in \{0, \delta, 2\delta, \cdots, \ell\delta, \cdots, 1\}$

$$\boldsymbol{z}^0 = 0$$
$$\boldsymbol{m}^\ell = f_\ell(\boldsymbol{z}^0, \cdots, \boldsymbol{z}^\ell)$$
$$\boldsymbol{z}^{\ell+1} = \frac{1}{\sqrt{N}}\boldsymbol{J}\boldsymbol{m}^\ell - \sum_{j=0}^{\ell} d_{\ell,j}\boldsymbol{m}^{j-1}$$

**State Evolution:** $\boldsymbol{z}^0, \cdots, \boldsymbol{z}^\ell$ converge to a centered Gaussian vector with cov.

$$\mathbb{E}\big[Z^{\ell+1}Z^{\ell'+1}\big] = \mathbb{E}\big[f_\ell(Z^0, \cdots, Z^\ell)f_{\ell'}(Z^0, \cdots, Z^{\ell'})\big].$$

**Choice of non-linearities:** $f_\ell : (z^0, \cdots, z^\ell) \mapsto m^\ell$

$$x^0 = 0$$
$$x^{j+1} - x^j = v(\delta j, x^j)\delta + (z^{j+1} - z^j) \quad \text{for} \quad 0 \le j \le \ell$$
$$m^\ell = \sum_{j=0}^{\ell-1} u(\delta j, x^j)(z^{j+1} - z^j)$$

# Self-consistency condition

Assume $Z^0, \cdots, Z^\ell$ has independent increments with $\mathbb{E}\big[(Z^{j+1} - Z^j)^2\big] = \delta \quad \forall j \leq \ell - 1$

Show $\quad \mathbb{E}\big[\big(Z^{\ell+1} - Z^\ell\big) Z^j\big] = 0 \quad \forall j \leq \ell \quad$ and $\quad \mathbb{E}\big[\big(Z^{\ell+1} - Z^\ell\big)^2\big] = \delta$

<u>Necessary condition:</u> $\quad \mathbb{E}\big[u(\delta\ell, X^\ell)^2\big] = 1 \quad \forall \ell$

$$\implies \qquad \mathbb{E}\big[u(t, X_t)^2\big] = 1 \quad \forall t \in [0, 1]$$

# An algorithm design question

Energy achieved: $$N^{-1}H_N(\boldsymbol{m}^{\lfloor \delta^{-1} \rfloor}) \xrightarrow[N\to\infty]{\delta\to 0} \int_0^1 \mathbb{E}\big[u(t,X_t)\big]\mathrm{d}t$$

Necessary condition: $$\mathbb{E}\big[u(t,X_t)^2\big] = 1 \quad \forall t \in [0,1]$$

Ising spins: $$M_1 = \int_0^1 u(s,X_s)\mathrm{d}B_s \in (-1,+1) \;\; \text{a.s.}$$

# An algorithm design question

$$\max_{u,v} \quad \int_0^1 \mathbb{E}\big[u(t, X_t)\big]\mathrm{d}t$$

subj. to

$$\mathbb{E}\big[u(t, X_t)^2\big] = 1 \quad \forall t \in [0, 1]$$

$$M_1 = \int_0^1 u(s, X_s)\mathrm{d}B_s \in (-1, +1) \quad \text{a.s.}$$

# An algorithm design question

$$\mathscr{E}_\star = \max_{u \in D[0,1]} \int_0^1 \mathbb{E}[u_t]\mathrm{d}t$$

subj. to

$$\mathbb{E}[u_t^2] = 1 \quad \forall t \in [0,1]$$

$$M_1 = \int_0^1 u_s \mathrm{d}B_s \in (-1,+1) \ \text{ a.s.}$$

# Relaxed stochastic control problem

$$\mathscr{E}_\star \leq \mathsf{REL}$$

$$\mathsf{REL} \equiv \sup_{u \in D[0,1]} \mathbb{E}\Big[\int_0^t u_s \mathrm{d}s + \frac{1}{2}\int_0^1 \nu(s)\big(u_s^2 - 1\big)\mathrm{d}s\Big]$$

$$\nu(t) = \int_t^1 \gamma(s)\mathrm{d}s \qquad \text{s.t.} \quad \int_0^1 u_s \mathrm{d}B_s \in (-1,1) \text{ a.s.}$$

# Relaxed stochastic control problem

$$\mathscr{E}_\star \leq \mathsf{REL}$$

$$\mathsf{REL} \equiv \sup_{u \in D[0,1]} \mathbb{E}\Big[\int_0^t u_s \mathrm{d}s + \frac{1}{2}\int_0^1 \nu(s)\big(u_s^2 - 1\big)\mathrm{d}s\Big]$$

$$\nu(t) = \int_t^1 \gamma(s)\mathrm{d}s \qquad \text{s.t.} \quad \int_0^1 u_s \mathrm{d}B_s \in (-1,1) \text{ a.s.}$$

**Theorem:** $\mathsf{REL} = \mathsf{P}(\gamma)$ and $\mathscr{E}_\star = \inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$ !!

# Relaxed stochastic control problem

$$\mathscr{E}_\star \leq \mathsf{REL}$$

$$\mathsf{REL} \equiv \sup_{u\in D[0,1]} \mathbb{E}\Big[\int_0^t u_s \mathrm{d}s + \frac{1}{2}\int_0^1 \nu(s)\big(u_s^2-1\big)\mathrm{d}s\Big]$$

$$\nu(t) = \int_t^1 \gamma(s)\mathrm{d}s \qquad \text{s.t.} \quad \int_0^1 u_s \mathrm{d}B_s \in (-1,1) \text{ a.s.}$$

**Theorem:** $\mathsf{REL} = \mathsf{P}(\gamma)$ and $\mathscr{E}_\star = \inf_{\gamma\in\mathscr{L}} \mathsf{P}(\gamma)$ !!

Optimal control: $u_t^* = \partial_x^2 \Phi_\gamma(t, X_t)$

$$v(t,x) = \gamma(t)\partial_x \Phi_\gamma(t,x)$$

$$\mathrm{d}X_t = \gamma(t)\partial_x\Phi_\gamma(t,X_t)\mathrm{d}t + \mathrm{d}B_t$$

# Main message

## Duality

Stochastic control problem

$$\sup_{u \in D[0,1]} \mathbb{E}\Big[\int_0^1 u_s \mathrm{d}s\Big]$$
$$\text{s.t.} \quad \mathbb{E}[u_t^2] = 1 \qquad \forall t \in [0,1],$$
$$\int_0^1 u_s \mathrm{d}B_s \in (-1,1) \text{ a.s.}$$

$\longleftrightarrow$

Parisi formula

$$\inf_{\gamma \in \mathscr{L}} \mathsf{P}(\gamma)$$

# Some interesting directions

- Physical interpretation of Extended Parisi Principle
- Explore the duality property.
- Other models: Perceptron, k-SAT, bipartite, Hopfield,…
- Sampling from the Gibbs measure
- Sparse models (MIN-BIS, MAX-CUT on random regular graph)

Thanks!