The Threshold Energy of Low Temperature Langevin Dynamics for Pure Spherical Spin Glasses

Mark Sellke
Department of Statistics
Harvard

Statistical Physics & Machine Learning Back Together Again
Cargèse, August 2023

arxiv:2305.07956

Plan for this Talk

1. Introduction and background
    - Spherical spin glasses and Langevin dynamics
    - Cugliandolo–Kurchan equations
    - Bounding flows
    - The threshold E_∞
2. Main result: threshold energy of low temperature dynamics
    - Upper bound: Lipschitz approximation and BOGP
    - Lower bound: climbing near saddles
3. Epilogue

Definition of Pure Spherical Spin Glasses

Pure p-spin Hamiltonian: random function $H_N : \mathbb{R}^N \to \mathbb{R}$ given by

$$H_N(\sigma) = N^{-(p-1)/2} \sum_{1 \le i_1, i_2, \dots, i_p \le N} J_{i_1,\dots,i_p} \sigma_{i_1} \dots \sigma_{i_p}$$

with i.i.d. Gaussian coefficients $J_{i_1,\dots,i_p} \sim \mathcal{N}(0,1)$.

Inputs σ will be on the sphere: $\mathcal{S}_N = \{\sigma \in \mathbb{R}^N \ : \ \sum_{i=1}^N \sigma_i^2 = N\}$.

Definition of Pure Spherical Spin Glasses

Pure p-spin Hamiltonian: random function $H_N : \mathbb{R}^N \to \mathbb{R}$ given by

$$H_N(\sigma) = N^{-(p-1)/2} \sum_{1 \le i_1, i_2, \dots, i_p \le N} J_{i_1,\dots,i_p} \sigma_{i_1} \dots \sigma_{i_p}$$

with i.i.d. Gaussian coefficients $J_{i_1,\dots,i_p} \sim \mathcal{N}(0,1)$.

Inputs σ will be on the sphere: $\mathcal{S}_N = \{\sigma \in \mathbb{R}^N \ : \ \sum_{i=1}^N \sigma_i^2 = N\}$.

Quick facts:

1. Rotationally invariant Gaussian process: $\mathbb{E} H_N(\sigma) H_N(\rho) = N \left(\frac{\langle \sigma, \rho \rangle}{N}\right)^p$.
2. Scaling: $\max_{\sigma \in \mathcal{S}_N} \|H_N\| \asymp N$, $\quad \|\nabla H_N(\sigma)\| \asymp \sqrt{N}$, $\quad \|\nabla^2 H_N(\sigma)\|_{\mathrm{op}} \asymp 1$.

Spherical Langevin Dynamics

Langevin dynamics on $\mathcal{S}_N$:

$$\mathrm{d}\boldsymbol{x}_t = \left(\beta \nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t) - \frac{(N-1)\boldsymbol{x}_t}{2N}\right) \mathrm{d}t + P^{\perp}_{\boldsymbol{x}_t}\, \mathrm{d}\boldsymbol{B}_t.$$

Invariant for Gibbs measure $\mu_\beta(\mathrm{d}\sigma) = e^{\beta H_N(\sigma)} \mathrm{d}\sigma / Z_N(\beta)$. Much is known about μ_β even at low temperature:

- Free energy is 1-RSB [Talagrand 06]
- Geometric description: orthogonal deep wells, extremes are Poisson-Dirichlet [Subag 17].

Langevin dynamics on $\mathcal{S}_N$:

$$\mathrm{d}\boldsymbol{x}_t = \left(\beta \nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t) - \frac{(N-1)\boldsymbol{x}_t}{2N}\right) \mathrm{d}t + P^{\perp}_{\boldsymbol{x}_t}\, \mathrm{d}\boldsymbol{B}_t.$$

Invariant for Gibbs measure $\mu_\beta(\mathrm{d}\sigma) = e^{\beta H_N(\sigma)}\mathrm{d}\sigma / Z_N(\beta)$. Much is known about μ_β even at low temperature:

- Free energy is 1-RSB [Talagrand 06]
- Geometric description: orthogonal deep wells, extremes are Poisson-Dirichlet [Subag 17].

$t_{mix}(\beta) \geq e^{\Omega(N)}$ for large β, so μ_β will not be realistically accessed [Ben Arous-Jagannath 18].

Study of $O(1)$-time dynamics since [Sompolinsky-Zippelius 82] (SK model).

Physics Predictions and Rigorous Results

1. Exact description via Cugliandolo-Kurchan equations [Crisanti-Horner-Sommers 93].
   - [Ben Arous-Dembo-Guionnet 06]: **Yes** (for soft spherical spins)

2. Fluctuation-dissipation relation & exponential decay of correlations at high temperature.
   - [Dembo-Guionnet-Mazza 07]: **Yes**, using 1.

Physics Predictions and Rigorous Results

1. Exact description via Cugliandolo-Kurchan equations [Crisanti-Horner-Sommers 93].
   - [Ben Arous-Dembo-Guionnet 06]: **Yes** (for soft spherical spins)
2. Fluctuation-dissipation relation & exponential decay of correlations at high temperature.
   - [Dembo-Guionnet-Mazza 07]: **Yes**, using 1.
3. Aging at low temperatures [Cugliandolo-Kurchan 93].
   - [Ben Arous-Dembo-Guionnet 01]: **Yes**, for $\underline{p = 2}$.

Physics Predictions and Rigorous Results

1. Exact description via Cugliandolo-Kurchan equations [Crisanti-Horner-Sommers 93].
   - [Ben Arous-Dembo-Guionnet 06]: **Yes** (for soft spherical spins)

2. Fluctuation-dissipation relation & exponential decay of correlations at high temperature.
   - [Dembo-Guionnet-Mazza 07]: **Yes**, using 1.

3. Aging at low temperatures [Cugliandolo-Kurchan 93].
   - [Ben Arous-Dembo-Guionnet 01]: **Yes**, for $\underline{p = 2}$.

4. Large time threshold energy $E_\infty(p) \equiv 2\sqrt{\frac{p-1}{p}}$ as $\beta \to \infty$ [Biroli 99].
   - [Ben Arous-Gheissari-Jagannath 18]: **Explicit bounds** via differential inequalities.

Cugliandolo-Kurchan Equations

Closed system of equations as $N \to \infty$ for:

$$C(s,t) \equiv \langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle / N,$$
$$R(s,t) \equiv \langle \boldsymbol{x}_s, \boldsymbol{B}_t \rangle / N.$$

Cugliandolo-Kurchan Equations

Closed system of equations as $N \to \infty$ for:

$$C(s,t) \equiv \langle \boldsymbol{x}_s, \boldsymbol{x}_t \rangle / N,$$
$$R(s,t) \equiv \langle \boldsymbol{x}_s, \boldsymbol{B}_t \rangle / N.$$

Tells you everything in principle, but hard to work with. For $s \geq t \geq 0$:

$$\partial_s R(s,t) = -\mu(s) R(s,t) + \beta^2 p(p-1) \int_t^s R(u,t) R(s,u) C(s,u)^{p-2} \, \mathrm{d}u,$$

$$\partial_s C(s,t) = -\mu(s) C(s,t) + \beta^2 p(p-1) \int_0^s C(u,t) R(s,u) C(s,u)^{p-2} \, \mathrm{d}u$$
$$+ \beta^2 p \int_0^t C(s,u)^{p-1} R(t,u) \, \mathrm{d}u;$$

$$\mu(s) \equiv \frac{1}{2} + \beta^2 p^2 \int_0^s C(s,u)^{p-1} R(s,u) \, \mathrm{d}u.$$

Bounding Flows Approach

Rigorously understanding the Cugliandolo-Kurchan equations is difficult at low temperature.

[Ben Arous-Gheissari-Jagannath 18]: **bounding flows** method of differential inequalities.

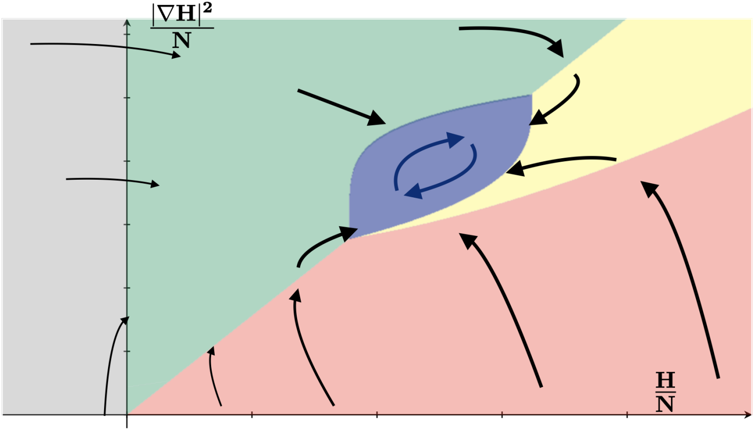
- Shows $\mathrm{d}(H_N(\boldsymbol{x}_t), \|\nabla H_N(\boldsymbol{x}_t)\|^2) \in \Gamma(H_N(\boldsymbol{x}_t), \|\nabla H_N(\boldsymbol{x}_t)\|^2) \subseteq \mathbb{R}^2$.
- Quantitative lower bounds on $H_N(\boldsymbol{x}_T)$, even for disorder dependent $\boldsymbol{x}_0 \in \mathcal{S}_N$.

Bounding Flows Approach

Rigorously understanding the Cugliandolo-Kurchan equations is difficult at low temperature.

[Ben Arous-Gheissari-Jagannath 18]: **bounding flows** method of differential inequalities.

- Shows $\mathrm{d}(H_N(\boldsymbol{x}_t), \|\nabla H_N(\boldsymbol{x}_t)\|^2) \in \Gamma(H_N(\boldsymbol{x}_t), \|\nabla H_N(\boldsymbol{x}_t)\|^2) \subseteq \mathbb{R}^2$.
- Quantitative lower bounds on $H_N(\boldsymbol{x}_T)$, even for disorder dependent $\boldsymbol{x}_0 \in \mathcal{S}_N$.



Intuition for E_∞

[Biroli 99]: $E_\infty = 2\sqrt{\frac{p-1}{p}}$ should be the threshold energy in the limit $\beta \to \infty$.

The Hessian's maximum eigenvalue "equals zero at zero temperature, as is expected for a dynamics in a rugged energy landscape".

Intuition for E_∞

[Biroli 99]: $E_\infty = 2\sqrt{\frac{p-1}{p}}$ should be the threshold energy in the limit $\beta \to \infty$.

The Hessian's maximum eigenvalue "equals zero at zero temperature, as is expected for a dynamics in a rugged energy landscape".

Explanation:

- For $\boldsymbol{x} \in \mathcal{S}_N$, the spherical Hessian $\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{x})$ is a **shifted GOE**:

$$\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{x}) \stackrel{d}{=} \sqrt{p(p-1)}\, GOE(N-1) - p \cdot \frac{H_N(\boldsymbol{x})}{N}.$$

- The prediction above says $\lambda_{\max}(\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{x}_T)) \approx 0$.
- Since $\lambda_{\max}(GOE(N-1)) \approx 2$, we should also predict energy E_∞, i.e.

$$\lim_{\beta, T \to \infty} \operatorname*{p-lim}_{N \to \infty} H_N(\boldsymbol{x}_T)/N = E_\infty.$$

New Results: E_∞ is the Threshold Energy as $\beta \to \infty$

Theorem (S 23, Upper Bound)

For any β there is $\delta > 0$ such that for any T, if $\boldsymbol{x}_0 \in \mathcal{S}_N$ is independent of H_N:

$$\mathbb{P}\left[\sup_{t \in [0,T]} H_N(\boldsymbol{x}_t)/N \leq E_\infty - \delta\right] \geq 1 - e^{-cN}.$$

New Results: E_∞ is the Threshold Energy as $\beta \to \infty$

Theorem (S 23, Upper Bound)

For any β there is $\delta > 0$ such that for any T, if $\mathbf{x}_0 \in \mathcal{S}_N$ is independent of H_N:

$$\mathbb{P}\left[\sup_{t\in[0,T]} H_N(\mathbf{x}_t)/N \le E_\infty - \delta\right] \ge 1 - e^{-cN}.$$

Theorem (S 23, Lower Bound)

For any $\eta > 0$, with $T_0 = T_0(\eta)$ and $\beta \ge \beta_0(\eta)$, ***even if $\mathbf{x}_0$ is disorder dependent****:*

$$\mathbb{P}\left[\inf_{t\in[T_0,T_0+e^{cN}]} H_N(\mathbf{x}_t)/N \ge E_\infty - \eta\right] \ge 1 - e^{-cN}.$$

New Results: E_∞ is the Threshold Energy as $\beta \to \infty$

Theorem (S 23, Upper Bound)

For any β there is $\delta > 0$ such that for any T, if $\mathbf{x}_0 \in \mathcal{S}_N$ is independent of H_N:

$$\mathbb{P}\left[\sup_{t \in [0,T]} H_N(\mathbf{x}_t)/N \le E_\infty - \delta\right] \ge 1 - e^{-cN}.$$

Theorem (S 23, Lower Bound)

*For any $\eta > 0$, with $T_0 = T_0(\eta)$ and $\beta \ge \beta_0(\eta)$, **even if $\mathbf{x}_0$ is disorder dependent**:*

$$\mathbb{P}\left[\inf_{t \in [T_0, T_0 + e^{cN}]} H_N(\mathbf{x}_t)/N \ge E_\infty - \eta\right] \ge 1 - e^{-cN}.$$

For large constant times $t \in [T_0, T]$ and large β, the energy stays uniformly just below E_∞:

$$H_N(\mathbf{x}_t)/N \in [E_\infty - \eta, E_\infty - \delta].$$

Some Comments

Once energy settles, the gradient stays small:

$$\lim_{N\to\infty} \mathbb{P}\left[\sup_{t\in[T_0,T]} \|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\|/\sqrt{N} \le \delta \right] = 1, \quad \forall\, \beta \ge \beta_0(\delta),\ \ T \ge T_0(\delta).$$

Some Comments

Once energy settles, the gradient stays small:

$$\lim_{N\to\infty} \mathbb{P}\left[\sup_{t\in[T_0,T]} \|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\|/\sqrt{N} \le \delta\right] = 1, \quad \forall\ \beta \ge \beta_0(\delta),\ \ T \ge T_0(\delta).$$

Mixed p-spin models with covariance $\xi(t) = \sum_{p\ge 2} \gamma_p^2 t^p$:

- Upper bound: $\mathrm{ALG}(\xi) = \int_0^1 \sqrt{\xi''(t)}\,\mathrm{d}t$.
- Lower bound: E_∞^- from [Auffinger-Ben Arous 2013].

Some Comments

Once energy settles, the gradient stays small:

$$\lim_{N\to\infty} \mathbb{P}\Big[\sup_{t\in[T_0,T]} \|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\|/\sqrt{N} \le \delta\Big] = 1, \quad \forall\, \beta \ge \beta_0(\delta),\ T \ge T_0(\delta).$$

Mixed p-spin models with covariance $\xi(t) = \sum_{p\ge 2} \gamma_p^2 t^p$:

- Upper bound: $\mathrm{ALG}(\xi) = \int_0^1 \sqrt{\xi''(t)}\,\mathrm{d}t$.
- Lower bound: E_∞^- from [Auffinger-Ben Arous 2013].
- In mixed models, $\lambda_{\max} \approx 0$ can hold in a range of energies, e.g. $[E_\infty^-, E_\infty^+]$.
- For **pure multi-species** spherical spin glasses, bounds continue to match at E_∞.

Some Comments

Once energy settles, the gradient stays small:

$$\lim_{N\to\infty} \mathbb{P}\left[\sup_{t\in[T_0,T]} \|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\|/\sqrt{N} \le \delta \right] = 1, \quad \forall\, \beta \ge \beta_0(\delta),\ T \ge T_0(\delta).$$

Mixed p-spin models with covariance $\xi(t) = \sum_{p\ge 2} \gamma_p^2 t^p$:

- Upper bound: $\mathrm{ALG}(\xi) = \int_0^1 \sqrt{\xi''(t)}\,\mathrm{d}t$.
- Lower bound: E_∞^- from [Auffinger-Ben Arous 2013].
- In mixed models, $\lambda_{\max} \approx 0$ can hold in a range of energies, e.g. $[E_\infty^-, E_\infty^+]$.
- For **pure multi-species** spherical spin glasses, bounds continue to match at E_∞.

Initializing via high-temperature dynamics changes neither bound.

- For pure models, threshold equals E_∞ regardless of early dynamics.
- [Folena–Franz–Ricci-Tersenghi 21]: this <u>does</u> change the eventual energy for mixed models.

Upper Bound via Hardness for Lipschitz Algorithms

Upper bound uses hardness for Lipschitz optimization algorithms (Brice's talk yesterday).

Definition

An L-**Lipschitz algorithm** is an $\mathcal{A}_N : \mathbb{R}^{N^p} \times \Omega \to \mathcal{S}_N$ which is L-Lipschitz in 1st coordinate.

Upper bound uses hardness for Lipschitz optimization algorithms (Brice's talk yesterday).

Definition

An L-**Lipschitz algorithm** is an $\mathcal{A}_N : \mathbb{R}^{N^p} \times \Omega \to \mathcal{S}_N$ which is L-Lipschitz in 1st coordinate.

Theorem (Huang-S 21 & 23)

Fix any $L, \eta > 0$. If $\mathcal{A}_N$ is an L-Lipschitz algorithm, then for N large enough,

$$\mathbb{P}[H_N(\mathcal{A}_N(H_N))/N \leq E_\infty + \eta] \geq 1 - e^{-cN}.$$

(Informally: Lipschitz algorithms cannot access energies above $E_\infty + o_N(1)$.)

Upper bound uses hardness for Lipschitz optimization algorithms (Brice's talk yesterday).

Definition

An L-**Lipschitz algorithm** is an $\mathcal{A}_N : \mathbb{R}^{N^p} \times \Omega \to \mathcal{S}_N$ which is L-Lipschitz in 1st coordinate.

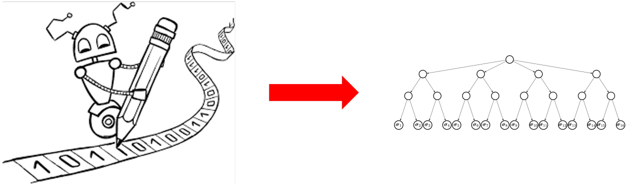
Theorem (Huang-S 21 & 23)

Fix any $L, \eta > 0$. If $\mathcal{A}_N$ is an L-Lipschitz algorithm, then for N large enough,

$$\mathbb{P}[H_N(\mathcal{A}_N(H_N))/N \leq E_\infty + \eta] \geq 1 - e^{-cN}.$$

(Informally: Lipschitz algorithms cannot access energies above $E_\infty + o_N(1)$.)

Proof: branching overlap gap property.
Run $\mathcal{A}_N$ on correlated copies of H_N.
Extends OGP from [Gamarnik-Sudan 14,...].



Upper Bound via Hardness for Lipschitz Algorithms

Remains to approximate $\boldsymbol{x}_T$ by an $L(\beta, T)$-Lipschitz function of $(J_{i_1,\dots,i_p})_{i_k=1}^N$ for each $\boldsymbol{B}_{[0,T]}$.

Previously known for soft spherical Langevin dynamics [Ben Arous-Dembo-Guionnet 06].
We approximate the hard dynamics pathwise by soft dynamics, which suffices.

Upper Bound via Hardness for Lipschitz Algorithms

Remains to approximate $\boldsymbol{x}_T$ by an $L(\beta, T)$-Lipschitz function of $(J_{i_1,\dots,i_p})_{i_k=1}^N$ for each $\boldsymbol{B}_{[0,T]}$.

Previously known for soft spherical Langevin dynamics [Ben Arous-Dembo-Guionnet 06].
We approximate the hard dynamics pathwise by soft dynamics, which suffices.

Combined with [Ben Arous-Dembo-Guionnet 06, Dembo-Guionnet-Mazza 07], a byproduct:

Corollary (S 23)

The hard spherical dynamics are described by Cugliandolo-Kurchan equations. They are given by a limit of those for the soft dynamics.

Upper Bound via Hardness for Lipschitz Algorithms

Remains to approximate $\boldsymbol{x}_T$ by an $L(\beta, T)$-Lipschitz function of $(J_{i_1,\dots,i_p})_{i_k=1}^N$ for each $\boldsymbol{B}_{[0,T]}$.

Previously known for soft spherical Langevin dynamics [Ben Arous-Dembo-Guionnet 06].
We approximate the hard dynamics pathwise by soft dynamics, which suffices.

Combined with [Ben Arous-Dembo-Guionnet 06, Dembo-Guionnet-Mazza 07], a byproduct:

Corollary (S 23)

The hard spherical dynamics are described by Cugliandolo-Kurchan equations. They are given by a limit of those for the soft dynamics.

Improving the upper bound from $E_\infty + o_N(1)$ to $E_\infty - o_\beta(1)$:

- [Ben Arous-Gheissari-Jagannath 18]: $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\| \geq \delta_1(\beta)\sqrt{N}$ for all times t.
- Hence a final noise-less gradient step slightly improves the energy.
- This modified algorithm is just as Lipschitz as before.

Lower Bound: Reaching Approximate Local Maxima

Definition

$x \in \mathcal{S}_N$ is an ε-**approximate local maximum** if both:

1. $\|\nabla_{\mathrm{sp}} H_N(x)\| \le \varepsilon\sqrt{N}$.
2. $\lambda_{\varepsilon N}\big(\nabla^2_{\mathrm{sp}} H_N(x)\big) \le \varepsilon$.

If 1 holds but 2 doesn't, then x is an ε-**approximate saddle**.

Lower Bound: Reaching Approximate Local Maxima

Definition

$x \in \mathcal{S}_N$ is an **ε-approximate local maximum** if both:

1. $\|\nabla_{\mathrm{sp}} H_N(x)\| \le \varepsilon\sqrt{N}$.
2. $\lambda_{\varepsilon N}(\nabla^2_{\mathrm{sp}} H_N(x)) \le \varepsilon$.

If 1 holds but 2 doesn't, then x is an **ε-approximate saddle**.

Proposition (Specific to Pure p-Spin Models)

With probability $1 - e^{-cN}$, all ε-approximate local maxima satisfy $H_N(x)/N \ge E_\infty - o_\varepsilon(1)$.

Lower Bound: Reaching Approximate Local Maxima

Definition

$x \in \mathcal{S}_N$ is an **ε-approximate local maximum** if both:

1. $\|\nabla_{\mathrm{sp}} H_N(x)\| \le \varepsilon\sqrt{N}$.
2. $\lambda_{\varepsilon N}\big(\nabla^2_{\mathrm{sp}} H_N(x)\big) \le \varepsilon$.

If 1 holds but 2 doesn't, then x is an **ε-approximate saddle**.

Proposition (Specific to Pure p-Spin Models)

With probability $1-e^{-cN}$, all ε-approximate local maxima satisfy $H_N(x)/N \ge E_\infty - o_\varepsilon(1)$.

Theorem (Only Uses 3rd-Order Smoothness of H_N; cf [ZLC 17, JNGKJ 21])

Suppose all ε-approximate local maxima satisfy $H_N(x)/N \ge E_*(\varepsilon)$.
Then for large T_0, β depending on ε, and disorder-dependent $x_0 \in \mathcal{S}_N$:

$$\mathbb{P}\left[\inf_{t\in[T_0,T_0+e^{cN}]} H_N(x_t)/N \ge E_*(\varepsilon) - o_\varepsilon(1)\right] \ge 1-e^{-cN}.$$

Energy Gain While Below $E_*(\varepsilon)$

We directly show $H_N(\boldsymbol{x}_t)$ increases while $H_N(\boldsymbol{x}_t)/N \le E_*(\varepsilon)$. This is formalized with a closely spaced sequence of stopping times

$$0 = \tau_0 < \tau_1 < \cdots < \tau_M \approx T$$

Energy Gain While Below $E_*(\varepsilon)$

We directly show $H_N(\boldsymbol{x}_t)$ increases while $H_N(\boldsymbol{x}_t)/N \leq E_*(\varepsilon)$. This is formalized with a closely spaced sequence of stopping times

$$0 = \tau_0 < \tau_1 < \cdots < \tau_M \approx T$$

Definition of $E_*(\varepsilon)$ leads to three cases:

1. Large energy: $H_N(\boldsymbol{x}_\tau)/N \geq E_*(\varepsilon)$.
2. Large gradient: $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \geq C\beta^{-1/2}\sqrt{N}$.
3. Approximate saddle: $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \leq C\beta^{-1/2}\sqrt{N}$ **and** $\lambda_{\varepsilon N}\left(\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\right) \geq \varepsilon$.

Energy Gain While Below $E_*(\varepsilon)$

We directly show $H_N(\boldsymbol{x}_t)$ increases while $H_N(\boldsymbol{x}_t)/N \le E_*(\varepsilon)$. This is formalized with a closely spaced sequence of stopping times

$$0 = \tau_0 < \tau_1 < \cdots < \tau_M \approx T$$

Definition of $E_*(\varepsilon)$ leads to three cases:

1. Large energy: $H_N(\boldsymbol{x}_\tau)/N \ge E_*(\varepsilon)$.
2. Large gradient: $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \ge C\beta^{-1/2}\sqrt{N}$.
3. Approximate saddle: $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \le C\beta^{-1/2}\sqrt{N}$ **and** $\lambda_{\varepsilon N}\big(\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\big) \ge \varepsilon$.

If $\boldsymbol{x}_\tau$ is in Case 1, simply stop once the energy drops below $E_*(\varepsilon)$.

In Cases 2, 3, we will show $H_N(\boldsymbol{x}_t)$ increases.

Energy Gain From Large Gradient

Lemma (Large Gradient Case)

If $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \geq C\beta^{-1/2}\sqrt{N}$ *then with probability* $1 - e^{-cN}$*:*

$$H_N(\boldsymbol{x}_{\tau+\beta^{-10}}) - H_N(\boldsymbol{x}_\tau) \geq \beta^{-10} N.$$

Energy Gain From Large Gradient

Lemma (Large Gradient Case)

If $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \geq C\beta^{-1/2}\sqrt{N}$ *then with probability* $1 - e^{-cN}$*:*

$$H_N(\boldsymbol{x}_{\tau+\beta^{-10}}) - H_N(\boldsymbol{x}_\tau) \geq \beta^{-10} N.$$

Proof: large gradient overwhelms the Itô term.

$$\mathrm{d}H_N(\boldsymbol{x}_t) = \Big(\underbrace{\beta\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\|^2 \pm O(N)\Big)}_{\geq CN \text{ on } \tau\leq t\leq\tau+\beta^{-10}} \mathrm{d}t + \underbrace{\beta\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\|}_{O(\sqrt{N})} \mathrm{d}B_t.$$

Energy Gain From Large Gradient

Lemma (Large Gradient Case)

If $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \geq C\beta^{-1/2}\sqrt{N}$ *then with probability* $1-e^{-cN}$*:*

$$H_N(\boldsymbol{x}_{\tau+\beta^{-10}}) - H_N(\boldsymbol{x}_\tau) \geq \beta^{-10} N.$$

Proof: large gradient overwhelms the Itô term.

$$\mathrm{d}H_N(\boldsymbol{x}_t) = \underbrace{\Big(\beta\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\|^2 \pm O(N)\Big)}_{\geq CN \text{ on } \tau \leq t \leq \tau+\beta^{-10}} \mathrm{d}t + \underbrace{\beta\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_t)\|}_{O(\sqrt{N})} \mathrm{d}B_t.$$

Noise level is so small that differential inequalities remain true with probability $1-e^{-cN}$.
(Similarly, bound growth of $\|\boldsymbol{x}_t - \boldsymbol{x}_\tau\|^2$ with another differential inequality.)

Remains to show the following (with $\beta \gg \overline{C}(\varepsilon) \gg 1/\varepsilon \gg C \asymp 1$).

Lemma

If $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \leq C\beta^{-1/2}\sqrt{N}$ ***and*** $\lambda_{\varepsilon N}\big(\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\big) \geq \varepsilon$*:*

$$H_N(\boldsymbol{x}_{\tau+\overline{C}(\varepsilon)\beta^{-1}}) - H_N(\boldsymbol{x}_\tau) \geq \beta^{-1} N.$$

Remains to show the following (with $\beta \gg \overline{C}(\varepsilon) \gg 1/\varepsilon \gg C \asymp 1$).

Lemma

If $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \leq C\beta^{-1/2}\sqrt{N}$ ***and*** $\lambda_{\varepsilon N}\left(\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\right) \geq \varepsilon$*:*

$$H_N(\boldsymbol{x}_{\tau+\overline{C}(\varepsilon)\beta^{-1}}) - H_N(\boldsymbol{x}_\tau) \geq \beta^{-1} N.$$

Wishful thinking: imagine H_N is quadratic and flatten the domain $\mathcal{S}_N$ to $\mathbb{R}^{N-1}$.

Gaining Energy Near Approximate Saddles

Remains to show the following (with $\beta \gg \overline{C}(\varepsilon) \gg 1/\varepsilon \gg C \asymp 1$).

Lemma

If $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \le C\beta^{-1/2}\sqrt{N}$ *and* $\lambda_{\varepsilon N}\left(\nabla^2_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\right) \ge \varepsilon$*:*

$$H_N(\boldsymbol{x}_{\tau+\overline{C}(\varepsilon)\beta^{-1}}) - H_N(\boldsymbol{x}_\tau) \ge \beta^{-1} N.$$

Wishful thinking: imagine H_N is quadratic and flatten the domain $\mathcal{S}_N$ to $\mathbb{R}^{N-1}$.

Then $\boldsymbol{x}_{\tau+t}$ would be a multi-dimensional OU process. Easy to analyze!

- Positive eigendirections: exponentially fast energy gain.
- Negative eigendirections: trapped or diffusive movement.
- Overall energy gain of $\Omega(N\beta^{-1})$ after time $\overline{C}(\varepsilon)\beta^{-1}$.
- (But, energy can initially drop. This is a problem for differential inequalities.)

Ornstein–Uhlenbeck Approximation via Taylor Expansion

In general: map $\mathcal{S}_N$ to $\mathbb{R}^{N-1}$ and **Taylor expand the SDE coefficients** near x_τ.

- A suitable approximation exactly yields a multi-dimensional OU process.
- Suffices to carefully estimate the approximation error.

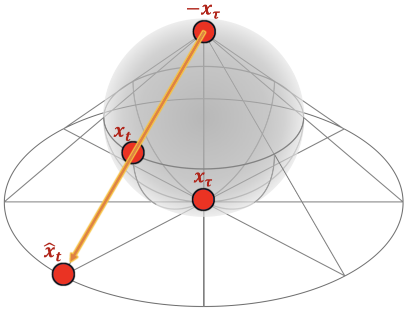
Ornstein–Uhlenbeck Approximation via Taylor Expansion

In general: map $\mathcal{S}_N$ to $\mathbb{R}^{N-1}$ and **Taylor expand the SDE coefficients** near x_τ.

- A suitable approximation exactly yields a multi-dimensional OU process.
- Suffices to carefully estimate the approximation error.

Use stereographic projection map $\Gamma_{\boldsymbol{x}_\tau}$ centered at $-\boldsymbol{x}_\tau$:

$$
\begin{aligned}
&\Gamma_{\boldsymbol{x}_\tau} : \mathcal{S}_N \backslash \{-\boldsymbol{x}_\tau\} \to \mathbb{R}^{N-1}, \\
&\Gamma_{\boldsymbol{x}_\tau}(\boldsymbol{x}_\tau) = \vec{0}, \\
&\Gamma_{\boldsymbol{x}_\tau}(\boldsymbol{x}_t) = \widehat{\boldsymbol{x}}_t, \\
&\widehat{H}_N(\widehat{\boldsymbol{x}}_t) = H_N(\Gamma_{\boldsymbol{x}_\tau}^{-1}(\widehat{\boldsymbol{x}}_t)) = H_N(\boldsymbol{x}_t)
\end{aligned}
$$



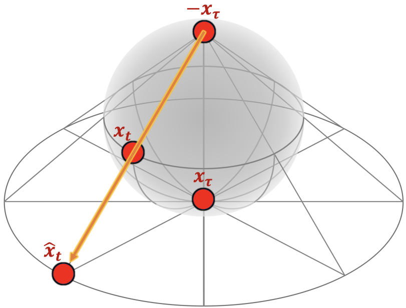
Ornstein–Uhlenbeck Approximation via Taylor Expansion

In general: map $\mathcal{S}_N$ to $\mathbb{R}^{N-1}$ and **Taylor expand the SDE coefficients** near x_τ.

- A suitable approximation exactly yields a multi-dimensional OU process.
- Suffices to carefully estimate the approximation error.

Use stereographic projection map $\Gamma_{\boldsymbol{x}_\tau}$ centered at $-\boldsymbol{x}_\tau$:

$$\Gamma_{\boldsymbol{x}_\tau} : \mathcal{S}_N \backslash \{-\boldsymbol{x}_\tau\} \to \mathbb{R}^{N-1},$$
$$\Gamma_{\boldsymbol{x}_\tau}(\boldsymbol{x}_\tau) = \vec{0},$$
$$\Gamma_{\boldsymbol{x}_\tau}(\boldsymbol{x}_t) = \widehat{\boldsymbol{x}}_t,$$
$$\widehat{H}_N(\widehat{\boldsymbol{x}}_t) = H_N(\Gamma^{-1}_{\boldsymbol{x}_\tau}(\widehat{\boldsymbol{x}}_t)) = H_N(\boldsymbol{x}_t)$$



Projected dynamics in $\mathbb{R}^{N-1}$ and quadratic approximation:

$$\mathrm{d}\widehat{\boldsymbol{x}}_t = \vec{b}_t(\widehat{\boldsymbol{x}}_t)\,\mathrm{d}t + \sigma_t \mathrm{d}\boldsymbol{W}_t,$$
$$\mathrm{d}\boldsymbol{x}_t^{(Q)} = \beta \nabla H_N^{(Q)}(\boldsymbol{x}_t^{(Q)})\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t.$$

Required Estimates for Ornstein–Uhlenbeck Approximation

We show $\boldsymbol{x}_t^{(Q)} \approx \widehat{\boldsymbol{x}}_t$ via more scalar approximate differential inequalities.

- Movement is small on $O(1/\beta)$ time-scales since $\|\nabla_{\mathrm{sp}} H_N(\boldsymbol{x}_\tau)\| \leq C\beta^{-1/2}\sqrt{N}$:

$$\|\widehat{\boldsymbol{x}}_{\tau+\overline{C}\beta^{-1}} - \widehat{\boldsymbol{x}}_t\| \leq O_{\overline{C}}(\beta^{-1/2}\sqrt{N}), \tag{1}$$

$$\implies \|\nabla \widehat{H}_N(\widehat{\boldsymbol{x}}_{\tau+\overline{C}\beta^{-1}})\| \leq O_{\overline{C}}(\beta^{-1/2}\sqrt{N}).$$

- Since $H_N^{(Q)}$ is a 2nd order Taylor approximation for $\widehat{H}_N$, (1) gives:

$$\left|H_N^{(Q)}(\boldsymbol{x}^{(Q)}_{\tau+\overline{C}\beta^{-1}}) - \widehat{H}_N(\boldsymbol{x}^{(Q)}_{\tau+\overline{C}\beta^{-1}})\right| \leq O_{\overline{C}}(\beta^{-3/2}N). \tag{2}$$

- Same-time approximation $\boldsymbol{x}_t^{(Q)} \approx \widehat{\boldsymbol{x}}_t$ turns out to be better since $\mathrm{d}B_t$ cancels:

$$\|\boldsymbol{x}^{(Q)}_{\tau+\overline{C}\beta^{-1}} - \widehat{\boldsymbol{x}}_{\tau+\overline{C}\beta^{-1}}\| \leq O_{\overline{C}}(\beta^{-1}\sqrt{N}).$$

- Combining the previous two,

$$\left|\widehat{H}_N(\widehat{\boldsymbol{x}}_{\tau+\overline{C}\beta^{-1}}) - \widehat{H}_N(\boldsymbol{x}^{(Q)}_{\tau+\overline{C}\beta^{-1}})\right| \leq O_{\overline{C}}(\beta^{-3/2}N). \tag{3}$$

- Energy gain of $H_N^{(Q)}(\boldsymbol{x}_t^{(Q)})$ is $\Omega(\beta^{-1}N)$ by explicit OU computation. Combining with (2), (3):

$$H_N(\boldsymbol{x}_{\tau+\overline{C}\beta^{-1}}) - H_N(\boldsymbol{x}_\tau) = \widehat{H}_N(\widehat{\boldsymbol{x}}_{\tau+\overline{C}\beta^{-1}}) - \widehat{H}_N(\widehat{\boldsymbol{x}}_\tau) \geq \Omega(\beta^{-1}N).$$

Conclusion

Pure p-spin Hamiltonian:

$$H_N(\sigma) = N^{-(p-1)/2} \sum_{1 \le i_1, i_2, \ldots, i_p \le N} J_{i_1,\ldots,i_p} \sigma_{i_1} \ldots \sigma_{i_p}$$

Main result: for spherical Langevin dynamics:

$$\lim_{T,\beta \to \infty} \operatorname*{p-lim}_{N \to \infty} H_N(x_T)/N = E_\infty(p) \equiv 2\sqrt{\frac{p-1}{p}}.$$

Upper bound holds for Lipschitz algorithms via branching overlap gap property.

Lower bound: dynamics reach approximate local maxima in general smooth landscapes.

- Holds for disorder-dependent $x_0 \in \mathcal{S}_N$, and uniformly in $t \in [T_0, T_0 + e^{cN}]$.

Still Open: monotonicity of asymptotic energy in time for **fixed** β? Existence of a limit?