Tight Algorithmic Thresholds for Optimizing Mean Field Spin Glass Hamiltonians

Mark Sellke (Stanford)

Simons Workshop on Rigorous Evidence for Information-Computation Trade-offs

September 15, 2021

Joint work with Brice Huang (MIT)

Plan for this Talk

1. Problem setup: optimizing a mean field spin glass Hamiltonian.

Plan for this Talk

1. Problem setup: optimizing a mean field spin glass Hamiltonian.
2. Background: Parisi formula, AMP, overlap gap property.

Plan for this Talk

1. Problem setup: optimizing a mean field spin glass Hamiltonian.
2. Background: Parisi formula, AMP, overlap gap property.
3. New result: a tight characterization of the best value achieved by a class of efficient algorithms.

Plan for this Talk

1. Problem setup: optimizing a mean field spin glass Hamiltonian.
2. Background: Parisi formula, AMP, overlap gap property.
3. New result: a tight characterization of the best value achieved by a class of efficient algorithms.
4. Some key ideas: ultrametricity and a branching OGP.

Mean Field Spin Glass Hamiltonians

Definition (Sherrington-Kirkpatrick 75,...)

Fix constants $\gamma_1, \gamma_2, \ldots, \gamma_K \geq 0$. The mixed p-spin Hamiltonian $H_N : \mathbb{R}^N \to \mathbb{R}$ is a random degree K polynomial defined by

$$H_N(\sigma_1, \ldots, \sigma_N) = \sum_{k=1}^{K} N^{-\frac{k-1}{2}} \gamma_k \sum_{1 \leq i_1, \ldots, i_k \leq N} J_{i_1, \ldots, i_k} \sigma_{i_1} \ldots \sigma_{i_k}.$$

Here $J_{i_1, \ldots, i_k}$ are IID standard gaussians.

Mean Field Spin Glass Hamiltonians

Definition (Sherrington-Kirkpatrick 75,...)

Fix constants $\gamma_1, \gamma_2, \ldots, \gamma_K \geq 0$. The mixed p-spin Hamiltonian $H_N : \mathbb{R}^N \to \mathbb{R}$ is a random degree K polynomial defined by

$$H_N(\sigma_1, \ldots, \sigma_N) = \sum_{k=1}^{K} N^{-\frac{k-1}{2}} \gamma_k \sum_{1 \leq i_1, \ldots, i_k \leq N} J_{i_1, \ldots, i_k} \sigma_{i_1} \ldots \sigma_{i_k}.$$

Here $J_{i_1, \ldots, i_k}$ are IID standard gaussians.

- Equivalently: let $\xi(t) = \sum_{k=1}^{K} \gamma_k^2 t^k$. Then H_N is the centered Gaussian process with covariance

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

- Two input sets will be considered:
  - Ising – $\boldsymbol{\sigma} \in \{-1, 1\}^N$
  - Spherical – $||\boldsymbol{\sigma}|| = \sqrt{N}$.

Motivations for the Model

- The Sherrington-Kirkpatrick model ($K = 2$) was introduced to study diluted magnetic alloys such as Copper Manganese.
- Magnetic interaction rapidly oscillates with distance, so use an Ising model with **random weights**: $H_N = \sum_{i,j} J_{i,j}\sigma_i\sigma_j$.

Motivations for the Model

- The Sherrington-Kirkpatrick model ($K = 2$) was introduced to study diluted magnetic alloys such as Copper Manganese.
- Magnetic interaction rapidly oscillates with distance, so use an Ising model with **random weights**: $H_N = \sum_{i,j} J_{i,j}\sigma_i\sigma_j$.
- Mean field: forget the 3-D structure, make all $J_{i,j}$ IID.
- Higher degree interactions and spherical inputs are mathematically natural extensions.

Motivations for the Model

- The Sherrington-Kirkpatrick model ($K = 2$) was introduced to study diluted magnetic alloys such as Copper Manganese.
- Magnetic interaction rapidly oscillates with distance, so use an Ising model with **random weights**: $H_N = \sum_{i,j} J_{i,j}\sigma_i\sigma_j$.
- Mean field: forget the 3-D structure, make all $J_{i,j}$ IID.
- Higher degree interactions and spherical inputs are mathematically natural extensions.
- Also arises as high-degree limit of random MaxCut, MaxSAT (Dembo-Montanari-Sen 17, Panchenko 18).
- Rich source of random non-convex functions, related to some neural network models (Gardner-Derrida 80s, Amit-Gutfreund-Sompolinsky 85, Choromanska-Henaff-Mathieu-Ben Arous-LeCun 15).

The Parisi Formula

Theorem (Parisi 82, Talagrand 06/10, Panchenko 14, Auffinger-Chen 17)

In both the Ising and spherical settings, the limit

$$\mathrm{OPT} \equiv \operatorname*{p-lim}_{N\to\infty} \max_{x} \frac{H_N(x)}{N} = \inf_{\zeta\in\mathcal{U}} \mathcal{P}_\xi(\zeta)$$

holds for explicit Parisi functionals $\mathcal{P}_\xi^{\mathrm{Is}}, \mathcal{P}_\xi^{\mathrm{Sp}}$. Here $\mathcal{U}$ is the set of ***non-decreasing*** *functions $\zeta : [0,1] \to \mathbb{R}^+$.*

The Parisi Formula

Theorem (Parisi 82, Talagrand 06/10, Panchenko 14, Auffinger-Chen 17)

In both the Ising and spherical settings, the limit

$$\mathrm{OPT} \equiv \operatorname*{p-lim}_{N\to\infty} \max_x \frac{H_N(x)}{N} = \inf_{\zeta\in\mathcal{U}} \mathcal{P}_\xi(\zeta)$$

holds for explicit Parisi functionals $\mathcal{P}_\xi^{\mathrm{Is}}, \mathcal{P}_\xi^{\mathrm{Sp}}$. Here $\mathcal{U}$ is the set of ***non-decreasing*** *functions $\zeta : [0,1] \to \mathbb{R}^+$.*

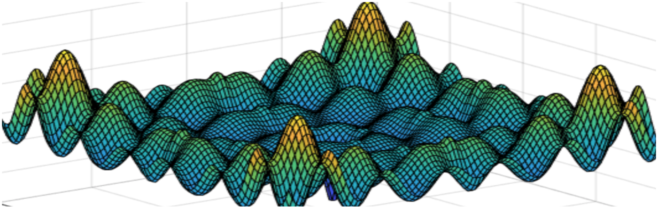
- Question: can efficient algorithms reach $(\mathrm{OPT} - \varepsilon)N$?
- If not, what can be done efficiently?
- Goal: compute $\boldsymbol{\sigma} = \mathcal{A}(H_N)$ with $H_N(\boldsymbol{\sigma})$ as large as possible.

A Look at the Landscape

- If H_N is close to convex, maybe gradient descent works.
- Not the case! On the sphere, H_N can have:

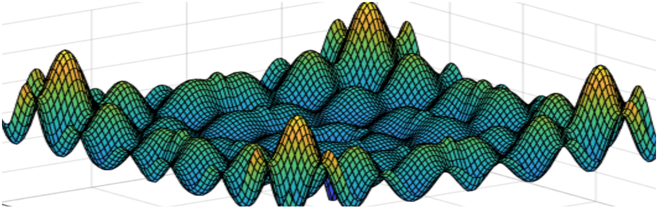
A Look at the Landscape

- If H_N is close to convex, maybe gradient descent works.
- Not the case! On the sphere, H_N can have:
  - Exponentially many near-optimal local maxima. ☺
  - Exponentially *more* suboptimal local maxima. ☹
  - Exponentially *more* suboptimal saddle points. ☹
    (Auffinger-Ben Arous 13, A-BA-Černý 13, Subag 17, Ben Arous-Mei-Montanari-Nica 19, Subag-Zeitouni 21)



A Look at the Landscape

- If H_N is close to convex, maybe gradient descent works.
- Not the case! On the sphere, H_N can have:
  - Exponentially many near-optimal local maxima.
  - Exponentially *more* suboptimal local maxima.
  - Exponentially *more* suboptimal saddle points.
    (Auffinger-Ben Arous 13, A-BA-Černý 13, Subag 17, Ben Arous-Mei-Montanari-Nica 19, Subag-Zeitouni 21)



- **Adversarial** H_N: reaching $\frac{\mathrm{OPT}}{\log(N)^c}$ is hard (Arora-Berger-Hazan-Kindler-Safra 05, Barak-Brandao-Harrow-Kelner-Steurer-Zhou 12).

AMP Algorithms Succeed under No Overlap Gap

Theorem (Subag 18, Montanari 19, El Alaoui-Montanari-**S** 20, **S** 21)

The asymptotic value

$$\mathrm{ALG} = \inf_{\zeta \in \mathcal{L}} \mathcal{P}_{\xi}(\zeta)$$

is achievable by AMP (assuming a minimizer $\zeta_* \in \mathcal{L}$ exists).
$\mathcal{L} \supsetneq \mathcal{U}$ contains all bounded variation functions $\zeta : [0,1] \to \mathbb{R}^+$.

AMP Algorithms Succeed under No Overlap Gap

Theorem (Subag 18, Montanari 19, El Alaoui-Montanari-S 20, S 21)

The asymptotic value

$$\mathrm{ALG} = \inf_{\zeta \in \mathcal{L}} \mathcal{P}_\xi(\zeta)$$

is achievable by AMP (assuming a minimizer $\zeta_* \in \mathcal{L}$ exists).
$\mathcal{L} \supsetneq \mathcal{U}$ contains all bounded variation functions $\zeta : [0,1] \to \mathbb{R}^+$.

- Approximate message passing (AMP) is really efficient.
    - Uses only $O(1)$ queries of ∇H_N. Great for tons of problems.
    - (Bolthausen 14, Donoho-Maleki-Montanari 09, Bayati-Montanari 11, Javanmard-Montanari 13, Bayati-Lelarge-Montanari 15, Rush-Venkataramanan 18, Montanari-Venkataramanan 21,. . .)

AMP Algorithms Succeed under No Overlap Gap

Theorem (Subag 18, Montanari 19, El Alaoui-Montanari-**S** 20, **S** 21)

The asymptotic value

$$\mathsf{ALG} = \inf_{\zeta \in \mathcal{L}} \mathcal{P}_\xi(\zeta)$$

is achievable by AMP (assuming a minimizer $\zeta_* \in \mathcal{L}$ exists).
$\mathcal{L} \supsetneq \mathcal{U}$ contains all bounded variation functions $\zeta : [0,1] \to \mathbb{R}^+$.

- Approximate message passing (AMP) is really efficient.
  - Uses only $O(1)$ queries of ∇H_N. Great for tons of problems.
  - (Bolthausen 14, Donoho-Maleki-Montanari 09, Bayati-Montanari 11, Javanmard-Montanari 13, Bayati-Lelarge-Montanari 15, Rush-Venkataramanan 18, Montanari-Venkataramanan 21,...)
- In brief: take small steps to simulate an SDE related to $\mathcal{P}_\xi$.
  - (AMS 20, roughly): No SDE-based AMP can reach $\mathsf{ALG} + \varepsilon$.
- Equality case $\mathsf{ALG} = \mathsf{OPT}$ corresponds to *no overlap gap*.

Algorithmic Hardness from the Overlap Gap Property

Theorem (Gamarnik-Jagannath 20, G-J-Wein 20&21, S 21)

*No **stable** algorithm can achieve* OPT *unless aforementioned AMP algorithms succeed (in even models with* $\gamma_3 = \gamma_5 = \cdots = 0$*).*

Algorithmic Hardness from the Overlap Gap Property

Theorem (Gamarnik-Jagannath 20, G-J-Wein 20&21, S 21)

*No **stable** algorithm can achieve* OPT *unless aforementioned AMP algorithms succeed (in even models with* $\gamma_3 = \gamma_5 = \cdots = 0$*).*

- Stable algorithms include:
  - $O(1)$ iterations of gradient descent or AMP
  - ... or *any* "constant-order method" querying $\nabla^{O(1)} H_N$
  - Langevin dynamics run for $O(1)$ time
  - Low degree polynomials
  - Poly-size circuits with depth at most $\frac{\log N}{2 \log \log N}$.

Algorithmic Hardness from the Overlap Gap Property

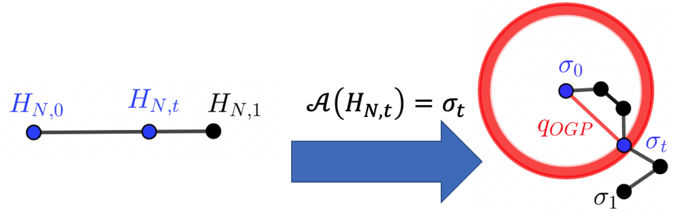
Theorem (Gamarnik-Jagannath 20, G-J-Wein 20&21, S 21)

No ***stable*** *algorithm can achieve* OPT *unless aforementioned AMP algorithms succeed (in even models with* $\gamma_3 = \gamma_5 = \cdots = 0$*).*

- Stable algorithms include:
  - $O(1)$ iterations of gradient descent or AMP
  - ... or *any* "constant-order method" querying $\nabla^{O(1)} H_N$
  - Langevin dynamics run for $O(1)$ time
  - Low degree polynomials
  - Poly-size circuits with depth at most $\frac{\log N}{2 \log \log N}$.
- Proof based on overlap gap property (OGP): a family of topological hardness criteria.

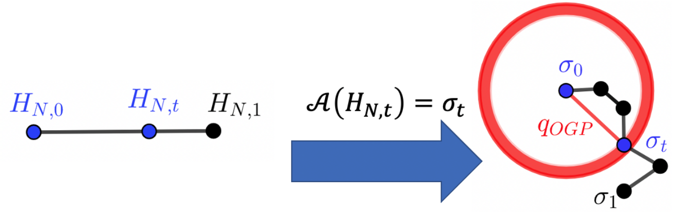
  (Achlioptas-Coja Oghlan 08, Gamarnik-Sudan 14, Gamarnik-Sudan 17 Rahman-Virag 17, Gamarnik-Zadik 17, Chen-Gamarnik-Panchenko-Rahman 17, Gamarnik-Jagannath-Sen 19, Wein 20, ...)

- Consider path $H_{N,t} = \sqrt{1-t}H_{N,0} + \sqrt{t}H_{N,1}$.
- $\boldsymbol{\sigma}_t = \mathcal{A}(H_{N,t})$ is morally continuous in t.



Overlap Gap Property for Spin Glasses: A Cartoon

- Consider path $H_{N,t} = \sqrt{1-t}H_{N,0} + \sqrt{t}H_{N,1}$.
- $\boldsymbol{\sigma}_t = \mathcal{A}(H_{N,t})$ is morally continuous in t.



- Overlap gap property: for some $q_{\text{OGP}} \in (0,1)$, if $||\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_t|| \approx q_{\text{OGP}}\sqrt{N}$, then either $\boldsymbol{\sigma}_0$ or $\boldsymbol{\sigma}_t$ is suboptimal.

$$\min\left(H_{N,0}(\boldsymbol{\sigma}_0), H_{N,t}(\boldsymbol{\sigma}_t)\right) \le (\text{OPT} - \varepsilon)N.$$

- "Continuity" implies $||\boldsymbol{\sigma}_t - \boldsymbol{\sigma}_0|| \approx q_{\text{OGP}}\sqrt{N}$ for some $t \in [0,1]$.

New Result: An Algorithmic Threshold

Theorem (Huang-S. 21+)

*No **overlap concentrated** algorithm can beat ALG (in even models).*

- Overlap concentrated algorithms include:
  - $O(1)$ iterations of gradient descent or AMP
  - ... or *any* constant-order method querying $\nabla^{O(1)} H_N$
  - Langevin dynamics run for $O(1)$ time
  - ~~Low degree polynomials~~
  - ~~Poly-size circuits with depth $\frac{\log N}{2 \log \log N}$~~.

New Result: An Algorithmic Threshold

Theorem (Huang-S. 21+)

*No **overlap concentrated** algorithm can beat ALG (in even models).*

- Overlap concentrated algorithms include:
  - $O(1)$ iterations of gradient descent or AMP
  - ... or *any* constant-order method querying $\nabla^{O(1)} H_N$
  - Langevin dynamics run for $O(1)$ time
  - ~~Low degree polynomials~~
  - ~~Poly-size circuits with depth $\frac{\log N}{2 \log \log N}$~~.
- For the algorithms listed above, result holds in a strong sense:

$$\mathbb{P}[H_N(\mathcal{A}(H_N)) \geq (\mathsf{ALG} + \varepsilon)N] \leq O(e^{-c(\varepsilon)N}).$$

New Result: An Algorithmic Threshold

Theorem (Huang-S. 21+)

*No **overlap concentrated** algorithm can beat ALG (in even models).*

- Overlap concentrated algorithms include:
  - $O(1)$ iterations of gradient descent or AMP
  - ... or *any* constant-order method querying $\nabla^{O(1)} H_N$
  - Langevin dynamics run for $O(1)$ time
  - ~~Low degree polynomials~~
  - ~~Poly-size circuits with depth $\frac{\log N}{2 \log \log N}$~~.
- For the algorithms listed above, result holds in a strong sense:

$$\mathbb{P}[H_N(\mathcal{A}(H_N)) \geq (\mathrm{ALG} + \varepsilon)N] \leq O(e^{-c(\varepsilon)N}).$$

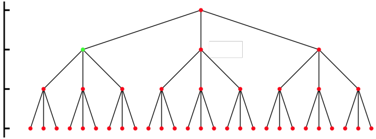
- Proof relies on a new **branching OGP**.
- In spherical models, branching OGP is in some sense **necessary** to rule out $\mathrm{ALG} + \varepsilon$. Simpler OGPs cannot.

Ultrametric Spaces and Trees

- Recall: ultrametric spaces X satisfy the ultrametric triangle inequality

$$d(x, y) \leq \max(d(x, z), d(y, z)), \quad \forall x, y, z \in X.$$

Equivalent to hierarchical clustering, or graph metrics of leaves of a rooted tree.
(All ultrametrics will be finite with sensible diameter.)

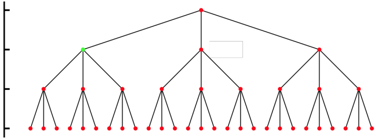


Ultrametric Spaces and Trees

- Recall: ultrametric spaces X satisfy the ultrametric triangle inequality

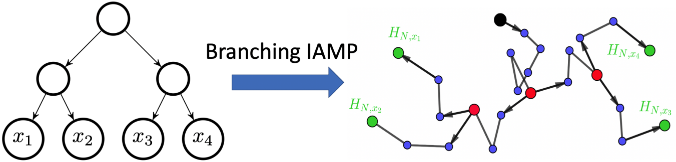
$$d(x,y) \le \max(d(x,z), d(y,z)), \quad \forall x,y,z \in X.$$

Equivalent to hierarchical clustering, or graph metrics of leaves of a rooted tree.
(All ultrametrics will be finite with sensible diameter.)



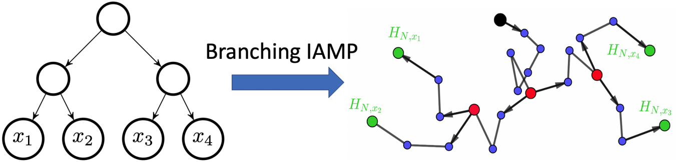
- For all $\beta > 0$, Gibbs measure $e^{\beta H_N(\boldsymbol{\sigma})} d\boldsymbol{\sigma}/Z$ is "$\approx$ ultrametric".
(Parisi 82, Mézard-Parisi-Sourlas-Toulouse-Virasoro 84, Derrida 85, Ruelle 87, Panchenko 13, Jagannath 17, Chatterjee-Sloman 20,...)

- Turns out algorithms can build ultrametric spaces!
- AMP algorithms for this problem explore using many small steps. Branch to get a multi-valued algorithm.



Algorithms and Ultrametrics

- Turns out algorithms can build ultrametric spaces!
- AMP algorithms for this problem explore using many small steps. Branch to get a multi-valued algorithm.



- Result: for any finite ultrametric X, branching IAMP can output a configuration $(\boldsymbol{\sigma}_x)_{x\in X}$ approximating X:

$$H_N(\boldsymbol{\sigma}_x) \approx \mathsf{ALG} \cdot N, \quad \forall x \in X,$$

$$||\boldsymbol{\sigma}_x - \boldsymbol{\sigma}_y|| \approx d_X(x,y)\sqrt{N}, \quad \forall x,y \in X.$$

(Subag 18, El Alaoui-Montanari 20, **S** 21)

Branching OGP

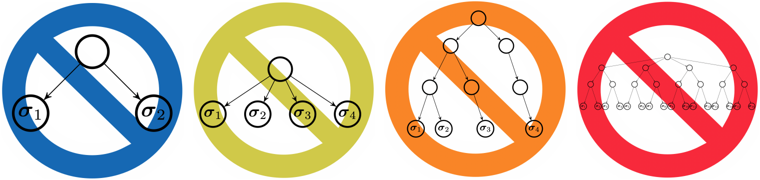
- If algorithms can recreate any ultrametric, we should use a stronger OGP.

Branching OGP

- If algorithms can recreate any ultrametric, we should use a stronger OGP.
- Classic OGP: rule out pairs $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ at distance $q_{\mathrm{OGP}}\sqrt{N}$.
- 1 layer OGP: tuples $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m)$ with all distances $q_{\mathrm{OGP}}\sqrt{N}$.
- Ladder OGP: $Dist(\boldsymbol{\sigma}_{i+1}, span(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_i)) = \delta\sqrt{N}$.

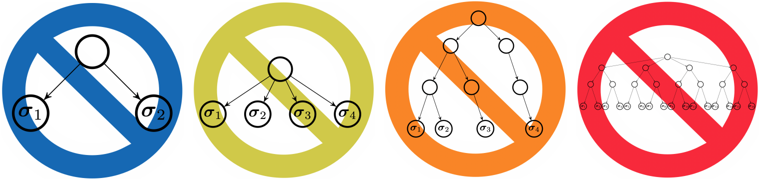
Branching OGP

- If algorithms can recreate any ultrametric, we should use a stronger OGP.
- Classic OGP: rule out pairs $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ at distance $q_{\text{OGP}}\sqrt{N}$.
- 1 layer OGP: tuples $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m)$ with all distances $q_{\text{OGP}}\sqrt{N}$.
- Ladder OGP: $Dist(\boldsymbol{\sigma}_{i+1}, span(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_i)) = \delta\sqrt{N}$.
- Branching OGP: rule out configurations $(\boldsymbol{\sigma}_x)_{x \in X}$ approximating an **arbitrarily complicated** ultrametric space.



Branching OGP

- If algorithms can recreate any ultrametric, we should use a stronger OGP.
- Classic OGP: rule out pairs $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ at distance $q_{\text{OGP}}\sqrt{N}$.
- 1 layer OGP: tuples $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m)$ with all distances $q_{\text{OGP}}\sqrt{N}$.
- Ladder OGP: $Dist(\boldsymbol{\sigma}_{i+1}, span(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_i)) = \delta\sqrt{N}$.
- Branching OGP: rule out configurations $(\boldsymbol{\sigma}_x)_{x \in X}$ approximating an **arbitrarily complicated** ultrametric space.



- But...**why would branching OGP imply hardness**?

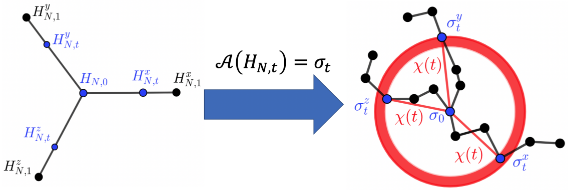
Overlap Concentrated Algorithms

Definition

An algorithm $\mathcal{A}$ is **overlap concentrated** if the random distance

$$\frac{||\mathcal{A}(H_{N,0}) - \mathcal{A}(H_{N,t})||}{\sqrt{N}}$$

tightly concentrates around its mean $\chi(t)$, uniformly over $t \in [0, 1]$.



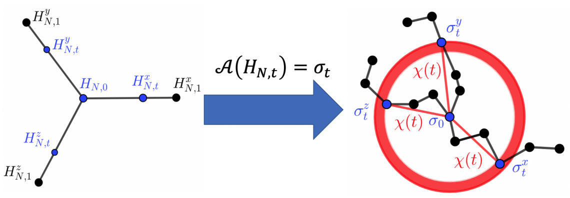
Overlap Concentrated Algorithms

Definition

An algorithm $\mathcal{A}$ is **overlap concentrated** if the random distance

$$\frac{||\mathcal{A}(H_{N,0}) - \mathcal{A}(H_{N,t})||}{\sqrt{N}}$$

tightly concentrates around its mean $\chi(t)$, uniformly over $t \in [0,1]$.

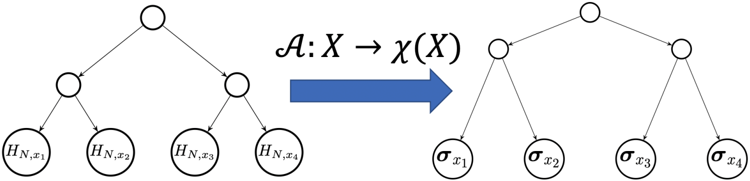


- Holds by concentration of measure if $\mathcal{A}$ is Lipschitz in H_N.
- $\implies$ gradient descent, AMP, ... are overlap concentrated.

Ultrametric Transformations from Overlap Concentration

- For ultrametric X, create correlated Hamiltonians $(H_{N,x})_{x \in X}$.
- Outputs $\boldsymbol{\sigma}_x = \mathcal{A}(H_{N,x})$ form a new ultrametric space $\chi(X)$:

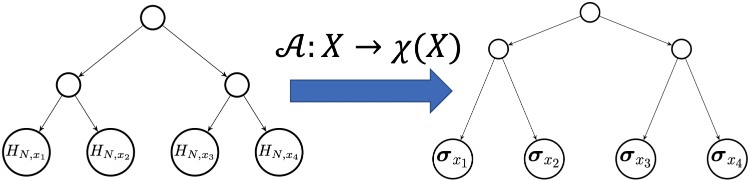
$$\frac{||\boldsymbol{\sigma}_{x_1} - \boldsymbol{\sigma}_{x_2}||}{\sqrt{N}} \approx \chi(d_X(x_1, x_2)), \quad x_1, x_2 \in X.$$



Ultrametric Transformations from Overlap Concentration

- For ultrametric X, create correlated Hamiltonians $(H_{N,x})_{x\in X}$.
- Outputs $\boldsymbol{\sigma}_x = \mathcal{A}(H_{N,x})$ form a new ultrametric space $\chi(X)$:

$$\frac{||\boldsymbol{\sigma}_{x_1} - \boldsymbol{\sigma}_{x_2}||}{\sqrt{N}} \approx \chi(d_X(x_1,x_2)), \quad x_1,x_2 \in X.$$



- χ is continuous, so we can make $\chi(X)$ **any ultrametric**.
- $\implies$ if $\mathcal{A}$ achieves $(\mathrm{ALG}+\varepsilon)N$, then there is a configuration $(\boldsymbol{\sigma}_x)_{x\in X}$ approximating **any desired ultrametric** $\chi(X)$ with

$$H_{N,x}(\boldsymbol{\sigma}_x) \geq (\mathrm{ALG}+\varepsilon)N, \quad \forall x \in X.$$

Ruling Out a Complicated Ultrametric

- Take $\chi(X)$ a k-ary tree branching at depths $[0, \delta, 2\delta, \ldots, 1]$.
  - These are "universal" for finite ultrametrics as $\delta \to 0$.

Ruling Out a Complicated Ultrametric

- Take $\chi(X)$ a k-ary tree branching at depths $[0, \delta, 2\delta, \ldots, 1]$.
  - These are "universal" for finite ultrametrics as $\delta \to 0$.
- Parisi formula upper bound generalizes to richer settings. Control average of $H_{N,x}(\boldsymbol{\sigma}_x)$ over all $\chi(X)$-configurations:

$$\max_{(\boldsymbol{\sigma}_x)_{x \in X}} \left\{ \frac{1}{|X|} \sum_{x \in X} H_{N,x}(\boldsymbol{\sigma}_x) : \frac{||\boldsymbol{\sigma}_{x_1} - \boldsymbol{\sigma}_{x_2}||}{\sqrt{N}} \approx d_{\chi(X)}(x_1, x_2) \quad \forall x_1, x_2 \in X \right\}.$$

- Upper bounds from any **increasing** function $\zeta : [0,1] \to \mathbb{R}^+$, expressed as multi-dimensional generalizations $\mathcal{P}_\xi^X$ of $\mathcal{P}_\xi$.

Ruling Out a Complicated Ultrametric

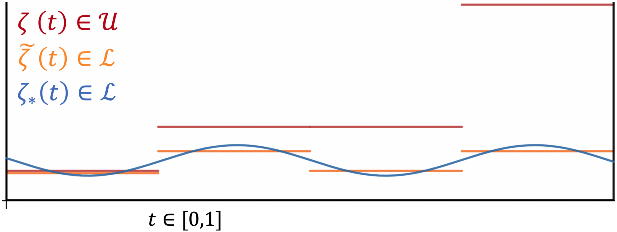
- Take $\chi(X)$ a k-ary tree branching at depths $[0, \delta, 2\delta, \ldots, 1]$.
  - These are "universal" for finite ultrametrics as $\delta \to 0$.
- Parisi formula upper bound generalizes to richer settings. Control average of $H_{N,x}(\boldsymbol{\sigma}_x)$ over all $\chi(X)$-configurations:

$$\max_{(\boldsymbol{\sigma}_x)_{x\in X}} \left\{ \frac{1}{|X|} \sum_{x \in X} H_{N,x}(\boldsymbol{\sigma}_x) : \frac{||\boldsymbol{\sigma}_{x_1} - \boldsymbol{\sigma}_{x_2}||}{\sqrt{N}} \approx d_{\chi(X)}(x_1, x_2) \;\; \forall x_1, x_2 \in X \right\}.$$

- Upper bounds from any **increasing** function $\zeta : [0,1] \to \mathbb{R}^+$, expressed as multi-dimensional generalizations $\mathcal{P}_\xi^X$ of $\mathcal{P}_\xi$.
- Eventually, obtain upper bound $\mathcal{P}_\xi(\widetilde{\zeta})$ in terms of the original Parisi functional.
- **Increasing** ζ transforms into **no-longer increasing** $\widetilde{\zeta}$.

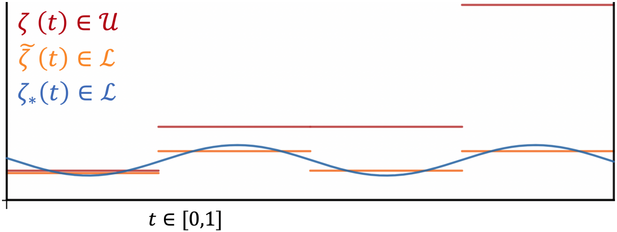
Ruling Out a Complicated Ultrametric

- The ratio $\widetilde{\zeta}/\zeta$ is piece-wise constant, shrinks at each $j\delta$.
- Hence $\widetilde{\zeta}$ approximates any $\zeta_* \in \mathcal{L}$, get upper bound $\mathsf{ALG} + \varepsilon$.



Ruling Out a Complicated Ultrametric

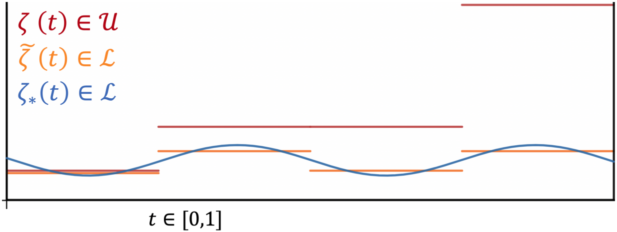
- The ratio $\tilde{\zeta}/\zeta$ is piece-wise constant, shrinks at each $j\delta$.
- Hence $\tilde{\zeta}$ approximates any $\zeta_* \in \mathcal{L}$, get upper bound ALG $+ \varepsilon$.



- Classic OGP: tree **branches once**, so $\tilde{\zeta}$ **decreases once**.
  - Cannot approximate general functions, cannot reach ALG.
  - This limitation is fundamental:

Ruling Out a Complicated Ultrametric

- The ratio $\tilde{\zeta}/\zeta$ is piece-wise constant, shrinks at each $j\delta$.
- Hence $\tilde{\zeta}$ approximates any $\zeta_* \in \mathcal{L}$, get upper bound $\mathrm{ALG} + \varepsilon$.



- Classic OGP: tree **branches once**, so $\tilde{\zeta}$ **decreases once**.
  - Cannot approximate general functions, cannot reach ALG.
  - This limitation is fundamental:

Theorem (Huang-S 21+)

On the sphere, to rule out $(\mathrm{ALG} + \varepsilon)$ *with forbidden ultrametric trees, the trees must contain full binary subtrees of unbounded size (as* $\varepsilon \to 0$*).*

Conclusion

- Problem: optimize a mean field spin glass Hamiltonian.
- Result: ALG is the best asymptotic value achievable by overlap concentrated algorithms, so existing AMP algorithms are optimal within this class (modulo technical points).

Conclusion

- Problem: optimize a mean field spin glass Hamiltonian.
- Result: ALG is the best asymptotic value achievable by overlap concentrated algorithms, so existing AMP algorithms are optimal within this class (modulo technical points).

- Overlap concentrated algorithms include constant-order methods such as GD and AMP with $O(1)$ iterations.
- Proof uses a **branching** OGP based on general ultrametric trees, which is in some sense necessary.

Conclusion

- Problem: optimize a mean field spin glass Hamiltonian.
- Result: ALG is the best asymptotic value achievable by overlap concentrated algorithms, so existing AMP algorithms are optimal within this class (modulo technical points).

- Overlap concentrated algorithms include constant-order methods such as GD and AMP with $O(1)$ iterations.
- Proof uses a **branching** OGP based on general ultrametric trees, which is in some sense necessary.

- A natural open direction: how generally does branching OGP identify the exact algorithmic threshold?