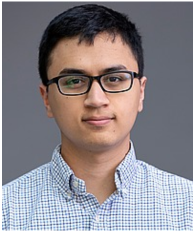
A constructive proof of the spherical Parisi formula

Brice Huang (MIT)

BIRS Workshop in Computational Complexity of Statistical Inference

Joint work with Mark Sellke (Harvard)

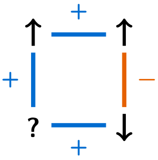


Mean-field spin glasses

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

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

- Ground state energy: typical max of $H_N(\boldsymbol{\sigma})/N$
- Random couplings $g_{i,j}$ lead to highly non-trivial behavior



Mixed p-spin models

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

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

Mixed p-spin models

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

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

- (Ising / spherical) mixed p-spin model: linear mixture of above

$$H_N(\boldsymbol{\sigma}) = \sum_{p \geq 1} \frac{\beta_p}{N^{(p-1)/2}} \sum_{\underline{i} \in [N]^p} g_{\underline{i}} (\boldsymbol{\sigma}^{\otimes p})_{\underline{i}}, \qquad \boldsymbol{\sigma} \in \{\pm 1\}^N \text{ or } S_N$$

$\xi =$ “model” encodes $(\beta_p)_{p \geq 1}$

Mixed p-spin models

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

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

- (Ising / spherical) mixed p-spin model: linear mixture of above

$$H_N(\boldsymbol{\sigma}) = \sum_{p \geq 1} \frac{\beta_p}{N^{(p-1)/2}} \sum_{\underline{i} \in [N]^p} g_{\underline{i}} (\boldsymbol{\sigma}^{\otimes p})_{\underline{i}}, \qquad \boldsymbol{\sigma} \in \{\pm 1\}^N \text{ or } S_N$$

ξ = "model" encodes $(\beta_p)_{p \geq 1}$

- Connections to tensor PCA, random max-k-SAT, high-dim inference, ...
(Ben Arous–Mei–Montanari-Nica 17, Dembo–Montanari–Sen 17, Panchenko 18, Fan–Mei–Montanari 21)

Free energy, Gibbs measure

- Free energy: softmax version of ground state. Let

$$Z_N = \int_{S_N} e^{\beta H_N(\boldsymbol{\sigma})} \, d\boldsymbol{\sigma} \qquad \textbf{(partition function)}$$

Main question: what is the limit of the **free energy** $\frac{1}{N} \log Z_N$?

Free energy, Gibbs measure

- Free energy: softmax version of ground state. Let

$$Z_N = \int_{S_N} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma} \qquad \textbf{(partition function)}$$

Main question: what is the limit of the **free energy** $\frac{1}{N} \log Z_N$?

Free energy, Gibbs measure

- Free energy: softmax version of ground state. Let

$$Z_N = \int_{S_N} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma} \qquad \textbf{(partition function)}$$

Main question: what is the limit of the **free energy** $\frac{1}{N} \log Z_N$?

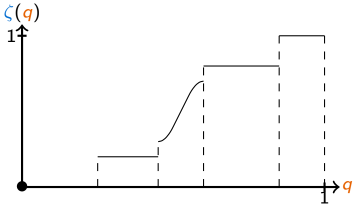
- Gibbs measure:

$$\mathrm{d}\mu_N(\boldsymbol{\sigma}) = \frac{1}{Z_N} e^{H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$$

Limiting behavior of μ_N? E.g. for replicas $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \sim \mu_N$, what is the law of the **overlap** $\frac{\langle \boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \rangle}{N} \in [-1, 1]$?

Physics predictions (now theorems)

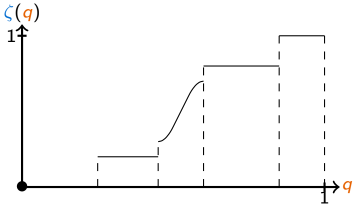
- Order parameter: probability measure ζ on $[0,1]$; identify with CDF:



Overlap distribution: $\zeta_* = \lim_{N\to\infty} \mathrm{Law}(\langle \boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \rangle / N)$

Physics predictions (now theorems)

- Order parameter: probability measure ζ on $[0,1]$; identify with CDF:



Overlap distribution: $\zeta_* = \lim_{N\to\infty} \mathrm{Law}(\langle \boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \rangle / N)$

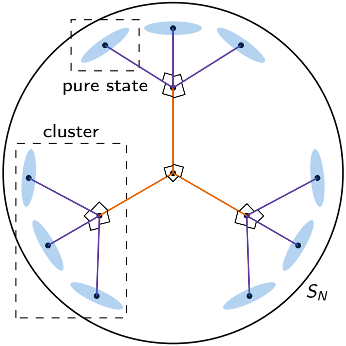
- **Parisi formula**:

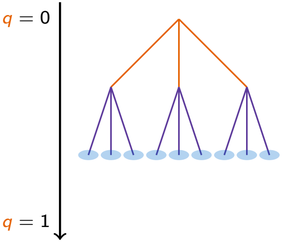
$$\lim_{N\to\infty} \frac{1}{N} \log Z_N = \min_{\zeta} \mathcal{P}(\zeta;\xi) = \mathcal{P}(\zeta_*;\xi)$$

where $\mathcal{P}$ = Parisi functional, and recall ξ = model

Ultrametricity

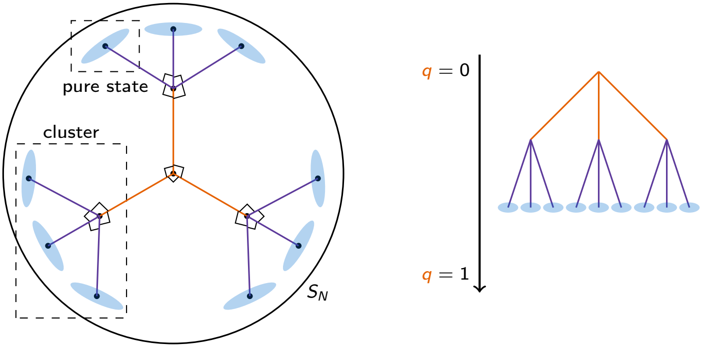
- μ_N convex combination of **pure states** with **hierarchical clustering**
  Parisi 83 (heuristic), Panchenko 13 (rigorous)





Ultrametricity

- μ_N convex combination of **pure states** with **hierarchical clustering**
  Parisi 83 (heuristic), Panchenko 13 (rigorous)



- Tree branches at radii in $\mathrm{supp}\ \zeta_*$ (cts support $\to$ cts branching, **full RSB**)

History of rigorous results

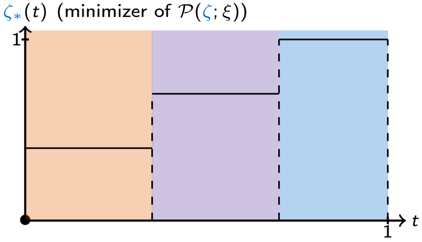
For both Ising, spherical mixed p-spin models:

- Guerra 03: $\limsup_{N\to\infty} \frac{1}{N} \log Z_N \le \mathcal{P}(\zeta_*; \xi)$ by interpolation argument
- Talagrand 06: matching lower bound, proves Parisi formula
  - **Analytic proof**, self-bounds the error in Guerra's upper bound

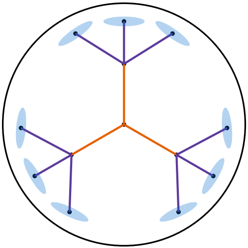
- Panchenko 13: ultrametricity of asymptotic Gibbs measures μ_N
  - $\Rightarrow$ New proof of Parisi formula
  - **Inductive proof on # spins**, focuses on understanding Gibbs measure precisely

- Jagannath 17, Subag 18, Chatterjee–Sloman 21: approximate ultrametricity for finite N
- Jagannath–Tobasco 18: ζ_* finitely many atomic & continuous pieces, in spherical models

- New proof of Parisi lower bound for spherical mixed p-spin models: $\liminf_{N\to\infty} \frac{1}{N} \log Z_N \geq \mathcal{P}(\zeta_*; \xi)$
- **Geometric approach**: directly constructs ultrametric tree of pure states in accordance with Parisi ansatz



$\Rightarrow$



Spherical mixed p-spin model

- Recall model: $\xi = (\beta_p)_{p \geq 1}$,

$$H_N(\boldsymbol{\sigma}) = \sum_{p \geq 1} \frac{\beta_p}{N^{(p-1)/2}} \sum_{\underline{i} \in [N]^p} g_{\underline{i}} (\boldsymbol{\sigma}^{\otimes p})_{\underline{i}} \qquad g_{\underline{i}} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

- Domain: $S_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N : \|\boldsymbol{\sigma}\| = \sqrt{N}\}$
- Free energy:

$$\frac{1}{N} \log \int_{S_N} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma}$$

Existence of ultrametric tree

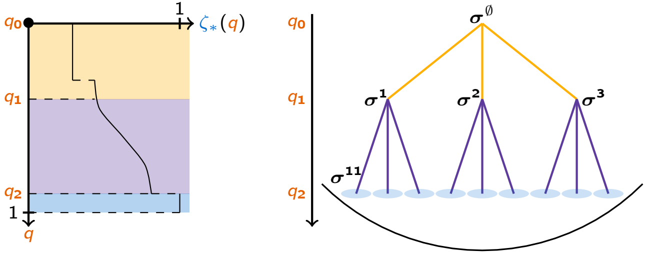
Theorem (H.-Sellke 23)

Let $q_0 < \cdots < q_D \in \operatorname{supp} \zeta_*$, $q_D = \max \operatorname{supp} \zeta_*$.

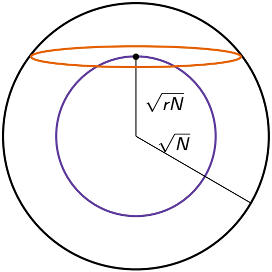
$\mathbb{T}_{k,D}$ k-ary tree of depth D. Whp, exists ultrametric $\{\boldsymbol{\sigma}^u : u \in \mathbb{T}_{k,D}\}$ such that each $\boldsymbol{\sigma}^u$, u leaf, is the center of a band B with

$$\frac{1}{N} \log \int_B e^{H_N(\boldsymbol{\sigma})}\, \mathrm{d}\boldsymbol{\sigma} \geq \mathcal{P}(\zeta_*; \xi) - \varepsilon.$$

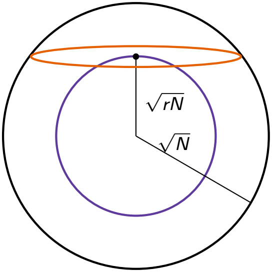
$\Rightarrow$ New proof of Parisi LB!

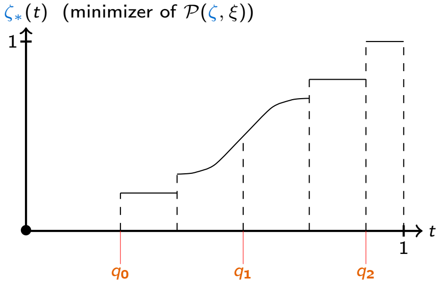


Proof idea 1: Decomposition into fundamental types

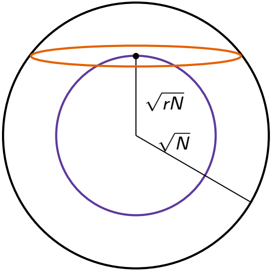


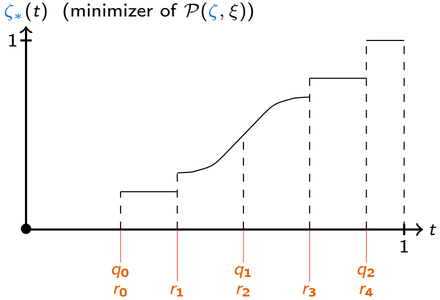
Proof idea 1: Decomposition into fundamental types



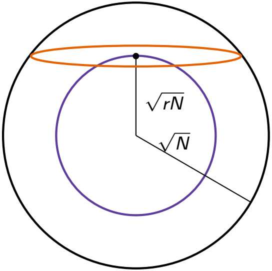


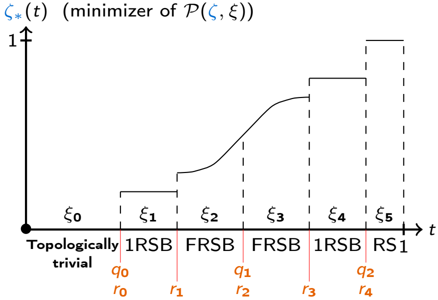
Proof idea 1: Decomposition into fundamental types



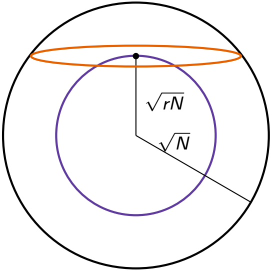


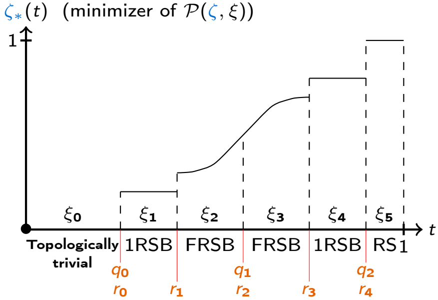
Proof idea 1: Decomposition into fundamental types





Proof idea 1: Decomposition into fundamental types

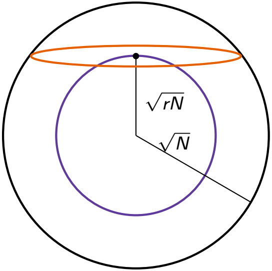


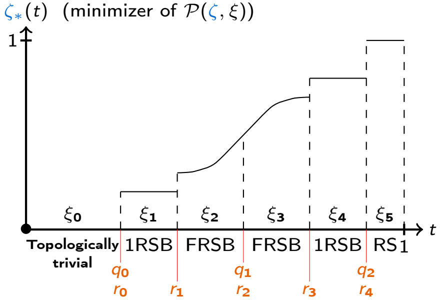


By following maxima of successive sub-models, can show: (Subag 18)

$$FE(\xi) \geq GSE(\xi_0) + \cdots + GSE(\xi_{D-1}) + FE(\xi_D)$$

Proof idea 1: Decomposition into fundamental types





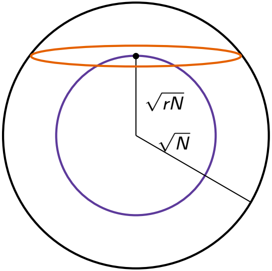
By following maxima of successive sub-models, can show: (Subag 18)

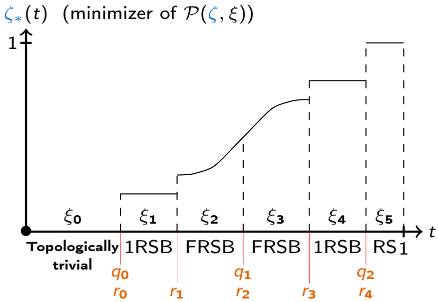
$$FE(\xi) \geq GSE(\xi_0) + \cdots + GSE(\xi_{D-1}) + FE(\xi_D)$$

This inequality is tight, since can show:

$$\mathcal{P}(\xi) = \mathcal{P}_{GS}(\xi_0) + \cdots + \mathcal{P}_{GS}(\xi_{D-1}) + \mathcal{P}(\xi_D)$$

Proof idea 1: Decomposition into fundamental types





By following maxima of successive sub-models, can show: (Subag 18)

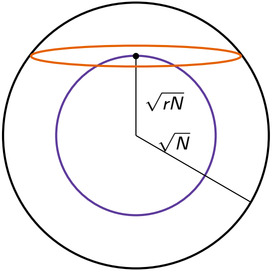
$$FE(\xi) \geq GSE(\xi_0) + \cdots + GSE(\xi_{D-1}) + FE(\xi_D)$$

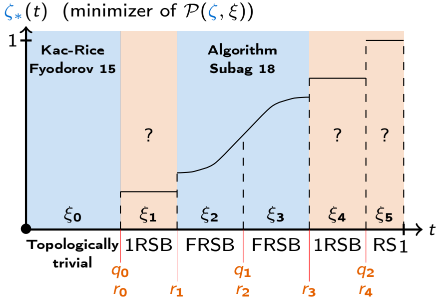
This inequality is tight, since can show:

$$\mathcal{P}(\xi) = \mathcal{P}_{GS}(\xi_0) + \cdots + \mathcal{P}_{GS}(\xi_{D-1}) + \mathcal{P}(\xi_D)$$

$\Rightarrow$ suffices to analyze fundamental types.

Proof idea 1: Decomposition into fundamental types





By following maxima of successive sub-models, can show: (Subag 18)

$$FE(\xi) \geq GSE(\xi_0) + \cdots + GSE(\xi_{D-1}) + FE(\xi_D)$$

This inequality is tight, since can show:

$$\mathcal{P}(\xi) = \mathcal{P}_{GS}(\xi_0) + \cdots + \mathcal{P}_{GS}(\xi_{D-1}) + \mathcal{P}(\xi_D)$$

$\Rightarrow$ suffices to analyze fundamental types. Top triv & FRSB known!

Proof idea 2: Truncated 2nd moment method

- For RS models, Parisi LB amounts to showing

$$Z_N = \int_{S_N} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma} \ge e^{-o(N)} \mathbb{E} Z_N$$

Proof idea 2: Truncated 2nd moment method

- For RS models, Parisi LB amounts to showing

$$Z_N = \int_{S_N} e^{H_N(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma} \ge e^{-o(N)} \mathbb{E} Z_N$$

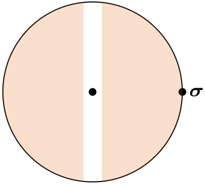
- Vanilla 2nd moment on Z_N: **doesn't always give sharp LB**

Proof idea 2: Truncated 2nd moment method

- For RS models, Parisi LB amounts to showing

$$Z_N = \int_{S_N} e^{H_N(\boldsymbol{\sigma})}\, \mathrm{d}\boldsymbol{\sigma} \geq e^{-o(N)} \mathbb{E} Z_N$$

- Vanilla 2nd moment on Z_N: **doesn't always give sharp LB**
Truncation: $\boldsymbol{\sigma}$ **typical** if $\{\boldsymbol{\rho} : \frac{1}{N}|\langle \boldsymbol{\rho}, \boldsymbol{\sigma}\rangle| \geq \delta\}$ accounts for $\leq e^{-\varepsilon N}$ of Z_N



Do 2nd moment on truncation: $\widetilde{Z}_N = \int_{\boldsymbol{\sigma}\ \textbf{typical}} e^{H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$

Proof idea 2: Truncated 2nd moment method (details)

- Truncated partition fn $\widetilde{Z}_N = \int_{\boldsymbol{\sigma}\ \textbf{typical}} e^{H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$
  $\mathbb{E}[\widetilde{Z}_N^2] \approx \mathbb{E}[Z_N]^2$ automatic; main work is to show $\mathbb{E}[\widetilde{Z}_N] \approx \mathbb{E}[Z_N]$

Proof idea 2: Truncated 2nd moment method (details)

- Truncated partition fn $\widetilde{Z}_N = \int_{\boldsymbol{\sigma} \textbf{ typical}} e^{H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$
  $\mathbb{E}[\widetilde{Z}_N^2] \approx \mathbb{E}[Z_N]^2$ automatic; main work is to show $\mathbb{E}[\widetilde{Z}_N] \approx \mathbb{E}[Z_N]$

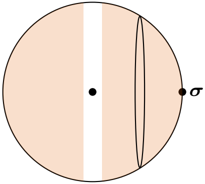
- Idea: $\mathbb{E}[Z_N]$ dominated by $\boldsymbol{\sigma}$ with energy $\approx E_* N$
  Conditional on $H_N(\boldsymbol{\sigma}) \approx E_* N$, H_N is spin glass with **planted spike** at $\boldsymbol{\sigma}$

Proof idea 2: Truncated 2nd moment method (details)

- Truncated partition fn $\widetilde{Z}_N = \int_{\boldsymbol{\sigma} \textbf{ typical}} e^{H_N(\boldsymbol{\sigma})} \mathrm{d}\boldsymbol{\sigma}$
  $\mathbb{E}[\widetilde{Z}_N^2] \approx \mathbb{E}[Z_N]^2$ automatic; main work is to show $\mathbb{E}[\widetilde{Z}_N] \approx \mathbb{E}[Z_N]$

- Idea: $\mathbb{E}[Z_N]$ dominated by $\boldsymbol{\sigma}$ with energy $\approx E_* N$
  Conditional on $H_N(\boldsymbol{\sigma}) \approx E_* N$, H_N is spin glass with **planted spike** at $\boldsymbol{\sigma}$

  Guerra's upper bound controls FE of each non-equatorial band:



  $\Rightarrow$ conditional on $H_N(\boldsymbol{\sigma}) \approx E_* N$, $\boldsymbol{\sigma}$ is whp typical

Conclusion

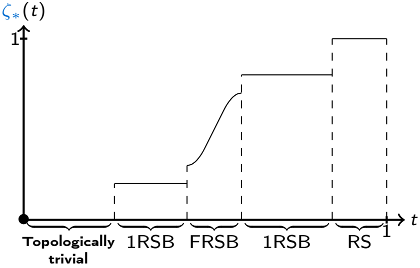
- Truncated 2nd moment $\Rightarrow$ FE of RS models

Conclusion

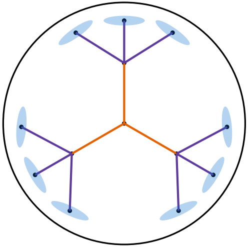
- Truncated 2nd moment $\Rightarrow$ FE of RS models
- Truncated 2nd moment on critical points $\Rightarrow$ GSE of 1RSB models.
  Also shows k orthogonal near-maximizers

Conclusion

- Truncated 2nd moment $\Rightarrow$ FE of RS models
- Truncated 2nd moment on critical points $\Rightarrow$ GSE of 1RSB models.
  Also shows k orthogonal near-maximizers
- Construct tree witnessing Parisi LB by taking k maximizers of each sub-model

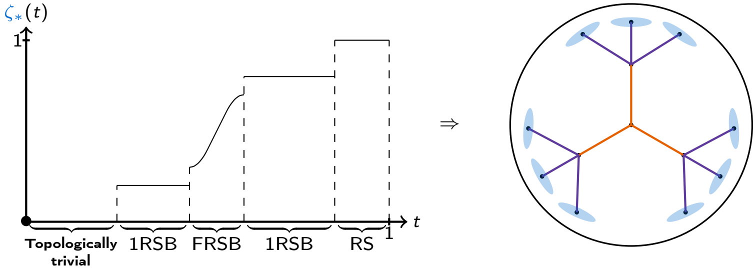


$\Rightarrow$



Conclusion

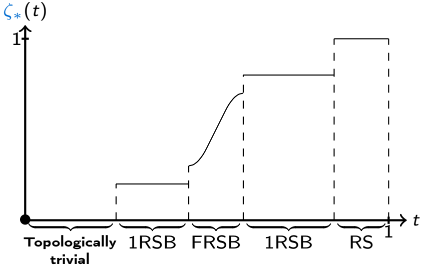
- Truncated 2nd moment ⇒ FE of RS models
- Truncated 2nd moment on critical points ⇒ GSE of 1RSB models.
  Also shows k orthogonal near-maximizers
- Construct tree witnessing Parisi LB by taking k maximizers of each sub-model



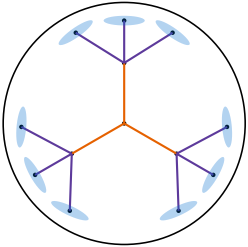
- Q: Can we construct a tree in this way that exhausts the Gibbs measure?

Conclusion

- Truncated 2nd moment $\Rightarrow$ FE of RS models
- Truncated 2nd moment on critical points $\Rightarrow$ GSE of 1RSB models.
  Also shows k orthogonal near-maximizers
- Construct tree witnessing Parisi LB by taking k maximizers of each sub-model



$\Rightarrow$



- Q: Can we construct a tree in this way that exhausts the Gibbs measure?

Thanks!