Spin glasses, algorithms, and inference

Brice Huang (MIT → Stanford → Yale)

Statistical physics & machine learning: moving forward

Cargèse institute | August 14, 2025

Thanks to wonderful collaborators

Mark Sellke

Nike Sun

Guy Bresler

Andrea Montanari

Huy Tuan Pham

Sidhanth Mohanty

Amit Rajaraman

David X. Wu

Lecture outline

1. Applications of planting in disordered models

2. A survey on the overlap gap property

Part I: applications of planting in disordered models

Planted models

In this conference, we've seen many problems about recovering **planted signal** from **signal + noise**:

Planted models

In this conference, we've seen many problems about recovering **planted signal** from **signal + noise**:

- **Planted clique**: find a k-clique planted in $G(N, 1/2)$ (Jerrum 92, Ma Wu 13, Brennan Bresler 18+19+20, Lee Pernice Rajaraman Zadik 25)
- **Tensor PCA**: recover rank 1 spike planted in gaussian p-tensor (Montanari Richard 14, Hopkins Shi Steurer 15, Wein Alaoui Moore 19, Ben Arous Gheissari Jagannath 20, Ben Arous Gerbelot Piccolo 24)
- **Single/multi-index models**: recover $\boldsymbol{W}^*$ from $y_i = f(\boldsymbol{W}^* \boldsymbol{x}_i, \varepsilon)$ (Damian Lee Soltanolkotabi 22, Damian Pillaud-Vivien Lee Bruna 24, DLB 25, Troiani Dandi Defilippis Zdeborová Loureiro Krzakala 25)

Planted models

In this conference, we've seen many problems about recovering **planted signal** from **signal + noise**:

- **Planted clique**: find a k-clique planted in $G(N, 1/2)$ (Jerrum 92, Ma Wu 13, Brennan Bresler 18+19+20, Lee Pernice Rajaraman Zadik 25)
- **Tensor PCA**: recover rank 1 spike planted in gaussian p-tensor (Montanari Richard 14, Hopkins Shi Steurer 15, Wein Alaoui Moore 19, Ben Arous Gheissari Jagannath 20, Ben Arous Gerbelot Piccolo 24)
- **Single/multi-index models**: recover $\boldsymbol{W}^*$ from $y_i = f(\boldsymbol{W}^* \boldsymbol{x}_i, \varepsilon)$ (Damian Lee Soltanolkotabi 22, Damian Pillaud-Vivien Lee Bruna 24, DLB 25, Troiani Dandi Defilippis Zdeborová Loureiro Krzakala 25)

This talk: planted models are also useful as a **proof device** for studying "null" models **without planted signal**

Outline of part I: applications of planting in disordered models

The classic planting trick: planting a Gibbs sample

Ground state large deviations in spherical spin glasses

TAP planting: capacity of the Ising perceptron

Gibbs measures: prototypical examples

Sherrington–Kirkpatrick model: for $\boldsymbol{\sigma} \in \{\pm 1\}^N$, $\boldsymbol{W} \sim \mathrm{GOE}(N)$:

$$H(\boldsymbol{\sigma}) = \frac{1}{2}(\boldsymbol{W}\boldsymbol{\sigma}, \boldsymbol{\sigma}) \qquad \text{Gibbs measure: } \mu_{\beta H}(\boldsymbol{\sigma}) = \frac{1}{Z} e^{\beta H(\boldsymbol{\sigma})}$$

Gibbs measures: prototypical examples

Sherrington–Kirkpatrick model: for $\boldsymbol{\sigma} \in \{\pm 1\}^N$, $\boldsymbol{W} \sim \mathrm{GOE}(N)$:

$$H(\boldsymbol{\sigma}) = \frac{1}{2}(\boldsymbol{W}\boldsymbol{\sigma}, \boldsymbol{\sigma}) \qquad \text{Gibbs measure: } \mu_{\beta H}(\boldsymbol{\sigma}) = \frac{1}{Z} e^{\beta H(\boldsymbol{\sigma})}$$

Random k-**NAE-SAT**: AND of not-all-equal clauses of size k, e.g.

$$H(\boldsymbol{\sigma}) = \mathrm{NAE}(\sigma_1, \bar{\sigma}_3, \sigma_7) \wedge \mathrm{NAE}(\sigma_2, \bar{\sigma}_3, \bar{\sigma}_5) \wedge \mathrm{NAE}(\bar{\sigma}_1, \bar{\sigma}_2, \sigma_6) \in \{\mathtt{T}, \mathtt{F}\}$$

Gibbs measures: prototypical examples

Sherrington–Kirkpatrick model: for $\boldsymbol{\sigma} \in \{\pm 1\}^N$, $\boldsymbol{W} \sim \mathrm{GOE}(N)$:

$$H(\boldsymbol{\sigma}) = \frac{1}{2}(\boldsymbol{W}\boldsymbol{\sigma}, \boldsymbol{\sigma}) \qquad \text{Gibbs measure: } \mu_{\beta H}(\boldsymbol{\sigma}) = \frac{1}{Z} e^{\beta H(\boldsymbol{\sigma})}$$

Random k-**NAE-SAT**: AND of not-all-equal clauses of size k, e.g.

$$H(\boldsymbol{\sigma}) = \mathrm{NAE}(\sigma_1, \bar{\sigma}_3, \sigma_7) \wedge \mathrm{NAE}(\sigma_2, \bar{\sigma}_3, \bar{\sigma}_5) \wedge \mathrm{NAE}(\bar{\sigma}_1, \bar{\sigma}_2, \sigma_6) \in \{\mathtt{T}, \mathtt{F}\}$$

Model: sample M clauses, with each literal $\overset{\mathrm{IID}}{\sim} \mathrm{unif}(\sigma_1, \bar{\sigma}_1 \ldots, \sigma_N, \bar{\sigma}_N)$.

Gibbs measures: prototypical examples

Sherrington–Kirkpatrick model: for $\boldsymbol{\sigma} \in \{\pm 1\}^N$, $\boldsymbol{W} \sim \mathrm{GOE}(N)$:

$$H(\boldsymbol{\sigma}) = \frac{1}{2}(\boldsymbol{W}\boldsymbol{\sigma}, \boldsymbol{\sigma}) \qquad \text{Gibbs measure: } \mu_{\beta H}(\boldsymbol{\sigma}) = \frac{1}{Z} e^{\beta H(\boldsymbol{\sigma})}$$

Random k-**NAE-SAT**: AND of not-all-equal clauses of size k, e.g.

$$H(\boldsymbol{\sigma}) = \mathrm{NAE}(\sigma_1, \bar{\sigma}_3, \sigma_7) \wedge \mathrm{NAE}(\sigma_2, \bar{\sigma}_3, \bar{\sigma}_5) \wedge \mathrm{NAE}(\bar{\sigma}_1, \bar{\sigma}_2, \sigma_6) \in \{\mathtt{T}, \mathtt{F}\}$$

Model: sample M clauses, with each literal $\overset{\mathrm{IID}}{\sim} \mathrm{unif}(\sigma_1, \bar{\sigma}_1 \ldots, \sigma_N, \bar{\sigma}_N)$.

$$\text{Gibbs measure: } \mu_H = \mathrm{unif}(\text{sat assignments of } H)$$

Gibbs measures: prototypical examples

Sherrington–Kirkpatrick model: for $\boldsymbol{\sigma} \in \{\pm 1\}^N$, $\boldsymbol{W} \sim \mathrm{GOE}(N)$:

$$H(\boldsymbol{\sigma}) = \frac{1}{2}\langle \boldsymbol{W}\boldsymbol{\sigma}, \boldsymbol{\sigma}\rangle \qquad \text{Gibbs measure: } \mu_{\beta H}(\boldsymbol{\sigma}) = \frac{1}{Z} e^{\beta H(\boldsymbol{\sigma})}$$

Random k-NAE-SAT: AND of not-all-equal clauses of size k, e.g.

$$H(\boldsymbol{\sigma}) = \mathrm{NAE}(\sigma_1, \bar{\sigma}_3, \sigma_7) \wedge \mathrm{NAE}(\sigma_2, \bar{\sigma}_3, \bar{\sigma}_5) \wedge \mathrm{NAE}(\bar{\sigma}_1, \bar{\sigma}_2, \sigma_6) \in \{\mathtt{T}, \mathtt{F}\}$$

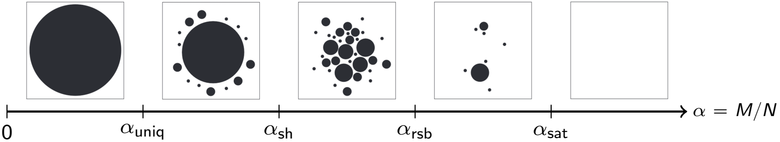
Model: sample M clauses, with each literal $\overset{\mathrm{IID}}{\sim} \mathrm{unif}(\sigma_1, \bar{\sigma}_1 \ldots, \sigma_N, \bar{\sigma}_N)$.

$$\text{Gibbs measure: } \mu_H = \mathrm{unif}(\text{sat assignments of } H)$$

Applications:

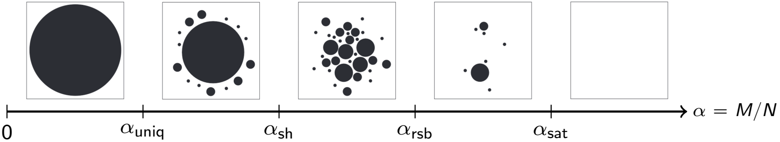
- Spin glasses: deep connections to free energy
- Bayesian inference: model of posteriors; sampling applications

Gibbs measures: (predicted) geometric phase transitions



(Image from Krzakala Montanari Ricci-Tersenghi Semerjian Zdeborová 06)

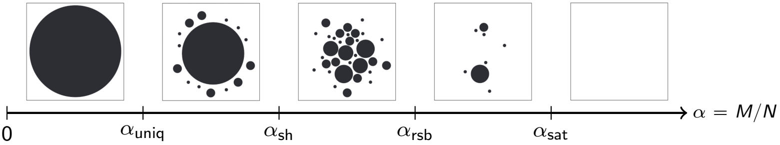
Gibbs measures: (predicted) geometric phase transitions



(Image from Krzakala Montanari Ricci-Tersenghi Semerjian Zdeborová 06)

- $\alpha \in (0, \alpha_{\text{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality

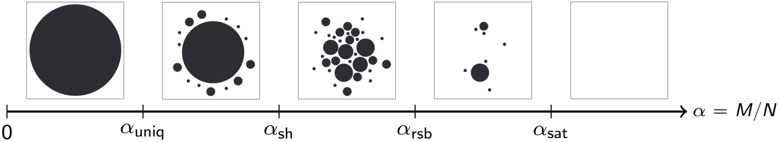
Gibbs measures: (predicted) geometric phase transitions



(Image from Krzakala Montanari Ricci-Tersenghi Semerjian Zdeborová 06)

- $\alpha \in (0, \alpha_{\text{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality
- $\alpha \in (\alpha_{\text{uniq}}, \alpha_{\text{sh}})$: rapid mixing from **random** but not worst-case start

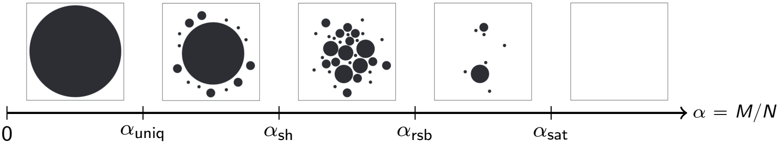
Gibbs measures: (predicted) geometric phase transitions



(Image from Krzakala Montanari Ricci-Tersenghi Semerjian Zdeborová 06)

- $\alpha \in (0, \alpha_{\text{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality
- $\alpha \in (\alpha_{\text{uniq}}, \alpha_{\text{sh}})$: rapid mixing from **random** but not worst-case start
- $\alpha \in (\alpha_{\text{sh}}, \alpha_{\text{rsb}})$: μ_H **shatters** into $e^{\Omega(N)}$ clusters of mass $e^{-\Omega(N)}$

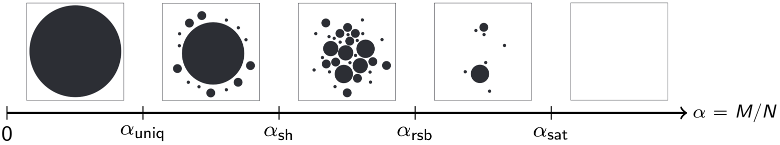
Gibbs measures: (predicted) geometric phase transitions



(Image from Krzakala Montanari Ricci-Tersenghi Semerjian Zdeborová 06)

- $\alpha \in (0, \alpha_{\text{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality
- $\alpha \in (\alpha_{\text{uniq}}, \alpha_{\text{sh}})$: rapid mixing from **random** but not worst-case start
- $\alpha \in (\alpha_{\text{sh}}, \alpha_{\text{rsb}})$: μ_H **shatters** into $e^{\Omega(N)}$ clusters of mass $e^{-\Omega(N)}$
- $\alpha \in (\alpha_{\text{rsb}}, \alpha_{\text{sat}})$: $e^{\Omega(N)}$ clusters but largest $O(1)$ dominate the mass (**condensation / RSB**)

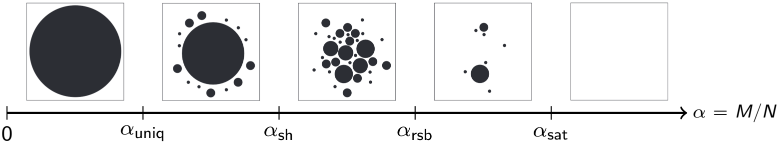
Gibbs measures: (predicted) geometric phase transitions



(Image from Krzakala Montanari Ricci-Tersenghi Semerjian Zdeborová 06)

- $\alpha \in (0, \alpha_{\text{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality
- $\alpha \in (\alpha_{\text{uniq}}, \alpha_{\text{sh}})$: rapid mixing from **random** but not worst-case start
- $\alpha \in (\alpha_{\text{sh}}, \alpha_{\text{rsb}})$: μ_H **shatters** into $e^{\Omega(N)}$ clusters of mass $e^{-\Omega(N)}$
- $\alpha \in (\alpha_{\text{rsb}}, \alpha_{\text{sat}})$: $e^{\Omega(N)}$ clusters but largest $O(1)$ dominate the mass (**condensation / RSB**)
- $\alpha \in (\alpha_{\text{sat}}, \infty)$: **unsat**, no solutions

Gibbs measures: (predicted) geometric phase transitions



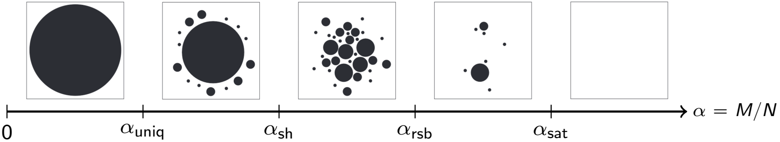
(Image from Krzakala Montanari Ricci-Tersenghi Semerjian Zdeborová 06)

- $\alpha \in (0, \alpha_{\text{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality
- $\alpha \in (\alpha_{\text{uniq}}, \alpha_{\text{sh}})$: rapid mixing from **random** but not worst-case start
- $\alpha \in (\alpha_{\text{sh}}, \alpha_{\text{rsb}})$: μ_H **shatters** into $e^{\Omega(N)}$ clusters of mass $e^{-\Omega(N)}$
- $\alpha \in (\alpha_{\text{rsb}}, \alpha_{\text{sat}})$: $e^{\Omega(N)}$ clusters but largest $O(1)$ dominate the mass (**condensation / RSB**)
- $\alpha \in (\alpha_{\text{sat}}, \infty)$: **unsat**, no solutions

---

Q: what does Gibbs measure μ_H look like around a typical $\boldsymbol{\sigma} \sim \mu_H$?

Gibbs measures: (predicted) geometric phase transitions



(Image from Krzakala Montanari Ricci-Tersenghi Semerjian Zdeborová 06)

- $\alpha \in (0, \alpha_{\text{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality
- $\alpha \in (\alpha_{\text{uniq}}, \alpha_{\text{sh}})$: rapid mixing from **random** but not worst-case start
- $\alpha \in (\alpha_{\text{sh}}, \alpha_{\text{rsb}})$: μ_H **shatters** into $e^{\Omega(N)}$ clusters of mass $e^{-\Omega(N)}$
- $\alpha \in (\alpha_{\text{rsb}}, \alpha_{\text{sat}})$: $e^{\Omega(N)}$ clusters but largest $O(1)$ dominate the mass (**condensation / RSB**)
- $\alpha \in (\alpha_{\text{sat}}, \infty)$: **unsat**, no solutions

---

Q: what does Gibbs measure μ_H look like around a typical $\boldsymbol{\sigma} \sim \mu_H$?
Challenge: $\boldsymbol{\sigma} \sim \mu_H$ not very explicit and hard to work with.

Planted model (constraint satisfaction setting)

Achlioptas Coja-Oghlan 08 had a beautiful idea. Consider joint distributions of $(H, \boldsymbol{\sigma})$:

Planted model (constraint satisfaction setting)

Achlioptas Coja-Oghlan 08 had a beautiful idea. Consider joint distributions of $(H, \boldsymbol{\sigma})$:

Null model:

- $H \sim \text{Law}(\text{random } k\text{-NAE-SAT})$
- $\boldsymbol{\sigma} \sim \text{unif}(\text{sat assignment of } H)$

Planted model (constraint satisfaction setting)

Achlioptas Coja-Oghlan 08 had a beautiful idea. Consider joint distributions of $(H, \boldsymbol{\sigma})$:

Null model:

- $H \sim \text{Law}(\text{random } k\text{-NAE-SAT})$
- $\boldsymbol{\sigma} \sim \text{unif}(\text{sat assignment of } H)$

Planted model:

- $\boldsymbol{\sigma} \sim \text{unif}(\{\pm 1\}^N)$

Planted model (constraint satisfaction setting)

Achlioptas Coja-Oghlan 08 had a beautiful idea. Consider joint distributions of $(H, \boldsymbol{\sigma})$:

Null model:

- $H \sim \text{Law}(\text{random } k\text{-NAE-SAT})$
- $\boldsymbol{\sigma} \sim \text{unif}(\text{sat assignment of } H)$

Planted model:

- $\boldsymbol{\sigma} \sim \text{unif}(\{\pm 1\}^N)$
- $H \sim \text{Law}(\text{random } k\text{-NAE-SAT} \mid \text{sat by } \boldsymbol{\sigma})$

Planted model (constraint satisfaction setting)

Achlioptas Coja-Oghlan 08 had a beautiful idea. Consider joint distributions of $(H, \boldsymbol{\sigma})$:

Null model:
- $H \sim \text{Law}(\text{random } k\text{-NAE-SAT})$
- $\boldsymbol{\sigma} \sim \text{unif}(\text{sat assignment of } H)$

Planted model:
- $\boldsymbol{\sigma} \sim \text{unif}(\{\pm 1\}^N)$
- $H \sim \text{Law}(\text{random } k\text{-NAE-SAT} \mid \text{sat by } \boldsymbol{\sigma})$

In certain regimes, planted / null models **contiguous**:

$$\mathbb{P}_{\text{planted}}(\boldsymbol{E}) = 1 - o(1) \quad \longleftrightarrow \quad \mathbb{P}_{\text{null}}(\boldsymbol{E}) = 1 - o(1) \qquad \forall \text{ event } \boldsymbol{E}$$

Planted model (constraint satisfaction setting)

Achlioptas Coja-Oghlan 08 had a beautiful idea. Consider joint distributions of $(H, \boldsymbol{\sigma})$:

Null model:

- $H \sim \mathrm{Law}(\text{random } k\text{-NAE-SAT})$
- $\boldsymbol{\sigma} \sim \mathrm{unif}(\text{sat assignment of } H)$

Planted model:

- $\boldsymbol{\sigma} \sim \mathrm{unif}(\{\pm 1\}^N)$
- $H \sim \mathrm{Law}(\text{random } k\text{-NAE-SAT} \mid \text{sat by } \boldsymbol{\sigma})$

In certain regimes, planted / null models **contiguous**:

$$\mathbb{P}_{\text{planted}}(\boldsymbol{E}) = 1 - o(1) \quad \longleftrightarrow \quad \mathbb{P}_{\text{null}}(\boldsymbol{E}) = 1 - o(1) \qquad \forall \text{ event } \boldsymbol{E}$$

If we want to prove a "whp" claim about the null model, equivalent to prove it in the planted model. This is often easier! (if contiguity holds)

Planted model (constraint satisfaction setting)

Achlioptas Coja-Oghlan 08 had a beautiful idea. Consider joint distributions of (H, σ):

Null model:
- $H \sim \text{Law}(\text{random } k\text{-NAE-SAT})$
- $\sigma \sim \text{unif}(\text{sat assignment of } H)$

(hard)

Planted model:
- $\sigma \sim \text{unif}(\{\pm 1\}^N)$
- $H \sim \text{Law}(\text{random } k\text{-NAE-SAT} \mid \text{sat by } \sigma)$

(easy)

In certain regimes, planted / null models **contiguous**:

$$\mathbb{P}_{\text{planted}}(\boldsymbol{E}) = 1 - o(1) \quad \longleftrightarrow \quad \mathbb{P}_{\text{null}}(\boldsymbol{E}) = 1 - o(1) \qquad \forall \text{ event } \boldsymbol{E}$$

If we want to prove a "whp" claim about the null model, equivalent to prove it in the planted model. This is often easier! (if contiguity holds)

Relation between planted and null models

$\sigma \in \{\pm 1\}^N$

H

	×	×		×	×		×		×		
			×	×		×				×	
×					×						×
		×	×					×			×
							×	×		×	
×	×					×			×		

× indicates that σ satisfies H

Relation between planted and null models

$\sigma \in \{\pm 1\}^N$

		×	×		×	×		×		×		
				×	×		×				×	
H	×					×						×
			×	×					×			×
								×	×		×	
	×	×					×			×		

× indicates that σ satisfies H

Null model: random row H, then
random × in that row

Relation between planted and null models

$\boldsymbol{\sigma} \in \{\pm 1\}^N$

H

	×	×		×	×		×		×		
			×	×		×				×	
×					×						×
		×	×					×			×
							×	×		×	
×	×					×			×		

× indicates that $\boldsymbol{\sigma}$ satisfies H

Null model: random row H, then
random × in that row

Planted model: random col $\boldsymbol{\sigma}$, then
random × in that col

Relation between planted and null models

$\boldsymbol{\sigma} \in \{\pm 1\}^N$

H

	×	×		×	×		×		×		
			×	×		×				×	
×					×						×
		×	×					×			×
							×	×		×	
×	×					×			×		

× indicates that $\boldsymbol{\sigma}$ satisfies H

Null model: random row H, then
random × in that row

Planted model: random col $\boldsymbol{\sigma}$, then
random × in that col

(all cols symmetric, but rows aren't)

Relation between planted and null models

$\sigma \in \{\pm 1\}^N$

H

	×	×		×	×		×		×		
			×	×		×				×	
×					×						×
		×	×					×			×
							×	×		×	
×	×					×			×		

× indicates that σ satisfies H

Null model: random row H, then random × in that row

Planted model: random col σ, then random × in that col

(all cols symmetric, but rows aren't)

Planted model weights H by $Z(H) = \#(\text{sat assignments of } H)$:

Relation between planted and null models

$\sigma \in \{\pm 1\}^N$

H

	×	×		×	×		×		×		
			×	×		×				×	
×					×						×
		×	×					×			×
							×	×		×	
×	×					×			×		

× indicates that σ satisfies H

Null model: random row H, then
random × in that row

Planted model: random col σ, then
random × in that col

(all cols symmetric, but rows aren't)

Planted model weights H by $Z(H) = \#(\text{sat assignments of } H)$:

$$\frac{d\mathbb{P}_{\text{planted}}}{d\mathbb{P}_{\text{null}}}(H, \sigma) = \frac{d\mathbb{P}_{\text{planted}}}{d\mathbb{P}_{\text{null}}}(H) = \frac{Z(H)}{\mathbb{E}Z(H)}$$

Relation between planted and null models

$\boldsymbol{\sigma} \in \{\pm 1\}^N$

H

	×	×		×	×		×		×		
			×	×		×				×	
×					×						×
		×	×					×			×
							×	×		×	
×	×					×			×		

× indicates that $\boldsymbol{\sigma}$ satisfies H

Null model: random row H, then
random × in that row

Planted model: random col $\boldsymbol{\sigma}$, then
random × in that col

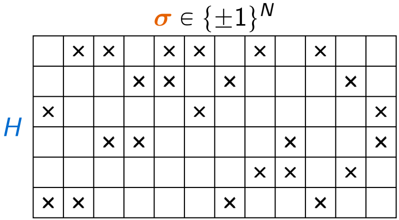
(all cols symmetric, but rows aren't)

Planted model weights H by $Z(H) = \#(\text{sat assignments of } H)$:

$$\frac{\mathrm{d}\mathbb{P}_{\text{planted}}}{\mathrm{d}\mathbb{P}_{\text{null}}}(H, \boldsymbol{\sigma}) = \frac{\mathrm{d}\mathbb{P}_{\text{planted}}}{\mathrm{d}\mathbb{P}_{\text{null}}}(H) = \frac{Z(H)}{\mathbb{E}Z(H)}$$

If this is $\Theta(1)$ whp, then planted / null models contiguous (Le Cam 60)

Relation between planted and null models



× indicates that $\boldsymbol{\sigma}$ satisfies H

Null model: random row H, then random × in that row

Planted model: random col $\boldsymbol{\sigma}$, then random × in that col

(all cols symmetric, but rows aren't)

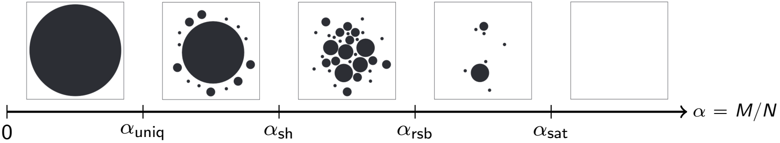
Planted model weights H by $Z(H) = \#(\text{sat assignments of } H)$:

$$\frac{d\mathbb{P}_{\text{planted}}}{d\mathbb{P}_{\text{null}}}(H, \boldsymbol{\sigma}) = \frac{d\mathbb{P}_{\text{planted}}}{d\mathbb{P}_{\text{null}}}(H) = \frac{Z(H)}{\mathbb{E}Z(H)}$$

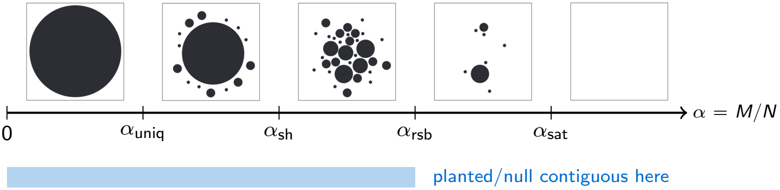
If this is $\Theta(1)$ whp, then planted / null models contiguous (Le Cam 60)

Holds for random k-NAE-SAT in **RS regime** $M/N < \alpha_{\text{rsb}}$

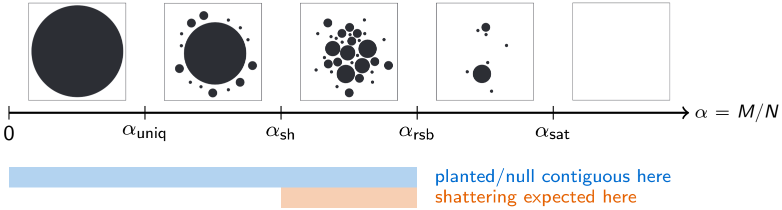
Application 1: shattering of random k-NAE-SAT



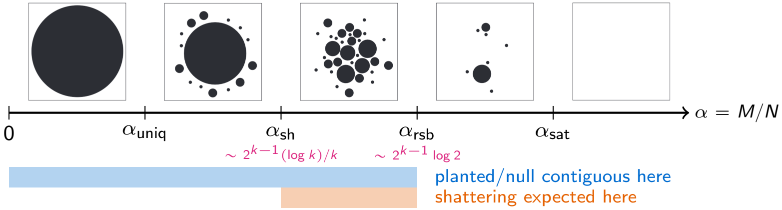
Application 1: shattering of random k-NAE-SAT



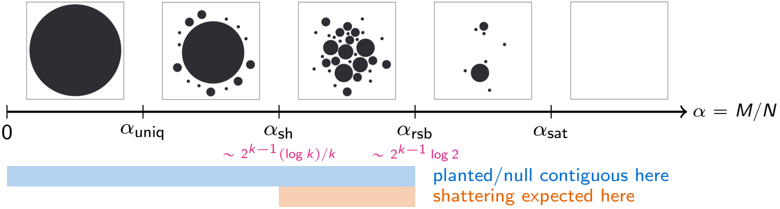
Application 1: shattering of random k-NAE-SAT



Application 1: shattering of random k-NAE-SAT



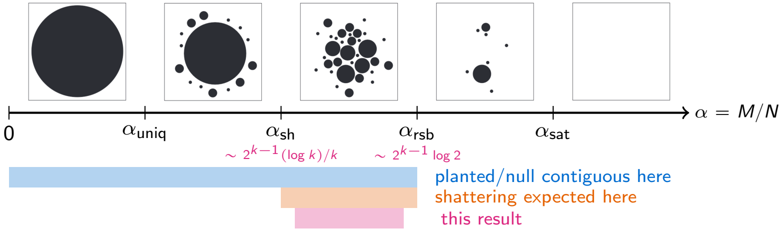
Application 1: shattering of random k-NAE-SAT



Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1 + o_k(1))\alpha_{\rm sh}, (1 - o_k(1))\alpha_{\rm rsb}]$,

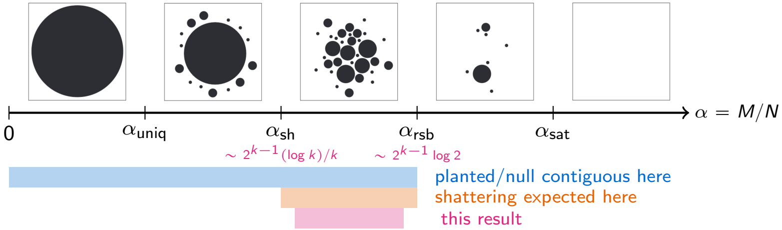
Application 1: shattering of random k-NAE-SAT



Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1 + o_k(1))\alpha_{\text{sh}}, (1 - o_k(1))\alpha_{\text{rsb}}]$,

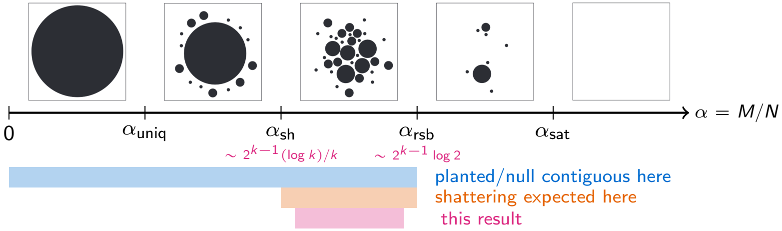
Application 1: shattering of random k-NAE-SAT



Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1 + o_k(1))\alpha_{\text{sh}}, (1 - o_k(1))\alpha_{\text{rsb}}]$,
whp over (k-*NAE-SAT instance* H, *Gibbs sample* $\boldsymbol{\sigma}$):

Application 1: shattering of random k-NAE-SAT

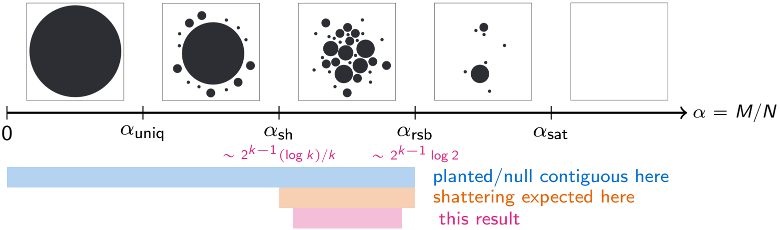


Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1 + o_k(1))\alpha_{\text{sh}}, (1 - o_k(1))\alpha_{\text{rsb}}]$, *whp over* ($k$*-NAE-SAT instance* H, *Gibbs sample* $\boldsymbol{\sigma}$)*:*

$$\#\Big\{ \text{sat assignments } \boldsymbol{\rho} \text{ of } H \text{ with } \underset{\substack{\uparrow \\ \text{Hamming dist}}}{\Delta}(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N] \Big\} = 0$$

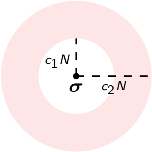
Application 1: shattering of random k-NAE-SAT



Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1 + o_k(1))\alpha_{\text{sh}}, (1 - o_k(1))\alpha_{\text{rsb}}]$*,*
whp over (k*-NAE-SAT instance* H*, Gibbs sample* $\boldsymbol{\sigma}$)*:*

$$\#\Big\{ \textit{sat assignments } \boldsymbol{\rho} \textit{ of } H \textit{ with } \underset{\substack{\uparrow \\ \textit{Hamming dist}}}{\Delta}(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N] \Big\} = 0$$



No sat assignments in ring around $\boldsymbol{\sigma}$ $\Rightarrow$ shattering

Application 1: shattering of random k-NAE-SAT

Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1+o_k(1))\alpha_{\mathsf{sh}}, (1-o_k(1))\alpha_{\mathsf{rsb}}]$, *whp over* $(k\text{-NAE-SAT instance } H, \text{Gibbs sample } \boldsymbol{\sigma})$:

$$\#\Big\{\text{sat assignments } \boldsymbol{\rho} \text{ of } H \text{ with } \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\} = 0$$

Application 1: shattering of random k-NAE-SAT

Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1+o_k(1))\alpha_{\mathsf{sh}}, (1-o_k(1))\alpha_{\mathsf{rsb}}]$*,*
whp over (k*-NAE-SAT instance* H, *Gibbs sample* $\boldsymbol{\sigma}$)*:*

$$\#\Big\{\textit{sat assignments } \boldsymbol{\rho} \textit{ of } H \textit{ with } \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\} = 0$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model.

Application 1: shattering of random k-NAE-SAT

Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1+o_k(1))\alpha_{\mathsf{sh}}, (1-o_k(1))\alpha_{\mathsf{rsb}}]$, *whp over* ($k$*-NAE-SAT instance* H, *Gibbs sample* $\boldsymbol{\sigma}$):

$$\#\Big\{\textit{sat assignments } \boldsymbol{\rho} \textit{ of } H \textit{ with } \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\} = 0$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model.

$H \sim \mathrm{Law}(\text{random } k\text{-NAE-SAT} \mid \text{sat by } \boldsymbol{\sigma})$ has **explicit description**:

Application 1: shattering of random k-NAE-SAT

Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1 + o_k(1))\alpha_{\mathsf{sh}}, (1 - o_k(1))\alpha_{\mathsf{rsb}}]$, *whp over* ($k$*-NAE-SAT instance* H, *Gibbs sample* $\boldsymbol{\sigma}$)*:*

$$\#\Big\{\textit{sat assignments } \boldsymbol{\rho} \textit{ of } H \textit{ with } \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\} = 0$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model.

$H \sim$ Law(random k-NAE-SAT | sat by $\boldsymbol{\sigma}$) has **explicit description**: clauses $\overset{IID}{\sim}$ Law(clause | sat by $\boldsymbol{\sigma}$).

Application 1: shattering of random k-NAE-SAT

Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1 + o_k(1))\alpha_{\mathsf{sh}}, (1 - o_k(1))\alpha_{\mathsf{rsb}}]$*,*
whp over (k*-NAE-SAT instance* H*, Gibbs sample* $\boldsymbol{\sigma}$*):*

$$\#\Big\{\textit{sat assignments } \boldsymbol{\rho} \textit{ of } H \textit{ with } \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\} = 0$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model.

$H \sim \mathrm{Law}(\text{random } k\text{-NAE-SAT} \mid \text{sat by } \boldsymbol{\sigma})$ has **explicit description**:
clauses $\overset{IID}{\sim} \mathrm{Law}(\text{clause} \mid \text{sat by } \boldsymbol{\sigma})$. Can calculate:

$$\mathbb{E}_{\mathsf{planted}} \#\Big\{\text{sat assignments } \boldsymbol{\rho} \text{ of } H \text{ with } \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\}$$

Application 1: shattering of random k-NAE-SAT

Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1+o_k(1))\alpha_{\mathrm{sh}}, (1-o_k(1))\alpha_{\mathrm{rsb}}]$,
whp over (k*-NAE-SAT instance* H, *Gibbs sample* $\boldsymbol{\sigma}$)*:*

$$\#\Big\{\textit{sat assignments } \boldsymbol{\rho} \textit{ of } H \textit{ with } \Delta(\boldsymbol{\sigma},\boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\} = 0$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model.

$H \sim \mathrm{Law}(\text{random } k\text{-NAE-SAT} \mid \text{sat by } \boldsymbol{\sigma})$ has **explicit description**: clauses $\overset{IID}{\sim}$ $\mathrm{Law}(\text{clause} \mid \text{sat by } \boldsymbol{\sigma})$. Can calculate:

$$\mathbb{E}_{\mathrm{planted}} \#\Big\{\text{sat assignments } \boldsymbol{\rho} \text{ of } H \text{ with } \Delta(\boldsymbol{\sigma},\boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\}$$
$$= \sum_{\boldsymbol{\rho} \in \{\pm 1\}^N : \Delta(\boldsymbol{\sigma},\boldsymbol{\rho}) \in [c_1 N, c_2 N]} \mathbb{P}_{\mathrm{planted}}(\boldsymbol{\rho} \text{ satisfies } H)$$

Application 1: shattering of random k-NAE-SAT

Theorem (Achlioptas Coja-Oghlan 08)

At constraint density $\alpha \in [(1 + o_k(1))\alpha_{\mathsf{sh}}, (1 - o_k(1))\alpha_{\mathsf{rsb}}]$, *whp over* ($k$*-NAE-SAT instance* H, *Gibbs sample* $\boldsymbol{\sigma}$)*:*

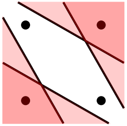
$$\#\Big\{\textit{sat assignments } \boldsymbol{\rho} \textit{ of } H \textit{ with } \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\} = 0$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model.

$H \sim \mathrm{Law}(\text{random } k\text{-NAE-SAT} \mid \text{sat by } \boldsymbol{\sigma})$ has **explicit description**: clauses $\overset{IID}{\sim} \mathrm{Law}(\text{clause} \mid \text{sat by } \boldsymbol{\sigma})$. Can calculate:

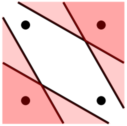
$$\mathbb{E}_{\text{planted}}\#\Big\{\text{sat assignments } \boldsymbol{\rho} \text{ of } H \text{ with } \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]\Big\}$$
$$= \sum_{\boldsymbol{\rho} \in \{\pm 1\}^N : \Delta(\boldsymbol{\sigma}, \boldsymbol{\rho}) \in [c_1 N, c_2 N]} \mathbb{P}_{\text{planted}}(\boldsymbol{\rho} \text{ satisfies } H) \ll 1$$

Application 2: symmetric Ising perceptron is frozen 1RSB



Symmetric Ising perceptron (Aubin Perkins Zdeborová 19):
intersection of discrete cube $\{\pm 1\}^N$ with IID symmetric slabs

Application 2: symmetric Ising perceptron is frozen 1RSB

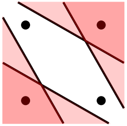


Symmetric Ising perceptron (Aubin Perkins Zdeborová 19): intersection of discrete cube $\{\pm 1\}^N$ with IID symmetric slabs

Formally: for $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M \overset{IID}{\sim} \mathcal{N}(0, I_N)$, $\kappa > 0$:

$$\boldsymbol{S} = \left\{ \boldsymbol{\sigma} \in \{\pm 1\}^N : |(\boldsymbol{g}^a, \boldsymbol{\sigma})| \leqslant \kappa\sqrt{N} \text{ for all } 1 \leqslant a \leqslant M \right\}$$

Application 2: symmetric Ising perceptron is frozen 1RSB



Symmetric Ising perceptron (Aubin Perkins Zdeborová 19): intersection of discrete cube $\{\pm 1\}^N$ with IID symmetric slabs

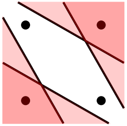
Formally: for $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M \overset{IID}{\sim} \mathcal{N}(0, I_N)$, $\kappa > 0$:

$$\boldsymbol{S} = \left\{ \boldsymbol{\sigma} \in \{\pm 1\}^N : |(\boldsymbol{g}^a, \boldsymbol{\sigma})| \leqslant \kappa\sqrt{N} \text{ for all } 1 \leqslant a \leqslant M \right\}$$

Theorem (Perkins Xu 21)

Whp over $\boldsymbol{G}$ *and* $\boldsymbol{\sigma} \sim \mathrm{unif}(\boldsymbol{S})$, $\boldsymbol{\sigma}$ *has* **Hamming distance $\Omega(N)$** *to all other elements of* $\boldsymbol{S}$ **(frozen 1RSB)**.

Application 2: symmetric Ising perceptron is frozen 1RSB



Symmetric Ising perceptron (Aubin Perkins Zdeborová 19): intersection of discrete cube $\{\pm 1\}^N$ with IID symmetric slabs

Formally: for $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M \overset{IID}{\sim} \mathcal{N}(0, I_N)$, $\kappa > 0$:

$$\boldsymbol{S} = \left\{ \boldsymbol{\sigma} \in \{\pm 1\}^N : |(\boldsymbol{g}^a, \boldsymbol{\sigma})| \leqslant \kappa\sqrt{N} \text{ for all } 1 \leqslant a \leqslant M \right\}$$

Theorem (Perkins Xu 21)

Whp over $\boldsymbol{G}$ and $\boldsymbol{\sigma} \sim \mathrm{unif}(\boldsymbol{S})$, $\boldsymbol{\sigma}$ has **Hamming distance $\boldsymbol{\Omega(N)}$** *to all other elements of $\boldsymbol{S}$* **(frozen 1RSB)**.

Planted model: $\boldsymbol{\sigma} \sim \mathrm{unif}(\{\pm 1\}^N)$, then sample IID $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M$ conditional on $|(\boldsymbol{g}^a, \boldsymbol{\sigma})| \leqslant \kappa\sqrt{N}$.

Application 3: shattering of pure spherical p-spin glass

Pure spherical p-spin model: for $g_{i_1,\ldots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$,

$$H(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1,\ldots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}$$

Application 3: shattering of pure spherical p-spin glass

Pure spherical p-spin model: for $g_{i_1,\dots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$,

$$H(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\dots,i_p=1}^{N} g_{i_1,\dots,i_p} \sigma_{i_1} \cdots \sigma_{i_p} \qquad \mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}\,\mathrm{d}\boldsymbol{\sigma}$$

on domain $\boldsymbol{\sigma} \in S_N = \sqrt{N}\mathbb{S}^{N-1}$

Application 3: shattering of pure spherical p-spin glass

Pure spherical p-spin model: for $g_{i_1,\dots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$,

$$H(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\dots,i_p=1}^{N} g_{i_1,\dots,i_p} \sigma_{i_1} \cdots \sigma_{i_p} \qquad \mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}\,\mathrm{d}\boldsymbol{\sigma}$$

on domain $\boldsymbol{\sigma} \in S_N = \sqrt{N}\mathbb{S}^{N-1}$

Planted model: $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$, then reweight probability density of H:

$$\frac{\rho_{\mathrm{pl}}(H \mid \boldsymbol{\sigma})}{\rho_{\mathrm{null}}(H)} \propto e^{\beta H(\boldsymbol{\sigma})}$$

Application 3: shattering of pure spherical p-spin glass

Pure spherical p-spin model: for $g_{i_1,\dots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$,

$$H(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\dots,i_p=1}^{N} g_{i_1,\dots,i_p} \sigma_{i_1} \cdots \sigma_{i_p} \qquad \mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}\,\mathrm{d}\boldsymbol{\sigma}$$

on domain $\boldsymbol{\sigma} \in S_N = \sqrt{N}\mathbb{S}^{N-1}$

Planted model: $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$, then reweight probability density of H:

$$\frac{\rho_{\mathrm{pl}}(H \mid \boldsymbol{\sigma})}{\rho_{\mathrm{null}}(H)} \propto e^{\beta H(\boldsymbol{\sigma})}$$

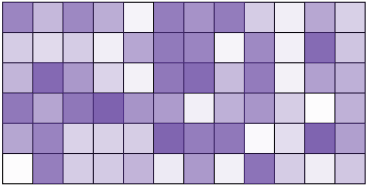
Equivalently: plant a spike

$$H(\boldsymbol{\rho}) = H_{\mathrm{null}}(\boldsymbol{\rho}) + N\beta R(\boldsymbol{\sigma}, \boldsymbol{\rho})^p \qquad R(\boldsymbol{\sigma}, \boldsymbol{\rho}) = \frac{(\boldsymbol{\sigma}, \boldsymbol{\rho})}{N}$$

Application 3: shattering of pure spherical p-spin glass

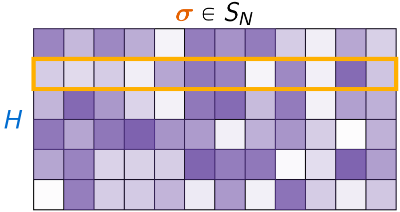
$\sigma \in S_N$

H



Heatmap indicates value of $e^{\beta H(\sigma)}$

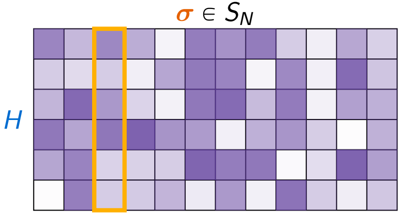
Application 3: shattering of pure spherical p-spin glass



Heatmap indicates value of $e^{\beta H(\sigma)}$

Null model: random row H, then col σ with prob $\propto e^{\beta H(\sigma)}$

Application 3: shattering of pure spherical p-spin glass

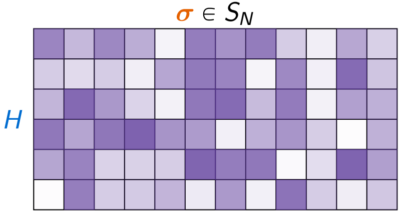


Heatmap indicates value of $e^{\beta H(\boldsymbol{\sigma})}$

Null model: random row H, then col $\boldsymbol{\sigma}$ with prob $\propto e^{\beta H(\boldsymbol{\sigma})}$

Planted model: random col $\boldsymbol{\sigma}$, then row H with prob $\propto e^{\beta H(\boldsymbol{\sigma})}$

Application 3: shattering of pure spherical p-spin glass



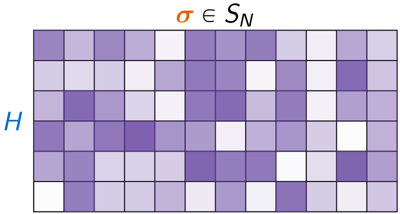
Heatmap indicates value of $e^{\beta H(\sigma)}$

Null model: random row H, then col σ with prob $\propto e^{\beta H(\sigma)}$

Planted model: random col σ, then row H with prob $\propto e^{\beta H(\sigma)}$

(all cols symmetric, but rows aren't)

Application 3: shattering of pure spherical p-spin glass



Heatmap indicates value of $e^{\beta H(\boldsymbol{\sigma})}$

Null model: random row H, then col $\boldsymbol{\sigma}$ with prob $\propto e^{\beta H(\boldsymbol{\sigma})}$

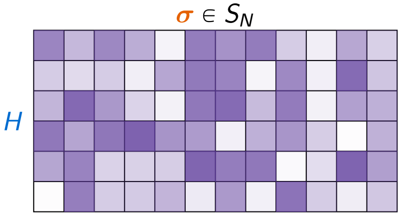
Planted model: random col $\boldsymbol{\sigma}$, then row H with prob $\propto e^{\beta H(\boldsymbol{\sigma})}$

(all cols symmetric, but rows aren't)

Planted model weights H by partition function ($\leftrightarrow$ row sum)

$$Z_\beta(H) = \int_{S_N} e^{\beta H(\boldsymbol{\sigma})} \, \mathrm{d}\boldsymbol{\sigma}$$

Application 3: shattering of pure spherical p-spin glass



Heatmap indicates value of $e^{\beta H(\boldsymbol{\sigma})}$

Null model: random row H, then col $\boldsymbol{\sigma}$ with prob $\propto e^{\beta H(\boldsymbol{\sigma})}$

Planted model: random col $\boldsymbol{\sigma}$, then row H with prob $\propto e^{\beta H(\boldsymbol{\sigma})}$

(all cols symmetric, but rows aren't)

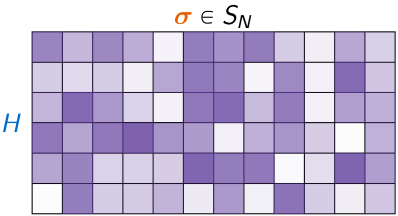
Planted model weights H by partition function ($\leftrightarrow$ row sum)

$$Z_\beta(H) = \int_{S_N} e^{\beta H(\boldsymbol{\sigma})}\, \mathrm{d}\boldsymbol{\sigma}$$

That is, (similarly to before)

$$\frac{\mathrm{d}\mathbb{P}_{\text{planted}}}{\mathrm{d}\mathbb{P}_{\text{null}}}(H, \boldsymbol{\sigma}) = \frac{\mathrm{d}\mathbb{P}_{\text{planted}}}{\mathrm{d}\mathbb{P}_{\text{null}}}(H) = \frac{Z_\beta(H)}{\mathbb{E} Z_\beta(H)}$$

Application 3: shattering of pure spherical p-spin glass



Heatmap indicates value of $e^{\beta H(\sigma)}$

Null model: random row H, then col σ with prob $\propto e^{\beta H(\sigma)}$

Planted model: random col σ, then row H with prob $\propto e^{\beta H(\sigma)}$

(all cols symmetric, but rows aren't)

Planted model weights H by partition function ($\leftrightarrow$ row sum)

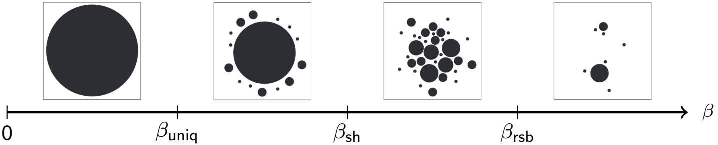
$$Z_\beta(H) = \int_{S_N} e^{\beta H(\sigma)} \, \mathrm{d}\sigma$$

That is, (similarly to before)

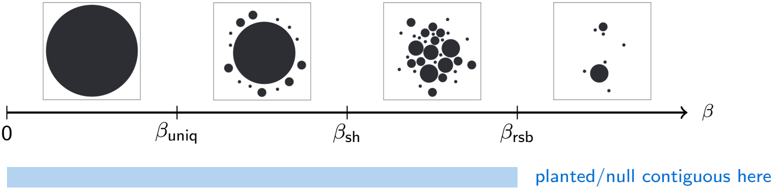
$$\frac{\mathrm{d}\mathbb{P}_{\text{planted}}}{\mathrm{d}\mathbb{P}_{\text{null}}}(H, \sigma) = \frac{\mathrm{d}\mathbb{P}_{\text{planted}}}{\mathrm{d}\mathbb{P}_{\text{null}}}(H) = \frac{Z_\beta(H)}{\mathbb{E} Z_\beta(H)}$$

Planted/null contiguous if this is $\Theta(1)$ whp. Holds for $\beta < \beta_{\text{rsb}}$.

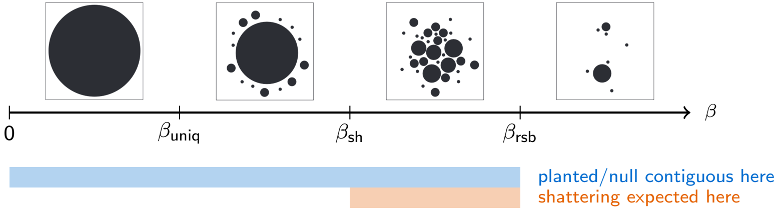
Application 3: shattering of pure spherical p-spin glass



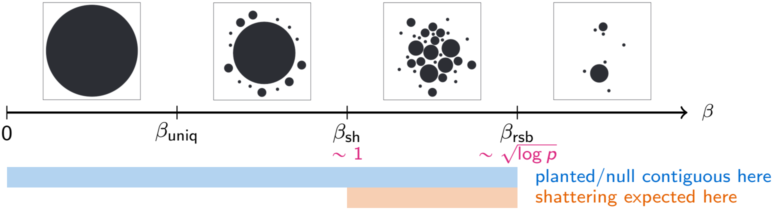
Application 3: shattering of pure spherical p-spin glass



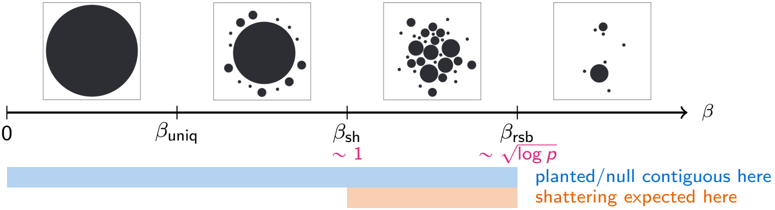
Application 3: shattering of pure spherical p-spin glass



Application 3: shattering of pure spherical p-spin glass



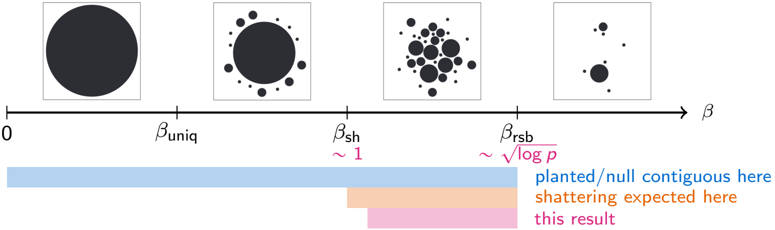
Application 3: shattering of pure spherical p-spin glass



Theorem (El Alaoui Montanari Sellke 23)

For $\beta \in [\beta_{\text{sh}} \cdot O(1), \beta_{\text{rsb}})$,

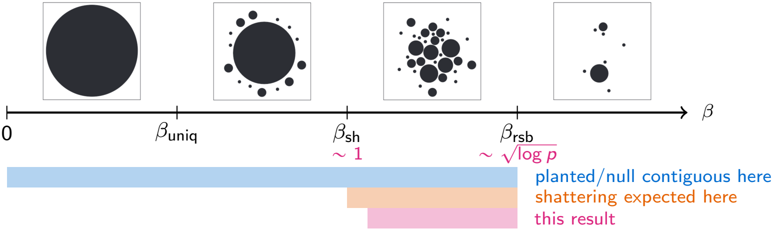
Application 3: shattering of pure spherical p-spin glass



Theorem (El Alaoui Montanari Sellke 23)

For $\beta \in [\beta_{\text{sh}} \cdot O(1), \beta_{\text{rsb}})$,

Application 3: shattering of pure spherical p-spin glass

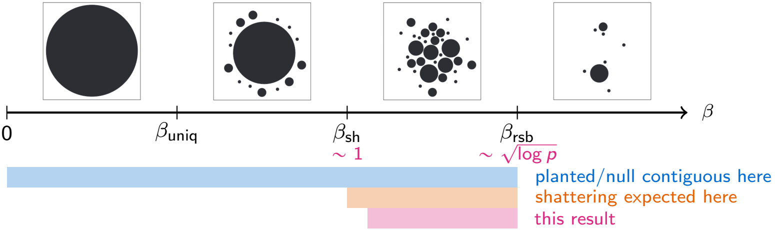


Theorem (El Alaoui Montanari Sellke 23)

For $\beta \in [\beta_{\text{sh}} \cdot O(1), \beta_{\text{rsb}})$, whp over $(H$, Gibbs sample $\boldsymbol{\sigma})$:

$$\int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) \in [c_1, c_2]\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho} \leqslant e^{-\Omega(N)} \cdot \int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) > c_2\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho}$$

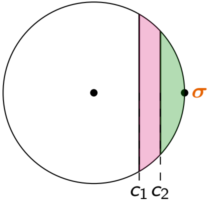
Application 3: shattering of pure spherical p-spin glass



Theorem (El Alaoui Montanari Sellke 23)

For $\beta \in [\beta_{\mathrm{sh}} \cdot O(1), \beta_{\mathrm{rsb}})$, whp over ($H$, Gibbs sample $\boldsymbol{\sigma}$):

$$\int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) \in [c_1, c_2]\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho} \leqslant e^{-\Omega(N)} \cdot \int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) > c_2\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho}$$



$\int e^{\beta H(\boldsymbol{\rho})}$ much larger in green region than red region $\Rightarrow$ shattering

Application 3: shattering of pure p-spin glasses

Theorem (El Alaoui Montanari Sellke 23)

For $\beta \in [\beta_{\mathsf{sh}} \cdot O(1), \beta_{\mathsf{rsb}})$, *whp over* $(H, \text{Gibbs sample } \boldsymbol{\sigma})$:

$$\int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) \in [c_1, c_2]\Big\} e^{\beta H(\boldsymbol{\rho})} \, \mathrm{d}\boldsymbol{\rho} \leqslant e^{-\Omega(N)} \cdot \int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) > c_2\Big\} e^{\beta H(\boldsymbol{\rho})} \, \mathrm{d}\boldsymbol{\rho}$$

Application 3: shattering of pure p-spin glasses

Theorem (El Alaoui Montanari Sellke 23)

For $\beta \in [\beta_{\mathsf{sh}} \cdot O(1), \beta_{\mathsf{rsb}})$, whp over ($H$, Gibbs sample $\boldsymbol{\sigma}$):

$$\int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho},\boldsymbol{\sigma}) \in [c_1, c_2]\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho} \leqslant e^{-\Omega(N)} \cdot \int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho},\boldsymbol{\sigma}) > c_2\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho}$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model:

Application 3: shattering of pure p-spin glasses

Theorem (El Alaoui Montanari Sellke 23)

For $\beta \in [\beta_{\mathrm{sh}} \cdot O(1), \beta_{\mathrm{rsb}})$, whp over ($H$, Gibbs sample $\boldsymbol{\sigma}$):

$$\int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) \in [c_1, c_2]\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho} \leqslant e^{-\Omega(N)} \cdot \int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) > c_2\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho}$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model:

$$\boldsymbol{\sigma} \sim \mathrm{unif}(S_N) \qquad H(\boldsymbol{\rho}) = H_{\mathrm{null}}(\boldsymbol{\rho}) + N\beta R(\boldsymbol{\sigma}, \boldsymbol{\rho})^p$$

This is an explicit **spiked** model.

Application 3: shattering of pure p-spin glasses

Theorem (El Alaoui Montanari Sellke 23)

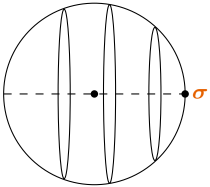
For $\beta \in [\beta_{\mathsf{sh}} \cdot O(1), \beta_{\mathsf{rsb}})$, whp over $(H$, Gibbs sample $\boldsymbol{\sigma})$:

$$\int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) \in [c_1, c_2]\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho} \leqslant e^{-\Omega(N)} \cdot \int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) > c_2\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho}$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model:

$$\boldsymbol{\sigma} \sim \mathrm{unif}(S_N) \qquad H(\boldsymbol{\rho}) = H_{\mathsf{null}}(\boldsymbol{\rho}) + N\beta R(\boldsymbol{\sigma}, \boldsymbol{\rho})^p$$

This is an explicit **spiked** model.



Application 3: shattering of pure p-spin glasses

Theorem (El Alaoui Montanari Sellke 23)

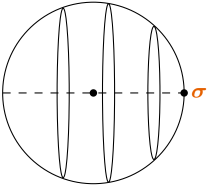
For $\beta \in [\beta_{\mathsf{sh}} \cdot O(1), \beta_{\mathsf{rsb}})$, whp over $(H$, Gibbs sample $\boldsymbol{\sigma})$:

$$\int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) \in [c_1, c_2]\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho} \leqslant e^{-\Omega(N)} \cdot \int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho}, \boldsymbol{\sigma}) > c_2\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho}$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model:

$$\boldsymbol{\sigma} \sim \mathsf{unif}(S_N) \qquad H(\boldsymbol{\rho}) = H_{\mathsf{null}}(\boldsymbol{\rho}) + N\beta R(\boldsymbol{\sigma}, \boldsymbol{\rho})^p$$

This is an explicit **spiked** model.



On each orthogonal band to $\boldsymbol{\sigma}$,
tools are known to estimate $\int e^{\beta H(\boldsymbol{\rho})}$

Application 3: shattering of pure p-spin glasses

Theorem (El Alaoui Montanari Sellke 23)

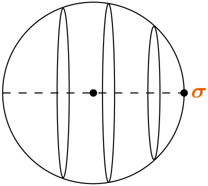
For $\beta \in [\beta_{\mathrm{sh}} \cdot O(1), \beta_{\mathrm{rsb}})$, whp over ($H$, Gibbs sample $\boldsymbol{\sigma}$):

$$\int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho},\boldsymbol{\sigma}) \in [c_1, c_2]\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho} \leqslant e^{-\Omega(N)} \cdot \int_{S_N} \mathbf{1}\Big\{R(\boldsymbol{\rho},\boldsymbol{\sigma}) > c_2\Big\} e^{\beta H(\boldsymbol{\rho})}\, \mathrm{d}\boldsymbol{\rho}$$

Proof: by contiguity, suffices to prove for $(H, \boldsymbol{\sigma}) \sim$ planted model:

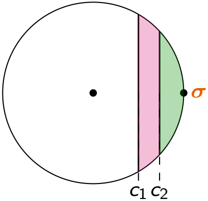
$$\boldsymbol{\sigma} \sim \mathrm{unif}(S_N) \qquad H(\boldsymbol{\rho}) = H_{\mathrm{null}}(\boldsymbol{\rho}) + N\beta R(\boldsymbol{\sigma},\boldsymbol{\rho})^p$$

This is an explicit **spiked** model.



On each orthogonal band to $\boldsymbol{\sigma}$,
tools are known to estimate $\int e^{\beta H(\boldsymbol{\rho})}$

these show green region $\gg$ red region



Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model
- $\boldsymbol{\sigma} \sim \mu$ from $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model
- $\boldsymbol{\sigma} \sim \mu$ from $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model
- $\boldsymbol{\sigma} \sim \mu$ from $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Nature reveals $(H, \boldsymbol{y})$. Goal: estimate $\mathbb{E}[\boldsymbol{\sigma} \mid H, \boldsymbol{y}]$.

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model
- $\boldsymbol{\sigma} \sim \mu$ from $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Planted generative process:

- $\boldsymbol{\sigma} \sim \text{unif}(S_N)$

Nature reveals $(H, \boldsymbol{y})$. Goal: estimate $\mathbb{E}[\boldsymbol{\sigma} \mid H, \boldsymbol{y}]$.

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model
- $\boldsymbol{\sigma} \sim \mu$ from $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Planted generative process:

- $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- $H \sim \mathrm{Law}(\text{planted model} \mid \boldsymbol{\sigma})$

$\leftrightarrow H(\boldsymbol{\rho}) = H_{\mathrm{null}}(\boldsymbol{\rho}) + NR(\boldsymbol{\rho}, \boldsymbol{\sigma})^p$

Nature reveals $(H, \boldsymbol{y})$. Goal: estimate $\mathbb{E}[\boldsymbol{\sigma} \mid H, \boldsymbol{y}]$.

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model
- $\boldsymbol{\sigma} \sim \mu$ from $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Planted generative process:

- $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- $H \sim$ Law(planted model $\mid \boldsymbol{\sigma}$)
$\leftrightarrow H(\boldsymbol{\rho}) = H_{\mathrm{null}}(\boldsymbol{\rho}) + NR(\boldsymbol{\rho}, \boldsymbol{\sigma})^p$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Nature reveals $(H, \boldsymbol{y})$. Goal: estimate $\mathbb{E}[\boldsymbol{\sigma} \mid H, \boldsymbol{y}]$.

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model
- $\boldsymbol{\sigma} \sim \mu$ from $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Planted generative process:

- $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- $H \sim \mathrm{Law}(\text{planted model} \mid \boldsymbol{\sigma})$
$\leftrightarrow H(\boldsymbol{\rho}) = H_{\mathrm{null}}(\boldsymbol{\rho}) + NR(\boldsymbol{\rho}, \boldsymbol{\sigma})^p$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Nature reveals $(H, \boldsymbol{y})$. Goal: estimate $\mathbb{E}[\boldsymbol{\sigma} \mid H, \boldsymbol{y}]$.

Tractable in planted model: $H, \boldsymbol{y}$ are independent gaussian-channel observations of $\boldsymbol{\sigma}$

Application 4: diffusion sampling from spin glasses

Main subroutine: mean ($\leftrightarrow$ score) estimation of following model

Generative process:

- $H \sim$ null model
- $\boldsymbol{\sigma} \sim \mu$ from $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Planted generative process:

- $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- $H \sim \mathrm{Law}(\text{planted model} \mid \boldsymbol{\sigma})$
$\leftrightarrow H(\boldsymbol{\rho}) = H_{\mathrm{null}}(\boldsymbol{\rho}) + NR(\boldsymbol{\rho}, \boldsymbol{\sigma})^p$
- $\boldsymbol{y} = t\boldsymbol{\sigma} + \sqrt{t}\boldsymbol{g}$, $\boldsymbol{g} \sim \mathcal{N}(0, I_N)$

Nature reveals $(H, \boldsymbol{y})$. Goal: estimate $\mathbb{E}[\boldsymbol{\sigma} \mid H, \boldsymbol{y}]$.

Tractable in planted model: $H, \boldsymbol{y}$ are independent gaussian-channel observations of $\boldsymbol{\sigma}$

El Alaoui Montanari Sellke 22+23, H Montanari Pham 24 use this to sample from Gibbs measure $\mu(\boldsymbol{\sigma}) \propto e^{H(\boldsymbol{\sigma})}$.

Other applications of planting

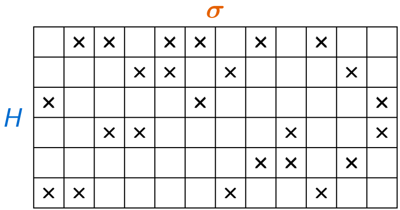
- Coja-Oghlan Krzakala Perkins Zdeborová 16
- Coja-Oghlan Efthymiou Jaafari Kang Kapetanopoulos 17
- Coja-Oghlan Kapetanopoulos Müller 18

} RS free energy of CSPs

- H Sellke 23: 2nd moment proof of RS free energy in spherical spin glasses

- Mossel Sly Sohn 24: sharp weak recovery threshold of sparse SBM

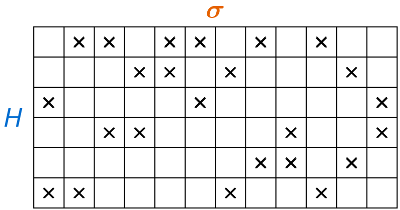
Classic planting requires centeredness + RS



Recall: planted model weights H by partition function $Z(H)$

$\frac{\mathrm{d}\mathbb{P}_{\mathrm{pl}}}{\mathrm{d}\mathbb{P}_{\mathrm{null}}} = \frac{Z(H)}{\mathbb{E}Z(H)}$, contiguous if this $\Theta(1)$ whp

Classic planting requires centeredness + RS

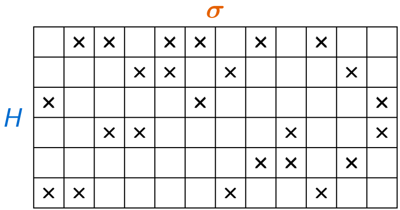


Recall: planted model weights H by partition function $Z(H)$

$$\frac{\mathrm{d}\mathbb{P}_{\mathrm{pl}}}{\mathrm{d}\mathbb{P}_{\mathrm{null}}} = \frac{Z(H)}{\mathbb{E}Z(H)}$$, contiguous if this $\Theta(1)$ whp

This proof strategy **requires contiguity**. Models where this holds:

Classic planting requires centeredness + RS



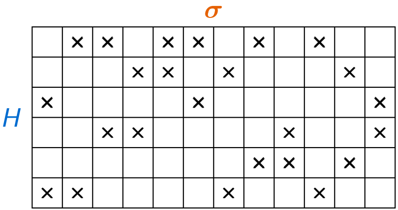
Recall: planted model weights H by partition function $Z(H)$

$$\frac{d\mathbb{P}_{\text{pl}}}{d\mathbb{P}_{\text{null}}} = \frac{Z(H)}{\mathbb{E}Z(H)},$$ contiguous if this $\Theta(1)$ whp

This proof strategy **requires contiguity**. Models where this holds:

- random NAE-SAT or XOR-SAT, but not (immediately) SAT

Classic planting requires centeredness + RS



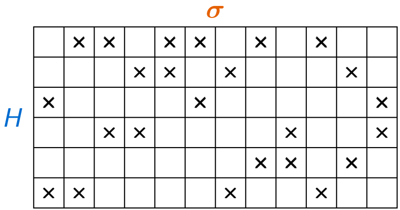
Recall: planted model weights H by partition function $Z(H)$

$$\frac{\mathrm{d}\mathbb{P}_{\mathrm{pl}}}{\mathrm{d}\mathbb{P}_{\mathrm{null}}} = \frac{Z(H)}{\mathbb{E}Z(H)},$$ contiguous if this $\Theta(1)$ whp

This proof strategy **requires contiguity**. Models where this holds:

- random NAE-SAT or XOR-SAT, but not (immediately) SAT
- symmetric perceptron, but not asymmetric perceptron

Classic planting requires centeredness + RS



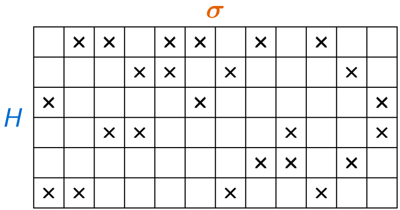
Recall: planted model weights H by partition function $Z(H)$

$$\frac{d\mathbb{P}_{\mathrm{pl}}}{d\mathbb{P}_{\mathrm{null}}} = \frac{Z(H)}{\mathbb{E}Z(H)}$$, contiguous if this $\Theta(1)$ whp

This proof strategy **requires contiguity**. Models where this holds:

- random NAE-SAT or XOR-SAT, but not (immediately) SAT
- symmetric perceptron, but not asymmetric perceptron
- (mixed) p-spin glass, but only without external field

Classic planting requires centeredness + RS



Recall: planted model weights H by partition function $Z(H)$

$$\frac{d\mathbb{P}_{\mathrm{pl}}}{d\mathbb{P}_{\mathrm{null}}} = \frac{Z(H)}{\mathbb{E}Z(H)},$$ contiguous if this $\Theta(1)$ whp

This proof strategy **requires contiguity**. Models where this holds:

- random NAE-SAT or XOR-SAT, but not (immediately) SAT
- symmetric perceptron, but not asymmetric perceptron
- (mixed) p-spin glass, but only without external field

Also need **free energy** $\approx$ **annealed free energy**: $\log Z = \log \mathbb{E}Z + O(1)$

We can plant things other than Gibbs samples!

We can plant things other than Gibbs samples!

Rest of this half: two applications that each reduce to analyzing a planted model

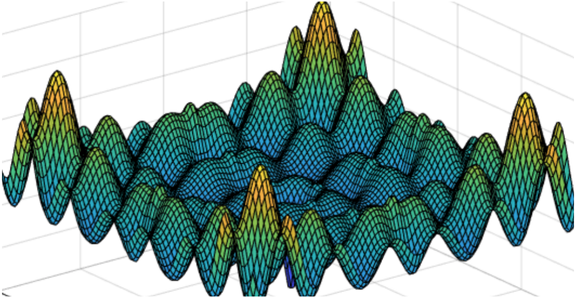
Outline of part I: applications of planting in disordered models

The classic planting trick: planting a Gibbs sample

Ground state large deviations in spherical spin glasses

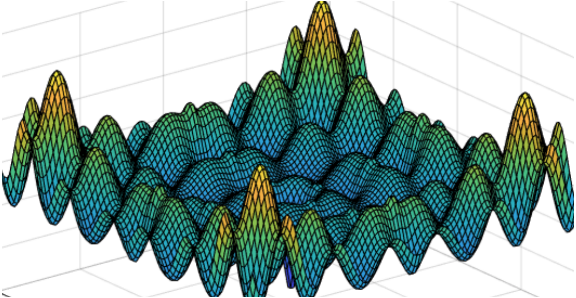
TAP planting: capacity of the Ising perceptron

How complicated is a random landscape?



Fyodorov 04, Auffinger Ben Arous Černý 13: study via # critical points,

How complicated is a random landscape?



Fyodorov 04, Auffinger Ben Arous Černý 13: study via $\#$ critical points, using the **Kac–Rice formula** to calculate quantities like $\mathbb{E}[\#\,\text{crit pts}]$

Landscape complexity

Huge amount of work studying wide range of models:

• Subag 17, Ben Arous Subag Zeitouni 20, Belius Černý Nakajima Schmidt 22: spherical spin glasses

• Sagun Güney Ben Arous LeCun 14: neural networks

• Ben Arous Mei Montanari Nica 17: spiked tensor model

• Fyodorov 16, Ben Arous Fyodorov Khoruzhenko 21, Subag 23, Kivimae 24: non gradient vector fields

• Maillard Ben Arous Biroli 20: generalized linear models

• Fan Mei Montanari 21: TAP free energy in $\mathbb{Z}_2$-synchronization

• Ben Arous Bourgade McKenna 24: elastic manifold

• Kivimae 23, McKenna 24, H Sellke 25 : bipartite / multi-species spherical spin glasses

Ground state of pure spin glasses, via complexity

Auffinger Ben Arous Černý 13: crit pt complexity of pure p-spin model

$$H(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1,\ldots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}\,, \qquad g_{i_1,\ldots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$$

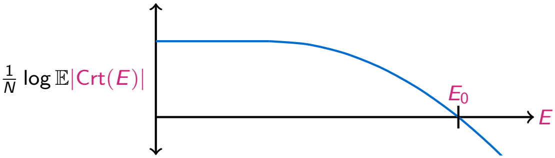
on $S_N = \sqrt{N}\mathbb{S}^{N-1}$.

Ground state of pure spin glasses, via complexity

Auffinger Ben Arous Černý 13: crit pt complexity of pure p-spin model

$$H(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1,\ldots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}, \qquad g_{i_1,\ldots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$$

on $S_N = \sqrt{N}\mathbb{S}^{N-1}$. Calculate $\mathbb{E}|\mathrm{Crt}(E)| \equiv \mathbb{E}|\{\text{crits with } H(\boldsymbol{\sigma})/N \geqslant E\}|$:

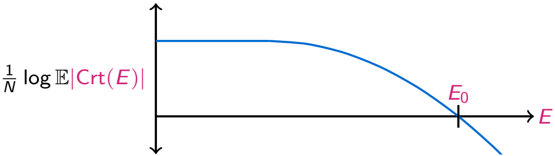


Ground state of pure spin glasses, via complexity

Auffinger Ben Arous Černý 13: crit pt complexity of pure p-spin model

$$H(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1,\ldots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}\,, \qquad g_{i_1,\ldots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$$

on $S_N = \sqrt{N}\mathbb{S}^{N-1}$. Calculate $\mathbb{E}|\mathrm{Crt}(E)| \equiv \mathbb{E}|\{\text{crits with } H(\boldsymbol{\sigma})/N \geqslant E\}|$:



Markov $\Rightarrow$ whp no crit pts with energy $> E_0$. Implies ground state UB:

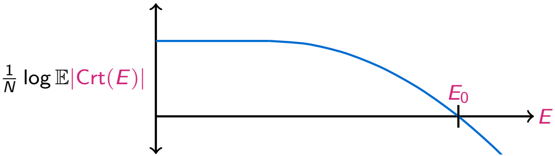
$$\mathrm{GS}_N \equiv \max_{\boldsymbol{\sigma}\in S_N} \frac{H(\boldsymbol{\sigma})}{N} \leqslant E_0 \qquad \text{whp}$$

Ground state of pure spin glasses, via complexity

Auffinger Ben Arous Černý 13: crit pt complexity of pure p-spin model

$$H(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1,\ldots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}\,, \qquad g_{i_1,\ldots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$$

on $S_N = \sqrt{N}\mathbb{S}^{N-1}$. Calculate $\mathbb{E}|\mathrm{Crt}(E)| \equiv \mathbb{E}|\{\text{crits with } H(\boldsymbol{\sigma})/N \geqslant E\}|$:



Markov $\Rightarrow$ whp no crit pts with energy $> E_0$. Implies ground state UB:

$$\mathrm{GS}_N \equiv \max_{\boldsymbol{\sigma}\in S_N} \frac{H(\boldsymbol{\sigma})}{N} \leqslant E_0 \qquad \text{whp}$$

This is sharp! E_0 matches ground state given by **Parisi formula**.

Related works on landscape complexity and ground state

- Subag 17: $\mathrm{GS}_N \geqslant E_0$ via 2nd moment analysis of $|\mathrm{Crt}(E)|$.

Thus $\mathrm{GS}_N \xrightarrow{p} E_0$, locating ground state independently of Parisi formula

Related works on landscape complexity and ground state

- Subag 17: $\mathrm{GS}_N \geqslant E_0$ via 2nd moment analysis of $|\mathrm{Crt}(E)|$.

Thus $\mathrm{GS}_N \xrightarrow{p} E_0$, locating ground state independently of Parisi formula

- Ben Arous Subag Zeitouni 20: similarly, $\mathrm{GS}_N \xrightarrow{p} E_0$ in some regime of mixed p-spin models (+ results on Gibbs measure geometry)

Related works on landscape complexity and ground state

- Subag 17: $\mathrm{GS}_N \geqslant E_0$ via 2nd moment analysis of $|\mathrm{Crt}(E)|$.
Thus $\mathrm{GS}_N \xrightarrow{p} E_0$, locating ground state independently of Parisi formula

- Ben Arous Subag Zeitouni 20: similarly, $\mathrm{GS}_N \xrightarrow{p} E_0$ in some regime of mixed p-spin models (+ results on Gibbs measure geometry)

- Subag Zeitouni 17: extremal process of crit pts in pure models

- Subag Zeitouni 21: asymptotics of $|\mathrm{Crt}(E)|$ at arbitrary E

- Belius Schmidt 22: asymptotics of # crit pts at given radial derivative

Related works on landscape complexity and ground state

- Subag 17: $\mathrm{GS}_N \geqslant E_0$ via 2nd moment analysis of $|\mathrm{Crt}(E)|$.
Thus $\mathrm{GS}_N \xrightarrow{p} E_0$, locating ground state independently of Parisi formula

- Ben Arous Subag Zeitouni 20: similarly, $\mathrm{GS}_N \xrightarrow{p} E_0$ in some regime of mixed p-spin models (+ results on Gibbs measure geometry)

- Subag Zeitouni 17: extremal process of crit pts in pure models

- Subag Zeitouni 21: asymptotics of $|\mathrm{Crt}(E)|$ at arbitrary E

- Belius Schmidt 22: asymptotics of # crit pts at given radial derivative

- Subag 17, Gheissari Jagannath 19, Ben Arous Jagannath 24: geometry of Gibbs measures in pure models

- Auffinger Chen 14, McKenna 21, Kivimae 23: crit pt complexity and ground state energy in pure spherical bipartite spin glasses

Upper-tail large deviations of ground state

For $E > E_0$, what is the large deviation rate

$$\Phi(E) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\mathrm{GS}(H) \geqslant E) \quad \text{where} \quad \mathrm{GS}(H) = \max_{\boldsymbol{\sigma}\in S_N} \frac{H(\boldsymbol{\sigma})}{N}\,?$$

Upper-tail large deviations of ground state

For $E > E_0$, what is the large deviation rate

$$\Phi(E) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\mathrm{GS}(H) \geqslant E) \quad \text{where} \quad \mathrm{GS}(H) = \max_{\boldsymbol{\sigma}\in S_N} \frac{H(\boldsymbol{\sigma})}{N}\,?$$

Trivial upper bound:

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) \leqslant \mathbb{E}|\mathrm{Crt}(E)| \equiv \mathbb{E}|\{\text{crit pts with } H(\boldsymbol{\sigma})/N \geqslant E\}|$$

Upper-tail large deviations of ground state

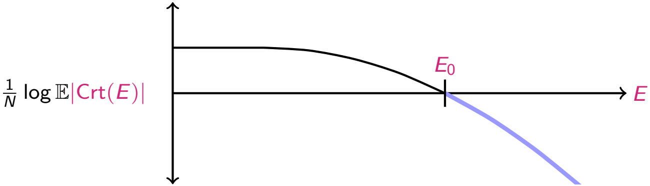
For $E > E_0$, what is the large deviation rate

$$\Phi(E) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\mathrm{GS}(H) \geqslant E) \quad \text{where} \quad \mathrm{GS}(H) = \max_{\boldsymbol{\sigma}\in S_N} \frac{H(\boldsymbol{\sigma})}{N} \,?$$

Trivial upper bound:

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) \leqslant \mathbb{E}|\mathrm{Crt}(E)| \equiv \mathbb{E}|\{\text{crit pts with } H(\boldsymbol{\sigma})/N \geqslant E\}|$$

Implies upper bound on $\Phi(E)$:



Upper-tail large deviations of ground state

For $E > E_0$, what is the large deviation rate

$$\Phi(E) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\mathrm{GS}(H) \geqslant E) \quad \text{where} \quad \mathrm{GS}(H) = \max_{\boldsymbol{\sigma}\in S_N} \frac{H(\boldsymbol{\sigma})}{N} \,?$$

Trivial upper bound:

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) \leqslant \mathbb{E}|\mathrm{Crt}(E)| \equiv \mathbb{E}|\{\text{crit pts with } H(\boldsymbol{\sigma})/N \geqslant E\}|$$

Q: Is this tight?

Upper-tail large deviations of ground state

For $E > E_0$, what is the large deviation rate

$$\Phi(E) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\mathrm{GS}(H) \geqslant E) \quad \text{where} \quad \mathrm{GS}(H) = \max_{\boldsymbol{\sigma}\in S_N} \frac{H(\boldsymbol{\sigma})}{N}\ ?$$

Trivial upper bound:

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) \leqslant \mathbb{E}|\mathrm{Crt}(E)| \equiv \mathbb{E}|\{\text{crit pts with } H(\boldsymbol{\sigma})/N \geqslant E\}|$$

Q: Is this tight?

Theorem (Subag 17; Ben Arous Subag Zeitoni 20; H Sellke 23)

Yes. In fact, in all pure p-spin models,

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) = (1 - o(1))\mathbb{E}|\mathrm{Crt}(E)|$$

Upper-tail large deviations of ground state

For $E > E_0$, what is the large deviation rate

$$\Phi(E) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\mathrm{GS}(H) \geqslant E) \quad \text{where} \quad \mathrm{GS}(H) = \max_{\boldsymbol{\sigma}\in S_N} \frac{H(\boldsymbol{\sigma})}{N}\ ?$$

Trivial upper bound:

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) \leqslant \mathbb{E}|\mathrm{Crt}(E)| \equiv \mathbb{E}|\{\text{crit pts with } H(\boldsymbol{\sigma})/N \geqslant E\}|$$

Q: Is this tight?

Theorem (Subag 17; Ben Arous Subag Zeitoni 20; H Sellke 23)

Yes. In fact, in all pure p-spin models,

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) = (1 - o(1))\mathbb{E}|\mathrm{Crt}(E)|$$

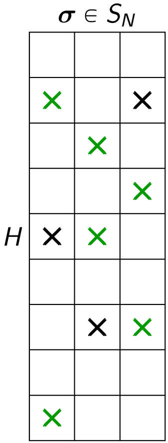
(H Sellke 23: also in a maximal regime of mixed p-spin models — later)

Upper-tail large deviations of ground state

$$\mathbb{P}(\mathsf{GS}(H) \geqslant E) = \mathbb{E}|\{\text{crit pts } \boldsymbol{\sigma} \text{ with } H(\boldsymbol{\sigma})/N \geqslant E \textbf{ and } \mathbf{H}(\boldsymbol{\sigma}) = \mathbf{max}(\mathbf{H})\}|$$
$$\equiv \mathbb{E}|\widetilde{\mathsf{Crt}}(E)|$$

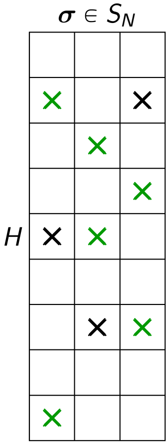
Upper-tail large deviations of ground state

$$\mathbb{P}(\mathsf{GS}(H) \geqslant E) = \mathbb{E}|\{\text{crit pts } \boldsymbol{\sigma} \text{ with } H(\boldsymbol{\sigma})/N \geqslant E \textbf{ and } \mathbf{H}(\boldsymbol{\sigma}) = \mathbf{max(H)}\}|$$
$$\equiv \mathbb{E}|\widetilde{\mathsf{Crt}}(E)|$$



Upper-tail large deviations of ground state

$$\mathbb{P}(\mathsf{GS}(H) \geqslant E) = \mathbb{E}|\{\text{crit pts } \boldsymbol{\sigma} \text{ with } H(\boldsymbol{\sigma})/N \geqslant E \textbf{ and } \mathbf{H}(\boldsymbol{\sigma}) = \mathbf{max(H)}\}|$$
$$\equiv \mathbb{E}|\widetilde{\mathsf{Crt}}(E)|$$

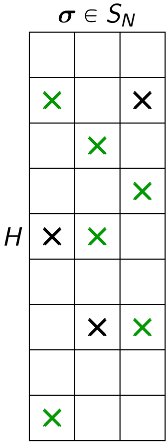


× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

Upper-tail large deviations of ground state

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) = \mathbb{E}|\{\text{crit pts } \boldsymbol{\sigma} \text{ with } H(\boldsymbol{\sigma})/N \geqslant E \textbf{ and } \mathbf{H}(\boldsymbol{\sigma}) = \mathbf{max(H)}\}|$$
$$\equiv \mathbb{E}|\widetilde{\mathrm{Crt}}(E)|$$

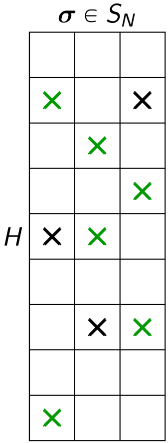


× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$
× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

$\mathbb{E}|\mathrm{Crt}(E)|$ = avg # × per row (including ×)
$\mathbb{E}|\widetilde{\mathrm{Crt}}(E)|$ = avg # × per row

Upper-tail large deviations of ground state

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) = \mathbb{E}|\{\text{crit pts } \boldsymbol{\sigma} \text{ with } H(\boldsymbol{\sigma})/N \geqslant E \textbf{ and } \mathbf{H}(\boldsymbol{\sigma}) = \mathbf{max(H)}\}|$$
$$\equiv \mathbb{E}|\widetilde{\mathrm{Crt}}(E)|$$



× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

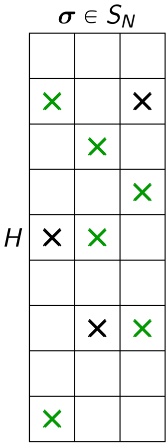
$\mathbb{E}|\mathrm{Crt}(E)| =$ avg # × per row (including ×)

$\mathbb{E}|\widetilde{\mathrm{Crt}}(E)| =$ avg # × per row

$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = \frac{\mathbb{E}|\widetilde{\mathrm{Crt}}(E)|}{\mathbb{E}|\mathrm{Crt}(E)|} = \text{(fraction of green × in grid)}$$

Upper-tail large deviations of ground state

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) = \mathbb{E}|\{\text{crit pts } \boldsymbol{\sigma} \text{ with } H(\boldsymbol{\sigma})/N \geqslant E \text{ and } \mathbf{H}(\boldsymbol{\sigma}) = \mathbf{max}(\mathbf{H})\}|$$
$$\equiv \mathbb{E}|\widetilde{\mathrm{Crt}}(E)|$$



× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

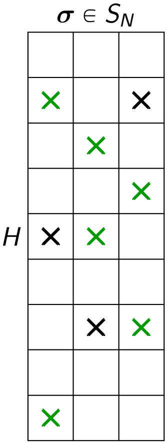
$\mathbb{E}|\mathrm{Crt}(E)| =$ avg # × per row (including ×)

$\mathbb{E}|\widetilde{\mathrm{Crt}}(E)| =$ avg # × per row

$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = \frac{\mathbb{E}|\widetilde{\mathrm{Crt}}(E)|}{\mathbb{E}|\mathrm{Crt}(E)|} = \text{(fraction of green × in grid)}$$
$$= \text{(fraction of green × in any column)}$$

Upper-tail large deviations of ground state

$$\mathbb{P}(\mathrm{GS}(H) \geqslant E) = \mathbb{E}|\{\text{crit pts } \boldsymbol{\sigma} \text{ with } H(\boldsymbol{\sigma})/N \geqslant E \textbf{ and } \mathbf{H(\boldsymbol{\sigma}) = \max(H)}\}|$$
$$\equiv \mathbb{E}|\widetilde{\mathrm{Crt}}(E)|$$



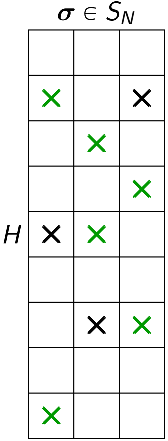
× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$
× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

$\mathbb{E}|\mathrm{Crt}(E)|$ = avg # × per row (including ×)
$\mathbb{E}|\widetilde{\mathrm{Crt}}(E)|$ = avg # × per row

$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = \frac{\mathbb{E}|\widetilde{\mathrm{Crt}}(E)|}{\mathbb{E}|\mathrm{Crt}(E)|} = \text{(fraction of green × in grid)}$$
$$= \text{(fraction of green × in any column)}$$

want to show: this is $1 - o(1)$

Re-interpretation: critical point planted model

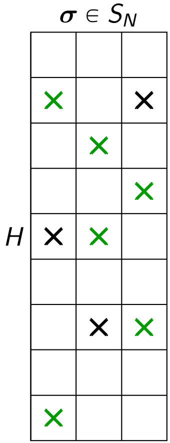


× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = \text{(fraction of green × in any col)}$$

Re-interpretation: critical point planted model



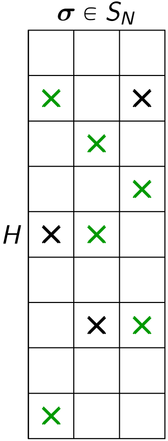
$\times$: $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

$\times$: subset of $\times$ where also $H(\boldsymbol{\sigma}) = \max(H)$

$$\frac{\mathbb{P}(\mathsf{GS}(H) \geqslant E)}{\mathbb{E}|\mathsf{Crt}(E)|} = \text{(fraction of green } \times \text{ in any col)}$$

Critical point planted model:

Re-interpretation: critical point planted model



× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

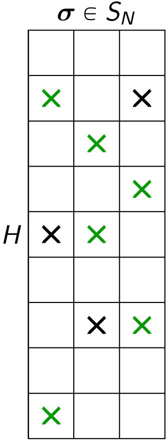
× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

$$\frac{\mathbb{P}(\mathsf{GS}(H) \geqslant E)}{\mathbb{E}|\mathsf{Crt}(E)|} = (\text{fraction of green } \times \text{ in any col})$$

Critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$

Re-interpretation: critical point planted model



× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

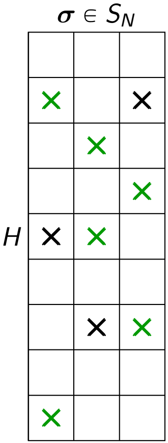
× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = (\text{fraction of green} \times \text{in any col})$$

Critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- sample H conditional on $\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma}) = \mathbf{0}$,
  $H(\boldsymbol{\sigma})/N \geqslant E$

Re-interpretation: critical point planted model



× : $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

× : subset of × where also $H(\boldsymbol{\sigma}) = \max(H)$

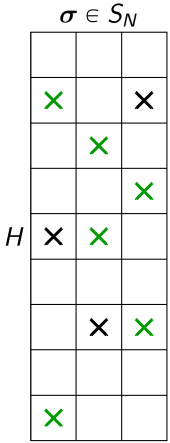
$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = \text{(fraction of green × in any col)}$$

Critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- sample H conditional on $\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma}) = \mathbf{0}$,
  $H(\boldsymbol{\sigma})/N \geqslant E$

$\leftrightarrow$ sample random col, then random × in col

Re-interpretation: critical point planted model



$\times$: $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

$\times$: subset of $\times$ where also $H(\boldsymbol{\sigma}) = \max(H)$

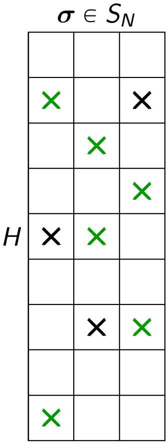
$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = (\text{fraction of green } \times \text{ in any col})$$
$$= \mathbb{P}_{\mathrm{planted}}(\text{sampled } \times \text{ is green})$$

Critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- sample H conditional on $\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma}) = \mathbf{0}$,
  $H(\boldsymbol{\sigma})/N \geqslant E$

$\leftrightarrow$ sample random col, then random $\times$ in col

Re-interpretation: critical point planted model



$\times$: $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

$\times$: subset of $\times$ where also $H(\boldsymbol{\sigma}) = \max(H)$

$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = (\text{fraction of green } \times \text{ in any col})$$

$$= \mathbb{P}_{\text{planted}}(\text{sampled } \times \text{ is green})$$

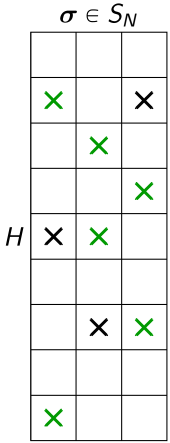
$$= \mathbb{P}_{\text{planted}}(H(\boldsymbol{\sigma}) = \max(H))$$

Critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- sample H conditional on $\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma}) = \mathbf{0}$,
  $H(\boldsymbol{\sigma})/N \geqslant E$

$\leftrightarrow$ sample random col, then random $\times$ in col

Re-interpretation: critical point planted model



$\times$: $\boldsymbol{\sigma}$ crit pt of H with $H(\boldsymbol{\sigma})/N \geqslant E$

$\times$: subset of $\times$ where also $H(\boldsymbol{\sigma}) = \max(H)$

$$\frac{\mathbb{P}(\mathrm{GS}(H) \geqslant E)}{\mathbb{E}|\mathrm{Crt}(E)|} = (\text{fraction of green } \times \text{ in any col})$$

$$= \mathbb{P}_{\text{planted}}(\text{sampled } \times \text{ is green})$$

$$= \mathbb{P}_{\text{planted}}(H(\boldsymbol{\sigma}) = \max(H))$$

Critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- sample H conditional on $\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma}) = \mathbf{0}$,

  $H(\boldsymbol{\sigma})/N \geqslant E$

$\leftrightarrow$ sample random col, then random $\times$ in col

Remains to show: $\mathbb{P}_{\text{planted}}(H(\boldsymbol{\sigma}) = \max(H)) = 1 - o(1)$

Planted large critical point is whp maximal

Recall critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- sample H conditional on $\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma}) = \mathbf{0}$ and $H(\boldsymbol{\sigma})/N \geqslant E$

Remains to show: $\mathbb{P}_{\mathrm{planted}}(H(\boldsymbol{\sigma}) = \max(H)) = 1 - o(1)$

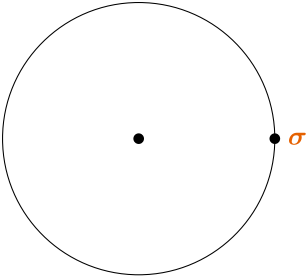
Planted large critical point is whp maximal

Recall critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- sample H conditional on $\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma}) = \mathbf{0}$ and $H(\boldsymbol{\sigma})/N \geqslant E$

Remains to show: $\mathbb{P}_{\mathrm{planted}}(H(\boldsymbol{\sigma}) = \max(H)) = 1 - o(1)$

Planted H is explicit **spiked** spherical spin glass



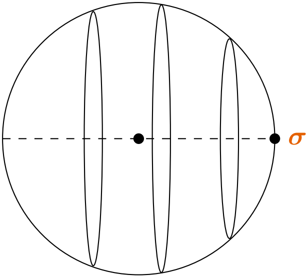
Planted large critical point is whp maximal

Recall critical point planted model:

- sample $\boldsymbol{\sigma} \sim \mathrm{unif}(S_N)$
- sample H conditional on $\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma}) = \mathbf{0}$ and $H(\boldsymbol{\sigma})/N \geqslant E$

Remains to show: $\mathbb{P}_{\mathrm{planted}}(H(\boldsymbol{\sigma}) = \max(H)) = 1 - o(1)$

Planted H is explicit **spiked** spherical spin glass



On each orthogonal band, max of H bounded by **Guerra's interpolation**

Beyond pure models

Q: does $\mathbb{P}(\mathrm{GS}(H) \geqslant E) = (1 - o(1))\mathbb{E}|\mathrm{Crt}(E)|$ in **mixed** p**-spin model**?

$$H(\boldsymbol{\sigma}) = \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1, \ldots, i_p = 1}^{N} g_{i_1, \ldots, i_p} \sigma_{i_1} \cdots \sigma_{i_p}$$

Beyond pure models

Q: does $\mathbb{P}(\mathrm{GS}(H) \geqslant E) = (1 - o(1))\mathbb{E}|\mathrm{Crt}(E)|$ in **mixed p-spin model**?

$$H(\boldsymbol{\sigma}) = \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1,\dots,i_p=1}^{N} g_{i_1,\dots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}$$

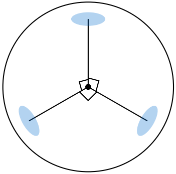
A: yes, for all "zero-temperature 1RSB" models (and this is maximal)

Beyond pure models

Q: does $\mathbb{P}(\mathrm{GS}(H) \geqslant E) = (1 - o(1))\mathbb{E}|\mathrm{Crt}(E)|$ in **mixed p-spin model**?

$$H(\boldsymbol{\sigma}) = \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1,\dots,i_p=1}^{N} g_{i_1,\dots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}$$

A: yes, for all "zero-temperature 1RSB" models (and this is maximal)



As $\beta \to \infty$, Gibbs measure

$$\mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}\,\mathrm{d}\boldsymbol{\sigma}$$

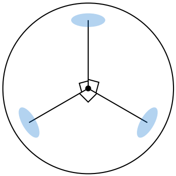
concentrates on **orthogonal** spherical caps whp.

Beyond pure models

Q: does $\mathbb{P}(\mathrm{GS}(H) \geqslant E) = (1 - o(1))\mathbb{E}|\mathrm{Crt}(E)|$ in **mixed *p*-spin model**?

$$H(\boldsymbol{\sigma}) = \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1,\ldots,i_p} \sigma_{i_1} \cdots \sigma_{i_p}$$

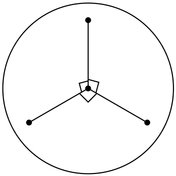
A: yes, for all “zero-temperature 1RSB” models (and this is maximal)



As $\beta \to \infty$, Gibbs measure

$$\mu_\beta(\mathrm{d}\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}\,\mathrm{d}\boldsymbol{\sigma}$$

concentrates on **orthogonal** spherical caps whp.

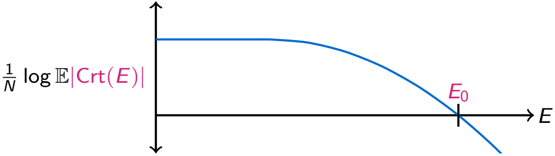


Equivalently: crit pts of H with value $\approx \mathrm{GS}_N$ are whp orthogonal.

(That is, they **do not cluster**)

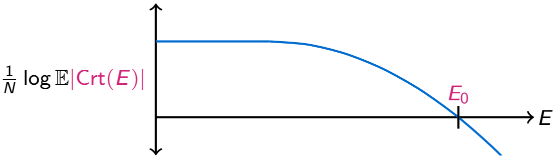
Complexity-based proof of ground state energy

Auffinger Ben Arous Černý 13 + Subag 17: in **pure p-spin models**, complexity-based proof of $\mathrm{GS}_N \xrightarrow{p} E_0$. Independent of Parisi formula.



Complexity-based proof of ground state energy

Auffinger Ben Arous Černý 13 + Subag 17: in **pure p-spin models**, complexity-based proof of $\mathrm{GS}_N \xrightarrow{p} E_0$. Independent of Parisi formula.

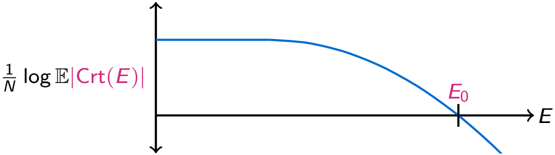


Ben Arous Subag Zeitouni 20: similarly $\mathrm{GS}_N \xrightarrow{p} E_0$ in **some regime of mixed p-spin models**

Q: for **which models** can complexity considerations show $\mathrm{GS}_N \xrightarrow{p} E_0$?

Complexity-based proof of ground state energy

Auffinger Ben Arous Černý 13 + Subag 17: in **pure p-spin models**, complexity-based proof of $\mathrm{GS}_N \xrightarrow{p} E_0$. Independent of Parisi formula.



Ben Arous Subag Zeitouni 20: similarly $\mathrm{GS}_N \xrightarrow{p} E_0$ in **some regime of mixed p-spin models**

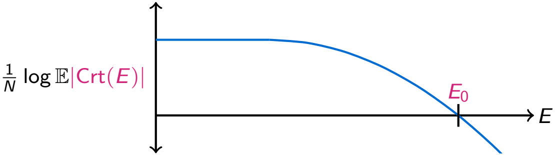
Q: for **which models** can complexity considerations show $\mathrm{GS}_N \xrightarrow{p} E_0$?

Corollary (H Sellke 23)

In all **zero-temperature 1RSB** *models (and this is maximal),* $\mathrm{GS}_N \xrightarrow{p} E_0$

Complexity-based proof of ground state energy

Auffinger Ben Arous Černý 13 + Subag 17: in **pure p-spin models**, complexity-based proof of $\mathrm{GS}_N \xrightarrow{p} E_0$. Independent of Parisi formula.



Ben Arous Subag Zeitouni 20: similarly $\mathrm{GS}_N \xrightarrow{p} E_0$ in **some regime of mixed p-spin models**

Q: for **which models** can complexity considerations show $\mathrm{GS}_N \xrightarrow{p} E_0$?

Corollary (H Sellke 23)

In all **zero-temperature 1RSB** *models (and this is maximal),* $\mathrm{GS}_N \xrightarrow{p} E_0$

(uses Guerra interpolation, but avoids more difficult Parisi formula LB)

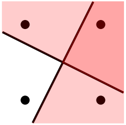
Outline of part I: applications of planting in disordered models

The classic planting trick: planting a Gibbs sample

Ground state large deviations in spherical spin glasses

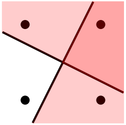
TAP planting: capacity of the Ising perceptron

The perceptron model



Intersection of discrete cube $\{\pm 1\}^N$
with M i.i.d. random half-spaces

The perceptron model

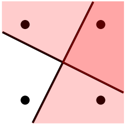


Intersection of discrete cube $\{\pm 1\}^N$ with M i.i.d. random half-spaces

Formally: for $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M \sim \mathcal{N}(0, I_N)$,

$$\boldsymbol{S} = \left\{ \boldsymbol{x} \in \{\pm 1\}^N : (\boldsymbol{g}^a, \boldsymbol{x}) \geqslant 0, \quad \forall 1 \leqslant a \leqslant M \right\}$$

The perceptron model



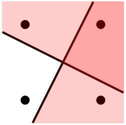
Intersection of discrete cube $\{\pm 1\}^N$
with $M = \alpha N$ i.i.d. random half-spaces

α = constraint density

Formally: for $g^1, \ldots, g^M \sim \mathcal{N}(0, I_N)$,

$$S = \left\{ x \in \{\pm 1\}^N : (g^a, x) \geqslant 0, \quad \forall 1 \leqslant a \leqslant M \right\}$$

The perceptron model



Intersection of discrete cube $\{\pm 1\}^N$
with $M = \alpha N$ i.i.d. random half-spaces

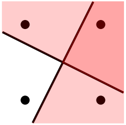
α = constraint density

Formally: for $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M \sim \mathcal{N}(0, I_N)$,

$$\boldsymbol{S} = \left\{ \boldsymbol{x} \in \{\pm 1\}^N : (\boldsymbol{g}^a, \boldsymbol{x}) \geqslant 0, \quad \forall 1 \leqslant a \leqslant M \right\}$$

Capacity problem: what is the critical $\alpha_\star$ where $\boldsymbol{S}$ goes from nonempty to empty (with high probability as $N \to \infty$)?

The perceptron model



Intersection of discrete cube $\{\pm 1\}^N$
with $M = \alpha N$ i.i.d. random half-spaces

α = constraint density

Formally: for $\boldsymbol{g}^1, \dots, \boldsymbol{g}^M \sim \mathcal{N}(0, I_N)$,

$$\boldsymbol{S} = \Big\{ \boldsymbol{x} \in \{\pm 1\}^N : (\boldsymbol{g}^a, \boldsymbol{x}) \geqslant 0, \quad \forall 1 \leqslant a \leqslant M \Big\}$$

Capacity problem: what is the critical $\alpha_\star$ where $\boldsymbol{S}$ goes from nonempty to empty (with high probability as $N \to \infty$)?

$\leftrightarrow$ memorization capacity of a neural network (Gardner 87)

Main result

Conjecture (Krauth Mézard 89)

For the Ising perceptron, $\alpha_\star = \alpha_{\mathsf{KM}} \approx 0.833$.

Main result

Conjecture (Krauth Mézard 89)

For the Ising perceptron, $\alpha_\star = \alpha_{\mathsf{KM}} \approx 0.833$.

Theorem (Ding Sun 18)

$\alpha_\star \geqslant \alpha_{\mathsf{KM}}$, *under condition that an explicit univariate function is* $\leqslant 0$.

Main result

Conjecture (Krauth Mézard 89)

For the Ising perceptron, $\alpha_\star = \alpha_{\mathsf{KM}} \approx 0.833$*.*

Theorem (Ding Sun 18)

$\alpha_\star \geqslant \alpha_{\mathsf{KM}}$*, under condition that an explicit univariate function is* $\leqslant 0$*.*

Theorem (H 24)

$\alpha_\star \leqslant \alpha_{\mathsf{KM}}$*, under condition that an explicit bivariate function is* $\leqslant 0$*.*

(next slide)

Main result

Conjecture (Krauth Mézard 89)

For the Ising perceptron, $\alpha_\star = \alpha_{\mathsf{KM}} \approx 0.833$.

Theorem (Ding Sun 18)

$\alpha_\star \geqslant \alpha_{\mathsf{KM}}$, *under condition that an explicit univariate function is* $\leqslant 0$.

Theorem (H 24)

$\alpha_\star \leqslant \alpha_{\mathsf{KM}}$, *under condition that an explicit bivariate function is* $\leqslant 0$.

(next slide)

Both results hold for more general model with **margin** $\kappa \in \mathbb{R}$:

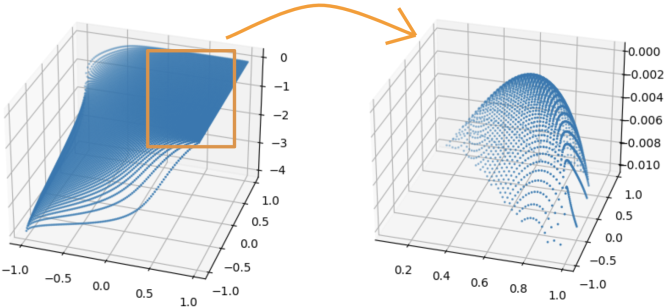
$$S = \left\{ \boldsymbol{x} \in \{\pm 1\}^N : (\boldsymbol{g}^a, \boldsymbol{x}) \geqslant \kappa\sqrt{N}, \quad \forall 1 \leqslant a \leqslant M \right\}$$

for analogous threshold $\alpha_{\mathsf{KM}}(\kappa)$, under further numerical conditions depending on κ.

The function in our numerical condition

$\mathscr{S}_\star(1,0) = 0$ local max, conjecturally unique global max

Plot of $\mathscr{S}_\star$ (domain $\mathbb{R}^2$ reparametrized to $[-1,1]^2$):



$x, y \in [-1, 1]$ $\mathscr{S}_* \geq -0.01$

Previous work

- Shcherbina Tirozzi 03, Stojnic 13: capacity of $\kappa \geqslant 0$ **spherical** perceptron (domain $\sqrt{N}\mathbb{S}^{N-1}$ instead of $\{\pm 1\}^N$)

Previous work

- Shcherbina Tirozzi 03, Stojnic 13: capacity of $\boldsymbol{\kappa} \geqslant 0$ **spherical** perceptron (domain $\sqrt{N}\mathbb{S}^{N-1}$ instead of $\{\pm 1\}^N$)

For **Ising** perceptron, $\boldsymbol{\kappa} = 0$:

- Talagrand 11, Xu 21, Nakajima Sun 23 sharp threshold **sequence** $\boldsymbol{\alpha}_\star(N)$ (non-explicit, doesn't imply $\boldsymbol{\alpha}_\star = \lim_{N\to\infty} \boldsymbol{\alpha}_\star(N)$ exists)

Previous work

- Shcherbina Tirozzi 03, Stojnic 13: capacity of $\kappa \geqslant 0$ **spherical** perceptron (domain $\sqrt{N}\mathbb{S}^{N-1}$ instead of $\{\pm 1\}^N$)

For **Ising** perceptron, $\kappa = 0$:

- Talagrand 11, Xu 21, Nakajima Sun 23 sharp threshold **sequence** $\alpha_\star(N)$ (non-explicit, doesn't imply $\alpha_\star = \lim_{N\to\infty} \alpha_\star(N)$ exists)

- Simple bound: $\alpha_\star \leqslant 1$ (in a few slides)

- Kim Roche 98, Talagrand 99 & 00: $0.005 \leqslant \alpha_\star \leqslant 0.9963$

Previous work

- Shcherbina Tirozzi 03, Stojnic 13: capacity of $\kappa \geqslant 0$ **spherical** perceptron (domain $\sqrt{N}\mathbb{S}^{N-1}$ instead of $\{\pm 1\}^N$)

For **Ising** perceptron, $\kappa = 0$:

- Talagrand 11, Xu 21, Nakajima Sun 23 sharp threshold **sequence** $\alpha_\star(N)$ (non-explicit, doesn't imply $\alpha_\star = \lim_{N\to\infty} \alpha_\star(N)$ exists)

- Simple bound: $\alpha_\star \leqslant 1$ (in a few slides)

- Kim Roche 98, Talagrand 99 & 00: $0.005 \leqslant \alpha_\star \leqslant 0.9963$

- Ding Sun 18: $\alpha_\star \geqslant \alpha_{\mathrm{KM}} \approx 0.833$

- Altschuler Tikhomirov 24: $\alpha_\star \leqslant 0.847$

Previous work

- Shcherbina Tirozzi 03, Stojnic 13: capacity of $\kappa \geqslant 0$ **spherical** perceptron (domain $\sqrt{N}\mathbb{S}^{N-1}$ instead of $\{\pm 1\}^N$)

For **Ising** perceptron, $\kappa = 0$:

- Talagrand 11, Xu 21, Nakajima Sun 23 sharp threshold **sequence** $\alpha_\star(N)$ (non-explicit, doesn't imply $\alpha_\star = \lim_{N\to\infty} \alpha_\star(N)$ exists)
- Simple bound: $\alpha_\star \leqslant 1$ (in a few slides)
- Kim Roche 98, Talagrand 99 & 00: $0.005 \leqslant \alpha_\star \leqslant 0.9963$
- Ding Sun 18: $\alpha_\star \geqslant \alpha_{\mathrm{KM}} \approx 0.833$
- Altschuler Tikhomirov 24: $\alpha_\star \leqslant 0.847$
- H 24: $\alpha_\star \leqslant \alpha_{\mathrm{KM}}$

Review: 1st/2nd moment method

- $\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha})| \ll 1 \Rightarrow$ no solution at constraint density $\boldsymbol{\alpha}$ (whp)
- $\mathbb{E}[|\boldsymbol{S}(N\boldsymbol{\alpha})|^2] = O(1) \cdot (\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha})|)^2 \Rightarrow \exists$ solution at density $\boldsymbol{\alpha}$ (with $\Omega(1)$ probability)

Review: 1st/2nd moment method

- $\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha})| \ll 1 \Rightarrow$ no solution at constraint density $\boldsymbol{\alpha}$ (whp)
- $\mathbb{E}[|\boldsymbol{S}(N\boldsymbol{\alpha})|^2] = O(1) \cdot (\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha})|)^2 \Rightarrow \exists$ solution at density $\boldsymbol{\alpha}$ (with $\Omega(1)$ probability)

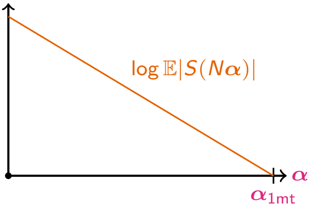
This provides a simple strategy to (try to) locate capacity:

Review: 1st/2nd moment method

- $\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha})| \ll 1 \Rightarrow$ no solution at constraint density $\boldsymbol{\alpha}$ (whp)
- $\mathbb{E}[|\boldsymbol{S}(N\boldsymbol{\alpha})|^2] = O(1) \cdot (\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha})|)^2 \Rightarrow \exists$ solution at density $\boldsymbol{\alpha}$ (with $\Omega(1)$ probability)

This provides a simple strategy to (try to) locate capacity:

- Let $\boldsymbol{\alpha}_{1\text{mt}}$ solve $\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha}_{1\text{mt}})| = 1$. (So no solns whp for $\boldsymbol{\alpha} > \boldsymbol{\alpha}_{1\text{mt}}$)

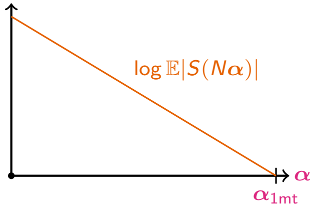


Review: 1st/2nd moment method

- $\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha})| \ll 1 \Rightarrow$ no solution at constraint density $\boldsymbol{\alpha}$ (whp)
- $\mathbb{E}[|\boldsymbol{S}(N\boldsymbol{\alpha})|^2] = O(1) \cdot (\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha})|)^2 \Rightarrow \exists$ solution at density $\boldsymbol{\alpha}$ (with $\Omega(1)$ probability)

This provides a simple strategy to (try to) locate capacity:

- Let $\boldsymbol{\alpha}_{1\text{mt}}$ solve $\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha}_{1\text{mt}})| = 1$. (So no solns whp for $\boldsymbol{\alpha} > \boldsymbol{\alpha}_{1\text{mt}}$)

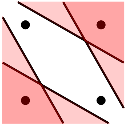


- (Hope to) show $\mathbb{E}[|\boldsymbol{S}(N\boldsymbol{\alpha}_{1\text{mt}})|^2] \asymp (\mathbb{E}|\boldsymbol{S}(N\boldsymbol{\alpha}_{1\text{mt}})|)^2 = 1$.

If so, $\boldsymbol{\alpha}_\star = \boldsymbol{\alpha}_{1\text{mt}}$.

1st/2nd moment method: a success story

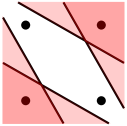
Symmetric Ising perceptron: constraints $|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}$
(Aubin Perkins Zdeborová 19, Perkins Xu 21, Abbe Li Sly 22, ...)



The moment method locates $\alpha_\star$ in this model!

1st/2nd moment method: a success story

Symmetric Ising perceptron: constraints $|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}$
(Aubin Perkins Zdeborová 19, Perkins Xu 21, Abbe Li Sly 22, ...)

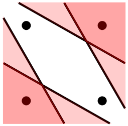


The moment method locates $\alpha_\star$ in this model!

Solution set: $\boldsymbol{S}(M) = \{\boldsymbol{x} \in \{\pm 1\}^N : \ |(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N} \quad \forall 1 \leqslant a \leqslant M\}$

1st/2nd moment method: a success story

Symmetric Ising perceptron: constraints $|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}$
(Aubin Perkins Zdeborová 19, Perkins Xu 21, Abbe Li Sly 22, . . .)



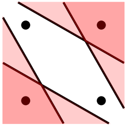
The moment method locates $\alpha_\star$ in this model!

Solution set: $\boldsymbol{S}(M) = \{\boldsymbol{x} \in \{\pm 1\}^N : \ |(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N} \quad \forall 1 \leqslant a \leqslant M\}$

$$\mathbb{E}|\boldsymbol{S}(M)| = 2^N \cdot \mathbb{P}(\text{a fixed } \boldsymbol{x} \text{ is in } \boldsymbol{S}(M))$$

1st/2nd moment method: a success story

Symmetric Ising perceptron: constraints $|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}$
(Aubin Perkins Zdeborová 19, Perkins Xu 21, Abbe Li Sly 22, ...)



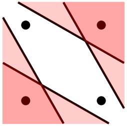
The moment method locates $\alpha_\star$ in this model!

Solution set: $\boldsymbol{S}(M) = \{\boldsymbol{x} \in \{\pm 1\}^N : \ |(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N} \quad \forall 1 \leqslant a \leqslant M\}$

$$\mathbb{E}|\boldsymbol{S}(M)| = 2^N \cdot \mathbb{P}(\text{a fixed } \boldsymbol{x} \text{ is in } \boldsymbol{S}(M))$$
$$= 2^N \cdot \prod_{a=1}^{M} \mathbb{P}\Big(|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}\Big)$$

1st/2nd moment method: a success story

Symmetric Ising perceptron: constraints $|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}$
(Aubin Perkins Zdeborová 19, Perkins Xu 21, Abbe Li Sly 22, ...)



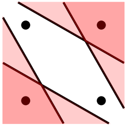
The moment method locates $\alpha_\star$ in this model!

Solution set: $\boldsymbol{S}(M) = \{\boldsymbol{x} \in \{\pm 1\}^N : \ |(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N} \quad \forall 1 \leqslant a \leqslant M\}$

$$\mathbb{E}|\boldsymbol{S}(M)| = 2^N \cdot \mathbb{P}(\text{a fixed } \boldsymbol{x} \text{ is in } \boldsymbol{S}(M))$$
$$= 2^N \cdot \prod_{a=1}^{M} \mathbb{P}\Big(|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}\Big) = 2^N \cdot \mathbb{P}(|\mathcal{N}(0,1)| \leqslant \boldsymbol{\kappa})^M$$

1st/2nd moment method: a success story

Symmetric Ising perceptron: constraints $|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}$
(Aubin Perkins Zdeborová 19, Perkins Xu 21, Abbe Li Sly 22, . . .)



The moment method locates $\boldsymbol{\alpha}_\star$ in this model!

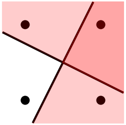
Solution set: $\boldsymbol{S}(M) = \{\boldsymbol{x} \in \{\pm 1\}^N : \ |(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N} \quad \forall 1 \leqslant a \leqslant M\}$

$$\mathbb{E}|\boldsymbol{S}(M)| = 2^N \cdot \mathbb{P}(\text{a fixed } \boldsymbol{x} \text{ is in } \boldsymbol{S}(M))$$
$$= 2^N \cdot \prod_{a=1}^{M} \mathbb{P}\Big(|(\boldsymbol{g}^a, \boldsymbol{x})| \leqslant \boldsymbol{\kappa}\sqrt{N}\Big) = 2^N \cdot \mathbb{P}(|\mathcal{N}(0,1)| \leqslant \boldsymbol{\kappa})^M$$

Can similarly calculate $\mathbb{E}[|\boldsymbol{S}(M)|^2]$, verify $\mathbb{E}[|\boldsymbol{S}(M)|^2] \asymp (\mathbb{E}|\boldsymbol{S}(M)|)^2$.

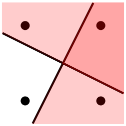
Direct moment method fails in asymmetric model

In our model, $\boldsymbol{S}(\alpha N) = \{\boldsymbol{x} \in \{\pm 1\}^N : \ (\boldsymbol{g}^a, \boldsymbol{x}) \geqslant 0 \quad \forall 1 \leqslant a \leqslant \alpha N\}$



Direct moment method fails in asymmetric model

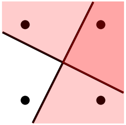
In our model, $\boldsymbol{S}(\alpha N) = \{\boldsymbol{x} \in \{\pm 1\}^N : (\boldsymbol{g}^a, \boldsymbol{x}) \geqslant 0 \quad \forall 1 \leqslant a \leqslant \alpha N\}$



$\mathbb{E}|\boldsymbol{S}(\alpha N)| = 2^N \cdot 2^{-\alpha N}$, so $\alpha_{1\mathrm{mt}} = 1$. This proves $\alpha_\star \leqslant 1$.

Direct moment method fails in asymmetric model

In our model, $\boldsymbol{S}(\alpha N) = \{\boldsymbol{x} \in \{\pm 1\}^N : \ (\boldsymbol{g}^a, \boldsymbol{x}) \geqslant 0 \quad \forall 1 \leqslant a \leqslant \alpha N\}$

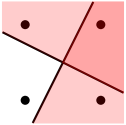


$\mathbb{E}|\boldsymbol{S}(\alpha N)| = 2^N \cdot 2^{-\alpha N}$, so $\alpha_{1\mathrm{mt}} = 1$. This proves $\alpha_\star \leqslant 1$.

But. . . this doesn't locate true threshold $\alpha_{\mathrm{KM}} \approx 0.833$ ☹

Direct moment method fails in asymmetric model

In our model, $\boldsymbol{S}(\alpha N) = \{\boldsymbol{x} \in \{\pm 1\}^N : \ (\boldsymbol{g}^a, \boldsymbol{x}) \geqslant 0 \quad \forall 1 \leqslant a \leqslant \alpha N\}$

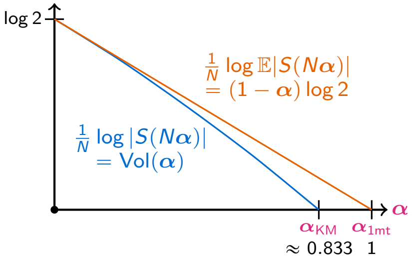


$\mathbb{E}|\boldsymbol{S}(\alpha N)| = 2^N \cdot 2^{-\alpha N}$, so $\alpha_{1\mathrm{mt}} = 1$. This proves $\alpha_\star \leqslant 1$.

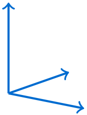
But... this doesn't locate true threshold $\alpha_{\mathrm{KM}} \approx 0.833$ ☹

Our approach: pass to a contiguous planted model in which 1st/2nd moment method locates capacity.
Next few slides motivate choice of planted model.

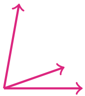
What goes wrong? A large deviations perspective



$\mathbb{E}|S(N\alpha)|$ dominated by events where the g^a are **atypically correlated**



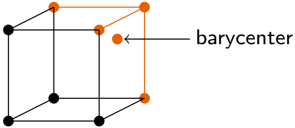
Typically: g^a orthogonal



Atypically: g^a correlated, which inflates # solutions

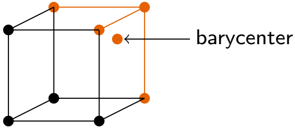
Remedy: conditional moment method

Key intuition of Ding Sun 18, Bolthausen 19: 1st mt failure caused by large deviation events in **barycenter** of solution set ***S***



Remedy: conditional moment method

Key intuition of Ding Sun 18, Bolthausen 19: 1st mt failure caused by large deviation events in **barycenter** of solution set $\boldsymbol{S}$

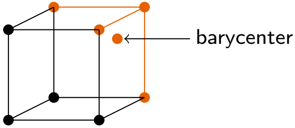


That is, $\mathbb{E}(|\boldsymbol{S}|) \gg (\text{typical } |\boldsymbol{S}|)$ but we expect, for **typical** realization of barycenter:

$$(\text{typical } |\boldsymbol{S}|) \quad \asymp \quad \mathbb{E}(|\boldsymbol{S}| \,|\, \text{barycenter})$$

Remedy: conditional moment method

Key intuition of Ding Sun 18, Bolthausen 19: 1st mt failure caused by large deviation events in **barycenter** of solution set $\boldsymbol{S}$

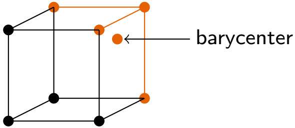


That is, $\mathbb{E}(|\boldsymbol{S}|) \gg$ (typical $|\boldsymbol{S}|$) but we expect, for **typical** realization of barycenter:

$$(\text{typical } |\boldsymbol{S}|) \asymp \mathbb{E}(|\boldsymbol{S}| \,|\, \text{barycenter}) \asymp \mathbb{E}(|\boldsymbol{S}|^2 \,|\, \text{barycenter})^{1/2}$$

Remedy: conditional moment method

Key intuition of Ding Sun 18, Bolthausen 19: 1st mt failure caused by large deviation events in **barycenter** of solution set $\boldsymbol{S}$



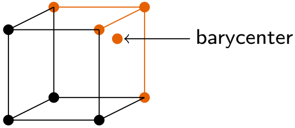
That is, $\mathbb{E}(|\boldsymbol{S}|) \gg (\text{typical } |\boldsymbol{S}|)$ but we expect, for **typical** realization of barycenter:

$$(\text{typical } |\boldsymbol{S}|) \asymp \mathbb{E}(|\boldsymbol{S}| \,|\, \text{barycenter}) \asymp \mathbb{E}(|\boldsymbol{S}|^2 \,|\, \text{barycenter})^{1/2}$$

Suggests plan: **condition** on barycenter, then 1st/2nd moment

Remedy: conditional moment method

Key intuition of Ding Sun 18, Bolthausen 19: 1st mt failure caused by large deviation events in **barycenter** of solution set $\boldsymbol{S}$



That is, $\mathbb{E}(|\boldsymbol{S}|) \gg (\text{typical } |\boldsymbol{S}|)$ but we expect, for **typical** realization of barycenter:

$$(\text{typical } |\boldsymbol{S}|) \quad \asymp \quad \mathbb{E}(|\boldsymbol{S}| \,|\, \text{barycenter}) \quad \asymp \quad \mathbb{E}(|\boldsymbol{S}|^2 \,|\, \text{barycenter})^{1/2}$$

Suggests plan: **condition** on barycenter, then 1st/2nd moment

Will implement by **planting** a certain **heuristic proxy** of barycenter

Heuristic description of barycenter

TAP equation (Thouless Anderson Palmer 77): nonlinear system in

- $\boldsymbol{G} \in \mathbb{R}^{M \times N}$ matrix with rows $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M$
- $\boldsymbol{m} \in \mathbb{R}^N$ barycenter of $\boldsymbol{S}$
- $\boldsymbol{n} \in \mathbb{R}^M$ average slacks of constraints: $n_a = \mathrm{avg}_{\boldsymbol{x} \in \boldsymbol{S}} \{ (\boldsymbol{g}^a, \boldsymbol{x})/\sqrt{N} \}$

Heuristic description of barycenter

TAP equation (Thouless Anderson Palmer 77): nonlinear system in

- $\boldsymbol{G} \in \mathbb{R}^{M\times N}$ matrix with rows $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M$
- $\boldsymbol{m} \in \mathbb{R}^N$ barycenter of $\boldsymbol{S}$
- $\boldsymbol{n} \in \mathbb{R}^M$ average slacks of constraints: $n_a = \text{avg}_{\boldsymbol{x}\in\boldsymbol{S}}\{\, (\boldsymbol{g}^a, \boldsymbol{x})/\sqrt{N}\,\}$

For explicit nonlinearities $\dot{F}, \widehat{F} : \mathbb{R} \to \mathbb{R}$, constants b, d:

$$\boldsymbol{m} = \dot{F}\left(\frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}\right) \qquad\qquad \boldsymbol{n} = \widehat{F}\left(\frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}\right)$$

Heuristic description of barycenter

TAP equation (Thouless Anderson Palmer 77): nonlinear system in

- $\boldsymbol{G} \in \mathbb{R}^{M \times N}$ matrix with rows $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M$
- $\boldsymbol{m} \in \mathbb{R}^N$ barycenter of $\boldsymbol{S}$
- $\boldsymbol{n} \in \mathbb{R}^M$ average slacks of constraints: $n_a = \text{avg}_{\boldsymbol{x} \in \boldsymbol{S}} \{ (\boldsymbol{g}^a, \boldsymbol{x})/\sqrt{N} \}$

For explicit nonlinearities $\dot{F}, \widehat{F} : \mathbb{R} \to \mathbb{R}$, constants b, d:

$$\boldsymbol{m} = \dot{F}\left(\frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}\right) \qquad \boldsymbol{n} = \widehat{F}\left(\frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}\right)$$

Physics prediction: whp over $\boldsymbol{G}$, this has a unique solution $(\boldsymbol{m}, \boldsymbol{n})$ (which has the physical meaning above)

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to $\text{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

---

TAP planted model:

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M\times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to $\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving $\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Under physics prediction: **planted ≈ null**!

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M\times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to $\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

Physics prediction $\Rightarrow$
$(\boldsymbol{m}, \boldsymbol{n})$ " $=$ " $f(\boldsymbol{G})$

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving $\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Under physics prediction: **planted $\approx$ null**!

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to $\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving $\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Under physics prediction: **planted ≈ null**!

Physics prediction ⇒ $(\boldsymbol{m}, \boldsymbol{n})$ " = " $f(\boldsymbol{G})$

$\mathrm{Law}(\boldsymbol{G})$

Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

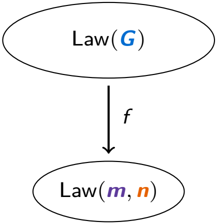
---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Under physics prediction: **planted ≈ null**!

Physics prediction $\Rightarrow$
$(\boldsymbol{m}, \boldsymbol{n})$ " = " $f(\boldsymbol{G})$



Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

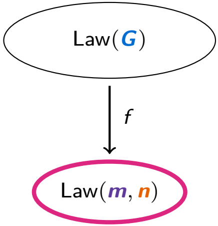
---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Under physics prediction: **planted ≈ null**!

Physics prediction ⇒ $(\boldsymbol{m}, \boldsymbol{n})$ " = " $f(\boldsymbol{G})$



Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

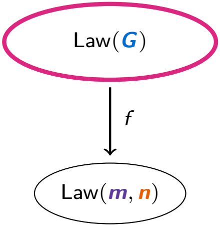
---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Under physics prediction: **planted $\approx$ null**!

Physics prediction $\Rightarrow$
$(\boldsymbol{m}, \boldsymbol{n})$ " = " $f(\boldsymbol{G})$



Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to $\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

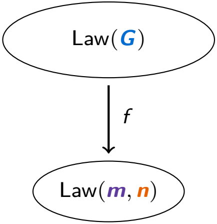
---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving $\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Under physics prediction: **planted ≈ null**!

Physics prediction ⇒
$(\boldsymbol{m}, \boldsymbol{n})$ " = " $f(\boldsymbol{G})$



Key idea: TAP planted model

Null model:

- $\boldsymbol{G} \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- $(\boldsymbol{m}, \boldsymbol{n})$ solution to TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(hard)**

Physics prediction: whp $(\boldsymbol{m}, \boldsymbol{n})$ exists & unique

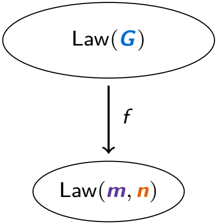
---

TAP planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law (explicit prediction)
- Sample $\boldsymbol{G}$ conditioned on solving TAP$(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$ **(easy)**

Under physics prediction: **planted ≈ null**!
Also: **1st/2nd mmt works in planted model**

Physics prediction $\Rightarrow$
$(\boldsymbol{m}, \boldsymbol{n})$ " $=$ " $f(\boldsymbol{G})$



Key idea: TAP planted model

Null model:

- $G \in \mathbb{R}^{M \times N} \sim$ IID $\mathcal{N}(0,1)$ entries
- (m, n) solution to TAP$(G; m, n)$ **(hard)**

Physics prediction: whp (m, n) exists & unique

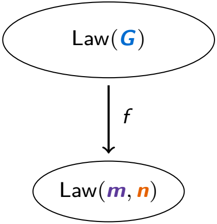
TAP planted model:

- Sample (m, n) from its law (explicit prediction)
- Sample G conditioned on solving TAP$(G; m, n)$ **(easy)**

Under physics prediction: **planted $\approx$ null**!
Also: **1st/2nd mmt works in planted model**

❗ existence/uniqueness of (m, n) is **not proven**.
We will need to justify that planted $\approx$ null.

Physics prediction $\Rightarrow$
(m, n) " = " $f(G)$



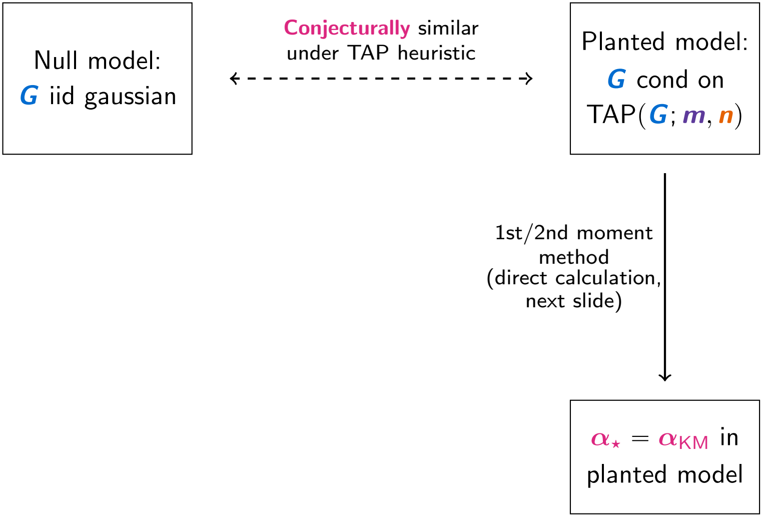
Proof roadmap

Null model:
G iid gaussian

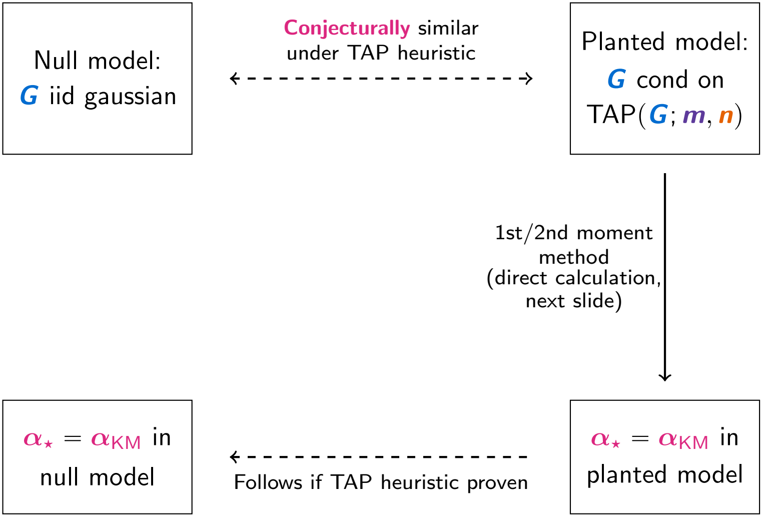
Conjecturally similar under TAP heuristic ⟷

Planted model:
G cond on TAP$(G; m, n)$

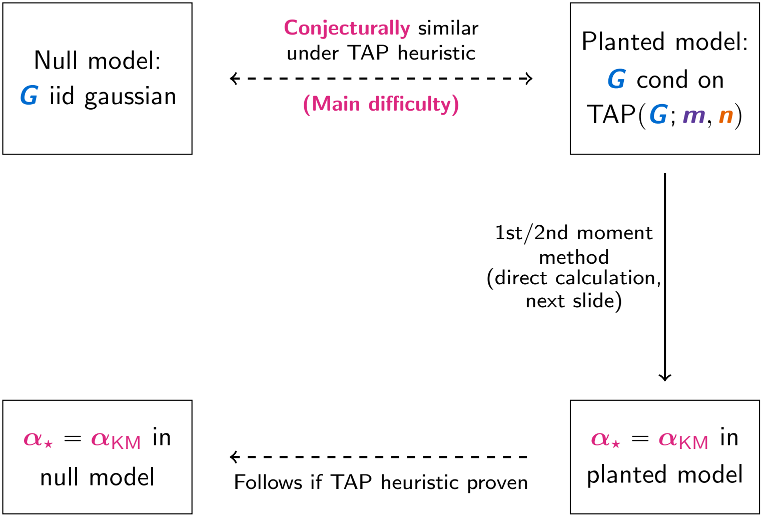
Proof roadmap



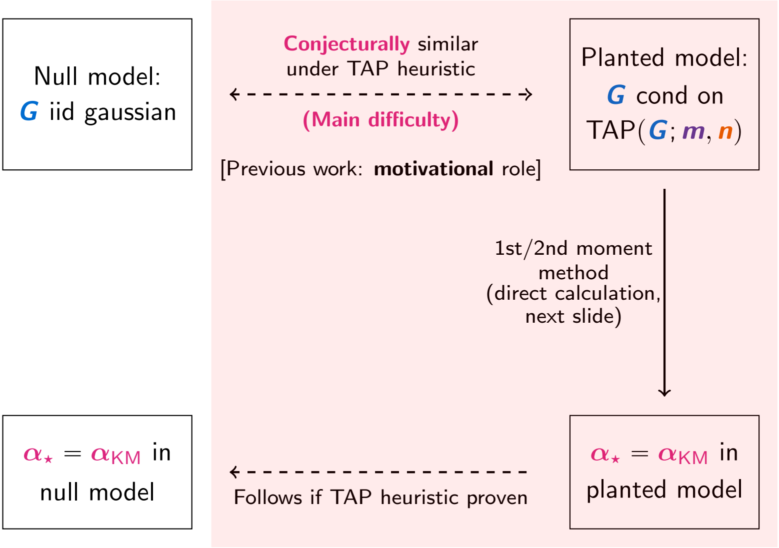
Proof roadmap



Proof roadmap



Proof roadmap



Proof roadmap

Null model:
$\boldsymbol{G}$ iid gaussian

Conjecturally similar
under TAP heuristic

←- - - - - - - - - - - - - - - - - -→

(Main difficulty)

[Previous work: **motivational** role]

Planted model:
$\boldsymbol{G}$ cond on
$\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$

DS18: lower bd by
1st/2nd mmt on
truncation of $|\boldsymbol{S}|$

1st/2nd moment
method
(direct calculation,
next slide)

$\alpha_\star = \alpha_{\mathrm{KM}}$ in
null model

←- - - - - - - - - - - - - - - - - -

Follows if TAP heuristic proven

$\alpha_\star = \alpha_{\mathrm{KM}}$ in
planted model

Proof roadmap

Null model:
$\boldsymbol{G}$ iid gaussian

Conjecturally similar
under TAP heuristic

Our work proves this. . .

Planted model:
$\boldsymbol{G}$ cond on
$\mathrm{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$

DS18: lower bd by
1st/2nd mmt on
truncation of $|\boldsymbol{S}|$

1st/2nd moment
method
(direct calculation,
next slide)

$\alpha_\star = \alpha_{\mathsf{KM}}$ in
null model

Follows if TAP heuristic proven

$\alpha_\star = \alpha_{\mathsf{KM}}$ in
planted model

Proof roadmap

Null model:
G iid gaussian

Conjecturally similar under TAP heuristic

Our work proves this...

Planted model:
G cond on $\mathrm{TAP}(G; m, n)$

DS18: lower bd by 1st/2nd mmt on **truncation** of $|S|$

1st/2nd moment method (direct calculation, next slide)

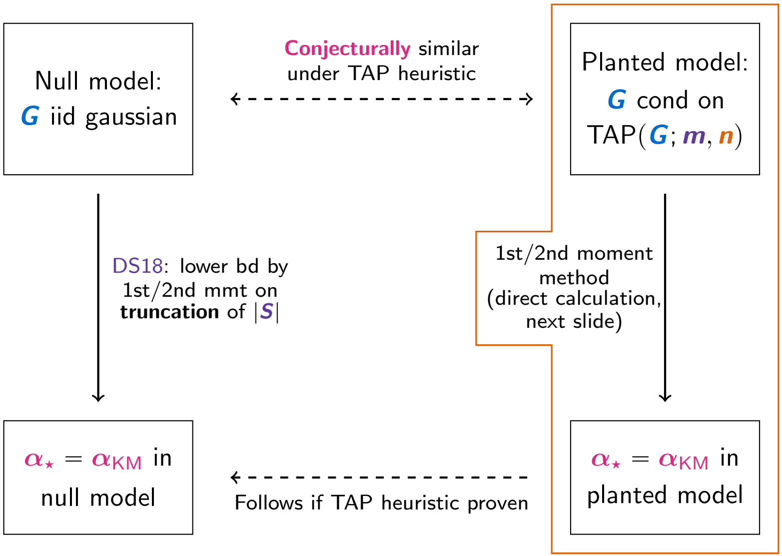
$\alpha_\star = \alpha_{\mathrm{KM}}$ in null model

...justifying this implication

Follows if TAP heuristic proven

$\alpha_\star = \alpha_{\mathrm{KM}}$ in planted model

Proof roadmap



1st/2nd moment works in planted model

Recall planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law
- Sample $\boldsymbol{G}$ conditional on $\dot{F}^{-1}(\boldsymbol{m}) = \frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}$, $\widehat{F}^{-1}(\boldsymbol{n}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}$

1st/2nd moment works in planted model

Recall planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law
- Sample $\boldsymbol{G}$ conditional on $\dot{F}^{-1}(\boldsymbol{m}) = \frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}$, $\widehat{F}^{-1}(\boldsymbol{n}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}$

This is **linear** constraint on $\boldsymbol{G}$

1st/2nd moment works in planted model

Recall planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law
- Sample $\boldsymbol{G}$ conditional on $\dot{F}^{-1}(\boldsymbol{m}) = \frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}$, $\widehat{F}^{-1}(\boldsymbol{n}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}$

This is **linear** constraint on $\boldsymbol{G}$ $\Rightarrow$ conditional on $(\boldsymbol{m}, \boldsymbol{n})$, $\boldsymbol{G}$ gaussian!

1st/2nd moment works in planted model

Recall planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law
- Sample $\boldsymbol{G}$ conditional on $\dot{F}^{-1}(\boldsymbol{m}) = \frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}$, $\widehat{F}^{-1}(\boldsymbol{n}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}$

This is **linear** constraint on $\boldsymbol{G}$ $\Rightarrow$ conditional on $(\boldsymbol{m}, \boldsymbol{n})$, $\boldsymbol{G}$ gaussian!

$\Rightarrow$ conditional moments of $|\boldsymbol{S}(\alpha N)|$ remain tractable.

1st/2nd moment works in planted model

Recall planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law
- Sample $\boldsymbol{G}$ conditional on $\dot{F}^{-1}(\boldsymbol{m}) = \frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}$, $\widehat{F}^{-1}(\boldsymbol{n}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}$

This is **linear** constraint on $\boldsymbol{G}$ $\Rightarrow$ conditional on $(\boldsymbol{m}, \boldsymbol{n})$, $\boldsymbol{G}$ gaussian!

$\Rightarrow$ conditional moments of $|\boldsymbol{S}(\alpha N)|$ remain tractable. For typical $(\boldsymbol{m}, \boldsymbol{n})$,

$$\mathbb{E}[|\boldsymbol{S}(\alpha N)||\boldsymbol{m}, \boldsymbol{n}] \quad \approx \quad \mathbb{E}[|\boldsymbol{S}(\alpha N)|^2|\boldsymbol{m}, \boldsymbol{n}]^{1/2}$$

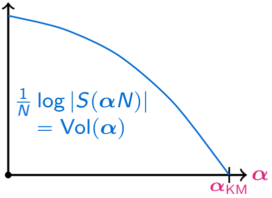
1st/2nd moment works in planted model

Recall planted model:
- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law
- Sample $\boldsymbol{G}$ conditional on $\dot{F}^{-1}(\boldsymbol{m}) = \frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}$, $\widehat{F}^{-1}(\boldsymbol{n}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}$

This is **linear** constraint on $\boldsymbol{G}$ $\Rightarrow$ conditional on $(\boldsymbol{m}, \boldsymbol{n})$, $\boldsymbol{G}$ gaussian!

$\Rightarrow$ conditional moments of $|\boldsymbol{S}(\alpha N)|$ remain tractable. For typical $(\boldsymbol{m}, \boldsymbol{n})$,

$$\mathbb{E}[|\boldsymbol{S}(\alpha N)||\boldsymbol{m}, \boldsymbol{n}] \quad \approx \quad \mathbb{E}[|\boldsymbol{S}(\alpha N)|^2|\boldsymbol{m}, \boldsymbol{n}]^{1/2} \quad \approx \quad \exp(N \,\mathrm{Vol}(\alpha))$$



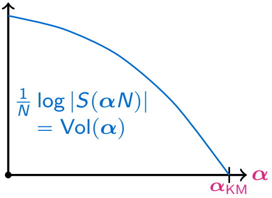
1st/2nd moment works in planted model

Recall planted model:
- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law
- Sample $\boldsymbol{G}$ conditional on $\dot{F}^{-1}(\boldsymbol{m}) = \frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}$, $\widehat{F}^{-1}(\boldsymbol{n}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}$

This is **linear** constraint on $\boldsymbol{G}$ $\Rightarrow$ conditional on $(\boldsymbol{m}, \boldsymbol{n})$, $\boldsymbol{G}$ gaussian!

$\Rightarrow$ conditional moments of $|\boldsymbol{S}(\alpha N)|$ remain tractable. For typical $(\boldsymbol{m}, \boldsymbol{n})$,

$$\mathbb{E}[|\boldsymbol{S}(\alpha N)||\boldsymbol{m}, \boldsymbol{n}] \quad \approx \quad \mathbb{E}[|\boldsymbol{S}(\alpha N)|^2|\boldsymbol{m}, \boldsymbol{n}]^{1/2} \quad \approx \quad \exp(N\,\mathrm{Vol}(\alpha))$$



$\Rightarrow$ planted model has capacity α_{KM}

1st/2nd moment works in planted model

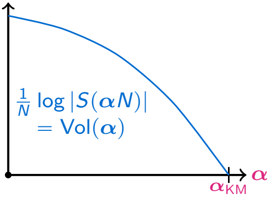
Recall planted model:

- Sample $(\boldsymbol{m}, \boldsymbol{n})$ from its law
- Sample $\boldsymbol{G}$ conditional on $\dot{F}^{-1}(\boldsymbol{m}) = \frac{\boldsymbol{G}^\top \boldsymbol{n}}{\sqrt{N}} - d\boldsymbol{m}$, $\widehat{F}^{-1}(\boldsymbol{n}) = \frac{\boldsymbol{G}\boldsymbol{m}}{\sqrt{N}} - b\boldsymbol{n}$

This is **linear** constraint on $\boldsymbol{G}$ $\Rightarrow$ conditional on $(\boldsymbol{m}, \boldsymbol{n})$, $\boldsymbol{G}$ gaussian!

$\Rightarrow$ conditional moments of $|\boldsymbol{S}(\alpha N)|$ remain tractable. For typical $(\boldsymbol{m}, \boldsymbol{n})$,

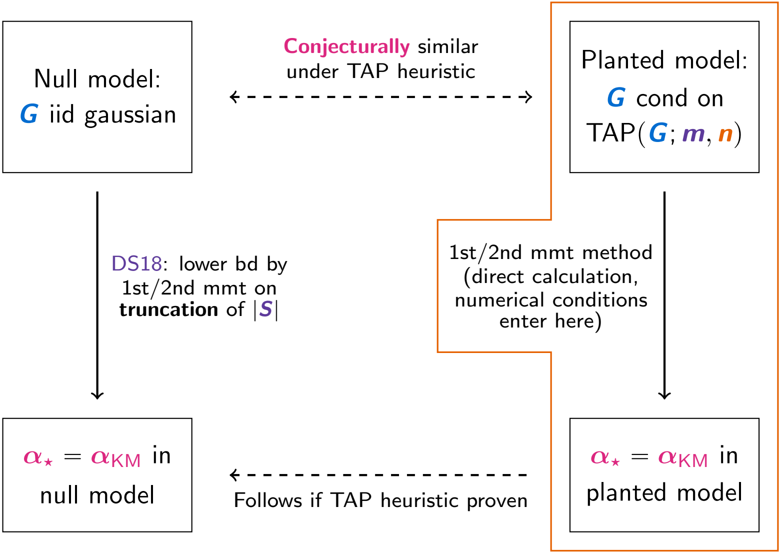
$$\mathbb{E}[|\boldsymbol{S}(\alpha N)||\boldsymbol{m}, \boldsymbol{n}] \quad \approx \quad \mathbb{E}[|\boldsymbol{S}(\alpha N)|^2|\boldsymbol{m}, \boldsymbol{n}]^{1/2} \quad \approx \quad \exp(N\,\mathrm{Vol}(\alpha))$$



$\Rightarrow$ planted model has capacity α_{KM}

(under our + DS18's numerical conditions)

Proof roadmap



Proof roadmap

Null model: G iid gaussian

Conjecturally similar under TAP heuristic

We now explain how to make this step rigorous.

Planted model: G cond on $\mathrm{TAP}(G; m, n)$

DS18: lower bd by 1st/2nd mmt on **truncation** of $|S|$

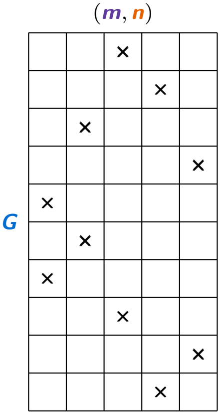
1st/2nd mmt method (direct calculation, numerical conditions enter here)

$\alpha_\star = \alpha_{\mathrm{KM}}$ in null model

Follows if TAP heuristic proven

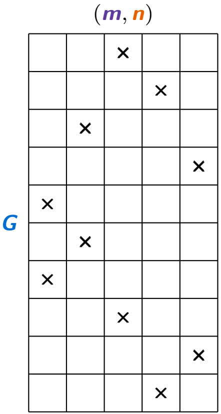
$\alpha_\star = \alpha_{\mathrm{KM}}$ in planted model

Key issue: linking true and planted models



× indicates $(\boldsymbol{m}, \boldsymbol{n})$ is TAP fixed point of $\boldsymbol{G}$

Key issue: linking true and planted models

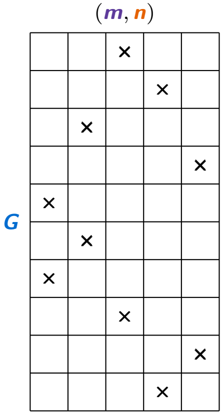


× indicates (m, n) is TAP fixed point of G

Null model: random row

Planted model: random col, then random × in col

Key issue: linking true and planted models

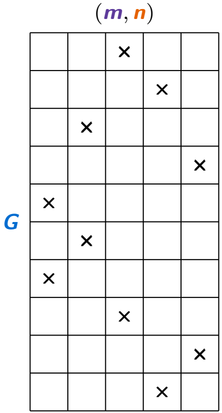


$\times$ indicates (m, n) is TAP fixed point of G

Null model: random row
Planted model: random col, then random $\times$ in col

TAP prediction: most rows have exactly one $\times$
so null $\approx$ planted

Key issue: linking true and planted models



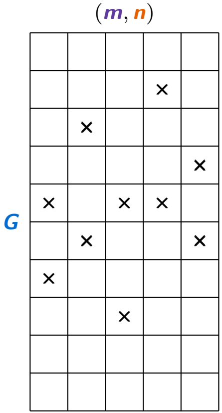
× indicates $(\boldsymbol{m}, \boldsymbol{n})$ is TAP fixed point of $\boldsymbol{G}$

Null model: random row
Planted model: random col, then random × in col

TAP prediction: most rows have exactly one ×
so null ≈ planted

but. . . **we don't actually know this** ☹

Key issue: linking true and planted models



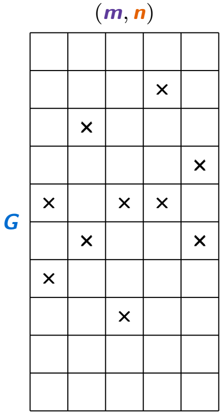
× indicates $(\boldsymbol{m}, \boldsymbol{n})$ is TAP fixed point of $\boldsymbol{G}$

Null model: random row
Planted model: random col, then random × in col

TAP prediction: most rows have exactly one ×
so null $\approx$ planted

but... **we don't actually know this** ☹

Key issue: linking true and planted models



× indicates (m, n) is TAP fixed point of G

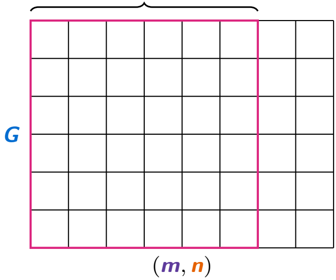
Null model: random row
Planted model: random col, then random × in col

TAP prediction: most rows have exactly one ×
so null ≈ planted

but. . . **we don't actually know this** ☹
⇒ planted / null models can a priori be different

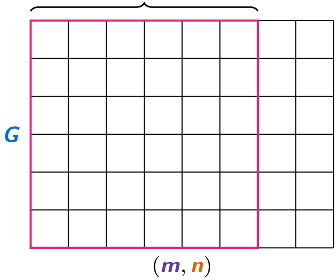
Our work: contiguity of null / planted models

$\boldsymbol{T} = \{\text{“typical”}\ (\boldsymbol{m}, \boldsymbol{n})\}$ (suitably defined set; whp in planted model)



Our work: contiguity of null / planted models

$T = \{$"typical" $(m, n)\}$ (suitably defined set; whp in planted model)

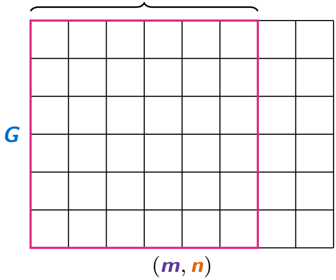


We show, for $G \sim$ null model:

- Existence: G has TAP fixed pt $(m, n) \in T$ whp (most rows have a × in T)

Our work: contiguity of null / planted models

$T = \{\text{"typical" } (m, n)\}$ (suitably defined set; whp in planted model)

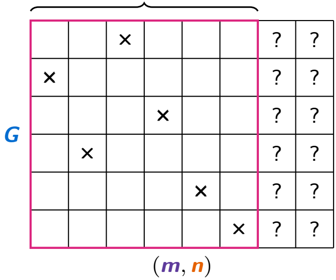


We show, for $G \sim$ null model:

- Existence: G has TAP fixed pt $(m, n) \in T$ whp (most rows have a × in T)
- Uniqueness: $\mathbb{E}[\#\text{TAP fixed pts in } T] = 1 + o(1)$ (on average, $1 + o(1)$ ×'s in T per row)

Our work: contiguity of null / planted models

$\boldsymbol{T} = \{$"typical" $(\boldsymbol{m}, \boldsymbol{n})\}$ (suitably defined set; whp in planted model)

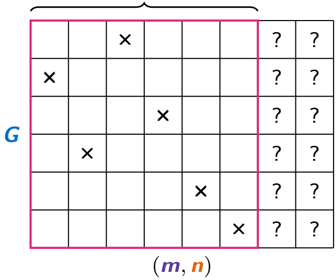


We show, for $\boldsymbol{G} \sim$ null model:

- Existence: $\boldsymbol{G}$ has TAP fixed pt $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T}$ whp (most rows have a × in $\boldsymbol{T}$)
- Uniqueness: $\mathbb{E}[\#\text{TAP fixed pts in } \boldsymbol{T}] = 1 + o(1)$ (on average, $1 + o(1)$ ×'s in $\boldsymbol{T}$ per row)

Our work: contiguity of null / planted models

$\boldsymbol{T} = \{\text{“typical”}\ (\boldsymbol{m}, \boldsymbol{n})\}$ (suitably defined set; whp in planted model)



We show, for $\boldsymbol{G} \sim$ null model:

- Existence: $\boldsymbol{G}$ has TAP fixed pt $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T}$ whp (most rows have a × in $\boldsymbol{T}$)
- Uniqueness: $\mathbb{E}[\#\text{TAP fixed pts in } \boldsymbol{T}] = 1 + o(1)$ (on average, $1 + o(1)$ ×'s in $\boldsymbol{T}$ per row)

This shows null $\approx$ planted. Formally,

$$\mathbb{P}_{\text{null}}(\boldsymbol{E}) \leqslant O(1) \cdot \sup_{(\boldsymbol{m},\boldsymbol{n}) \in \boldsymbol{T}} \mathbb{P}_{\text{planted}}(\boldsymbol{E} | \boldsymbol{m}, \boldsymbol{n}) + o(1) \qquad \text{for all event } \boldsymbol{E}$$

Existence: algorithmic proof

Goal: $\boldsymbol{G} \sim$ null model has TAP fixed pt $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T}$ whp

Existence: algorithmic proof

Goal: $\boldsymbol{G} \sim$ null model has TAP fixed pt $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T}$ whp

Approximate message passing (AMP) finds such a point:

$$\boldsymbol{m}^{k+1} = \dot{F}\left(\frac{\boldsymbol{G}^\top \boldsymbol{n}^k}{\sqrt{N}} - d\boldsymbol{m}^k\right) \qquad \boldsymbol{n}^k = \widehat{F}\left(\frac{\boldsymbol{G}\boldsymbol{m}^k}{\sqrt{N}} - b\boldsymbol{n}^{k-1}\right)$$

Existence: algorithmic proof

Goal: $\boldsymbol{G} \sim$ null model has TAP fixed pt $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T}$ whp

Approximate message passing (AMP) finds such a point:

$$\boldsymbol{m}^{k+1} = \dot{F}\left(\frac{\boldsymbol{G}^\top \boldsymbol{n}^k}{\sqrt{N}} - d\,\boldsymbol{m}^k\right) \qquad \boldsymbol{n}^k = \widehat{F}\left(\frac{\boldsymbol{G}\,\boldsymbol{m}^k}{\sqrt{N}} - b\,\boldsymbol{n}^{k-1}\right)$$

Follows from existing tools to analyze AMP:

- **AMP state evolution** (Bayati Montanari 11, Bolthausen 14, ...)
- **Local concavity of TAP free energy** near late AMP iterates (Celentano Fan Mei 21, Celentano 22, Celentano Fan Lin Mei 23)

Uniqueness: double-counting argument

Goal: for $\boldsymbol{G} \sim$ null model, $\mathbb{E}[\#\text{TAP fixed pts of } \boldsymbol{G} \text{ in } \boldsymbol{T}] = 1 + o(1)$

Uniqueness: double-counting argument

Goal: for $\boldsymbol{G} \sim$ null model, $\mathbb{E}[\#\text{TAP fixed pts of } \boldsymbol{G} \text{ in } \boldsymbol{T}] = 1 + o(1)$

This also has algorithmic proof! Following claim implies uniqueness:

Uniqueness: double-counting argument

Goal: for $\boldsymbol{G} \sim$ null model, $\mathbb{E}[\#\text{TAP fixed pts of } \boldsymbol{G} \text{ in } \boldsymbol{T}] = 1 + o(1)$

This also has algorithmic proof! Following claim implies uniqueness:

Fix $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T} = \{\text{typical pts}\}$. Sample $\boldsymbol{G}$ conditioned on $\text{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$.

AMP run on $\boldsymbol{G}$ finds **the planted point** $(\boldsymbol{m}, \boldsymbol{n})$ whp

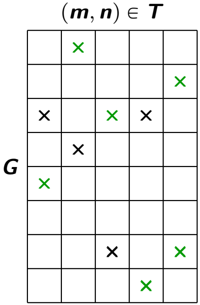
Uniqueness: double-counting argument

Goal: for $\boldsymbol{G} \sim$ null model, $\mathbb{E}[\#\text{TAP fixed pts of } \boldsymbol{G} \text{ in } \boldsymbol{T}] = 1 + o(1)$

This also has algorithmic proof! Following claim implies uniqueness:

Fix $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T} = \{\text{typical pts}\}$. Sample $\boldsymbol{G}$ conditioned on $\text{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$.

AMP run on $\boldsymbol{G}$ finds **the planted point** $(\boldsymbol{m}, \boldsymbol{n})$ whp



× : $(\boldsymbol{m}, \boldsymbol{n})$ TAP fixed pt of $\boldsymbol{G}$
× : subset of × where $\text{AMP}(\boldsymbol{G})$ finds $(\boldsymbol{m}, \boldsymbol{n})$

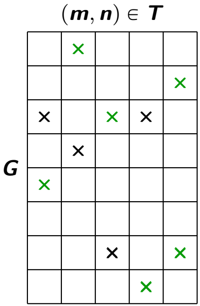
Uniqueness: double-counting argument

Goal: for $\boldsymbol{G} \sim$ null model, $\mathbb{E}[\#\text{TAP fixed pts of } \boldsymbol{G} \text{ in } \boldsymbol{T}] = 1 + o(1)$

This also has algorithmic proof! Following claim implies uniqueness:

Fix $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T} = \{\text{typical pts}\}$. Sample $\boldsymbol{G}$ conditioned on $\text{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$.

AMP run on $\boldsymbol{G}$ finds **the planted point** $(\boldsymbol{m}, \boldsymbol{n})$ whp



× : $(\boldsymbol{m}, \boldsymbol{n})$ TAP fixed pt of $\boldsymbol{G}$

× : subset of × where $\text{AMP}(\boldsymbol{G})$ finds $(\boldsymbol{m}, \boldsymbol{n})$

$\Rightarrow$ at most 1 × per row

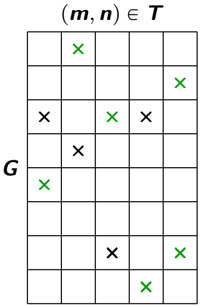
Uniqueness: double-counting argument

Goal: for $\boldsymbol{G} \sim$ null model, $\mathbb{E}[\#\text{TAP fixed pts of } \boldsymbol{G} \text{ in } \boldsymbol{T}] = 1 + o(1)$

💡 This also has algorithmic proof! Following claim implies uniqueness:

Fix $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T} = \{\text{typical pts}\}$. Sample $\boldsymbol{G}$ conditioned on $\text{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$.

AMP run on $\boldsymbol{G}$ finds **the planted point** $(\boldsymbol{m}, \boldsymbol{n})$ whp



× : $(\boldsymbol{m}, \boldsymbol{n})$ TAP fixed pt of $\boldsymbol{G}$

× : subset of × where $\text{AMP}(\boldsymbol{G})$ finds $(\boldsymbol{m}, \boldsymbol{n})$

$\Rightarrow$ at most 1 × per row

Claim $\Rightarrow$ in each col, fraction of green × $= 1 - o(1)$

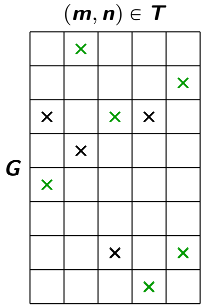
Uniqueness: double-counting argument

Goal: for $\boldsymbol{G} \sim$ null model, $\mathbb{E}[\#\text{TAP fixed pts of } \boldsymbol{G} \text{ in } \boldsymbol{T}] = 1 + o(1)$

This also has algorithmic proof! Following claim implies uniqueness:

Fix $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T} = \{\text{typical pts}\}$. Sample $\boldsymbol{G}$ conditioned on $\text{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$.

AMP run on $\boldsymbol{G}$ finds **the planted point** $(\boldsymbol{m}, \boldsymbol{n})$ whp



× : $(\boldsymbol{m}, \boldsymbol{n})$ TAP fixed pt of $\boldsymbol{G}$

× : subset of × where $\text{AMP}(\boldsymbol{G})$ finds $(\boldsymbol{m}, \boldsymbol{n})$

$\Rightarrow$ at most 1 × per row

Claim $\Rightarrow$ in each col, fraction of green × $= 1 - o(1)$

$\Rightarrow$ in whole grid, fraction of green × $= 1 - o(1)$

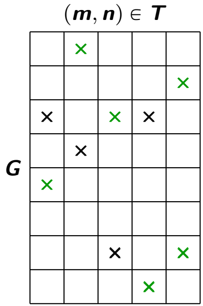
Uniqueness: double-counting argument

Goal: for $\boldsymbol{G} \sim$ null model, $\mathbb{E}[\#\text{TAP fixed pts of } \boldsymbol{G} \text{ in } \boldsymbol{T}] = 1 + o(1)$

This also has algorithmic proof! Following claim implies uniqueness:

Fix $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T} = \{\text{typical pts}\}$. Sample $\boldsymbol{G}$ conditioned on $\text{TAP}(\boldsymbol{G}; \boldsymbol{m}, \boldsymbol{n})$.

AMP run on $\boldsymbol{G}$ finds **the planted point** $(\boldsymbol{m}, \boldsymbol{n})$ whp



× : $(\boldsymbol{m}, \boldsymbol{n})$ TAP fixed pt of $\boldsymbol{G}$

× : subset of × where $\text{AMP}(\boldsymbol{G})$ finds $(\boldsymbol{m}, \boldsymbol{n})$

$\Rightarrow$ at most 1 × per row

Claim $\Rightarrow$ in each col, fraction of green × $= 1 - o(1)$

$\Rightarrow$ in whole grid, fraction of green × $= 1 - o(1)$

$\Rightarrow$ on average, at most $1 + o(1)$ ×'s per row

Uniqueness: AMP returns home in planted model

Remains to show: for $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T}$, $\boldsymbol{G}$ conditioned on $\mathrm{TAP}(\boldsymbol{G}, \boldsymbol{m}, \boldsymbol{n})$,

AMP run on $\boldsymbol{G}$ finds **the planted point** $(\boldsymbol{m}, \boldsymbol{n})$ whp

Uniqueness: AMP returns home in planted model

Remains to show: for $(\boldsymbol{m}, \boldsymbol{n}) \in \boldsymbol{T}$, $\boldsymbol{G}$ conditioned on $\mathrm{TAP}(\boldsymbol{G}, \boldsymbol{m}, \boldsymbol{n})$,

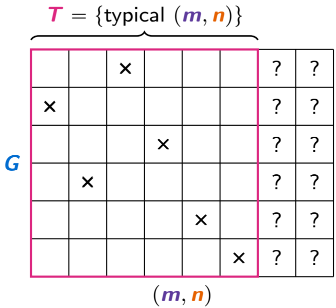
AMP run on $\boldsymbol{G}$ finds **the planted point** $(\boldsymbol{m}, \boldsymbol{n})$ whp

---

This can be proved by the same AMP state evolution + local concavity of TAP free energy analyses.

Crucially: recall $\mathrm{Law}_{\mathrm{planted}}(\boldsymbol{G} \mid \boldsymbol{m}, \boldsymbol{n})$ remains gaussian. This provides enough structure to adapt these techniques.

Recap: contiguity of null / planted models

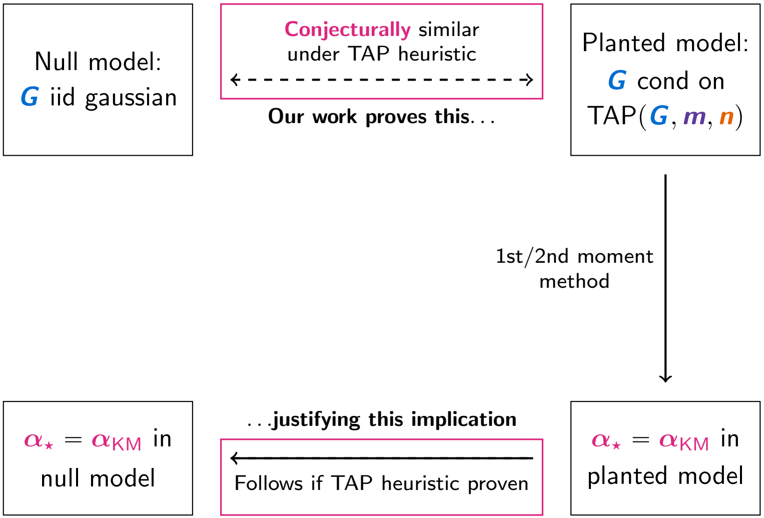


We show, for $G \sim$ null model:

- Existence: G has TAP fixed pt $(m, n) \in T$ whp (most rows have a × in T)
- Uniqueness: $\mathbb{E}[\#\text{TAP fixed pts in } T] = 1 + o(1)$ (on average, $1 + o(1)$ ×'s in T per row)

This shows null $\approx$ planted.

Recap: proof roadmap



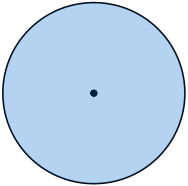
Other uses of TAP planting

"AMP returns home in planted model $\rightarrow$ uniqueness" is general method, enables passing to TAP planted model

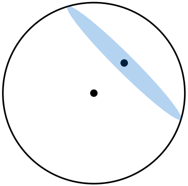
Other uses of TAP planting

"AMP returns home in planted model → uniqueness" is general method, enables passing to TAP planted model

Centered (and RS) Gibbs measures are simpler than non-centered ones:



measure μ has mean $\mathbf{0}$ **(easy)**

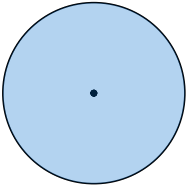


measure μ concentrates on band with **random** center, which is a TAP fixed pt **(hard)**

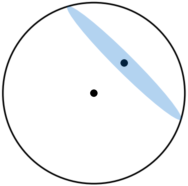
Other uses of TAP planting

"AMP returns home in planted model → uniqueness" is general method, enables passing to TAP planted model

Centered (and RS) Gibbs measures are simpler than non-centered ones:



measure μ has mean **0**
(easy)



measure μ concentrates on band with **random** center, which is a TAP fixed pt **(hard)**

TAP planting lets you **condition on the random center**, effectively reducing to the mean-zero case.

Other uses of TAP planting

"AMP returns home in planted model → uniqueness" is general method, enables passing to TAP planted model

Centered (and RS) Gibbs measures are simpler than non-centered ones:

measure μ has mean $\mathbf{0}$ **(easy)**

measure μ concentrates on band with **random** center, which is a TAP fixed pt **(hard)**

TAP planting lets you **condition on the random center**, effectively reducing to the mean-zero case. Usages in spin glass sampling:

- High-precision estimation of mean(μ) (H Montanari Pham 24)
- Covariance bound $\|\mathrm{cov}(\mu)\|_{\mathrm{op}} = O(1)$ (H Mohanty Rajaraman Wu 24)

Conclusion

Null model:

- $H \sim \text{Law}(\text{problem})$
- $\boldsymbol{\sigma} \sim \text{Gibbs}(H)$ **(hard)**

Planted model:

- $\boldsymbol{\sigma} \sim \text{uniform}$
- $H \sim \text{Law}(\text{problem} \mid \boldsymbol{\sigma})$ **(easy)**

Conclusion

Null model:

- $H \sim \text{Law}(\text{problem})$
- $\boldsymbol{\sigma} \sim \text{Gibbs}(H)$ **(hard)**

Planted model:

- $\boldsymbol{\sigma} \sim \text{uniform}$
- $H \sim \text{Law}(\text{problem} \,|\, \boldsymbol{\sigma})$ **(easy)**

Classic planting trick: if null / planted models are contiguous, often easier to prove things about planted model.

Conclusion

Null model:

- $H \sim \text{Law}(\text{problem})$
- $\boldsymbol{\sigma} \sim \text{Gibbs}(H)$ **(hard)**

Planted model:

- $\boldsymbol{\sigma} \sim \text{uniform}$
- $H \sim \text{Law}(\text{problem} \mid \boldsymbol{\sigma})$ **(easy)**

Classic planting trick: if null / planted models are contiguous,
often easier to prove things about planted model.

Can also plant other objects, like critical points or TAP fixed points.

Conclusion

Null model:

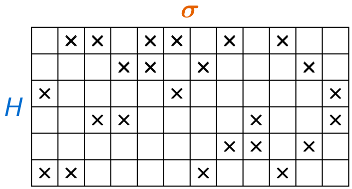
- $H \sim \text{Law}(\text{problem})$
- $\boldsymbol{\sigma} \sim \text{Gibbs}(H)$ **(hard)**

Planted model:

- $\boldsymbol{\sigma} \sim \text{uniform}$
- $H \sim \text{Law}(\text{problem} \,|\, \boldsymbol{\sigma})$ **(easy)**

Classic planting trick: if null / planted models are contiguous, often easier to prove things about planted model.

Can also plant other objects, like critical points or TAP fixed points.



Applications:

- shattering & RS free energy of many models
- spin glass diffusion sampling
- ground state large deviation & 1RSB ground state energy
- capacity of Ising perceptron

Part II: a survey on the overlap gap property

Part II: a survey on the overlap gap property

Outline of part II: a survey on the overlap gap property

Introduction and motivating problems

Overlap gap property: the basics

More OGPs and algorithm classes

Further enhancements

Hardness of finding strict local maxima

Strong low degree hardness

Random optimization problems

How well can an efficient algorithm optimize a random objective?

Random optimization problems

How well can an efficient algorithm optimize a random objective?

Extensively studied models of disordered objectives:

Random optimization problems

How well can an efficient algorithm optimize a random objective?
Extensively studied models of disordered objectives:

• **Constraint satisfaction**: max cut / max independent set on $G(N,p)$, random instances of (max)-k-SAT, . . .

Random optimization problems

How well can an efficient algorithm optimize a random objective?
Extensively studied models of disordered objectives:

- **Constraint satisfaction**: max cut / max independent set on $G(N, p)$, random instances of (max)-k-SAT, ...

- **Mean-field spin glass**: polynomial with IID gaussian coefs, e.g. cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k} \sigma_i \sigma_j \sigma_k \,, \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

Maximize over domain $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising)

Random optimization problems

How well can an efficient algorithm optimize a random objective?
Extensively studied models of disordered objectives:

• **Constraint satisfaction**: max cut / max independent set on $G(N, p)$, random instances of (max)-k-SAT, ...

• **Mean-field spin glass**: polynomial with IID gaussian coefs, e.g. cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k} \sigma_i \sigma_j \sigma_k , \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

Maximize over domain $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising)
Degree 2 on S_N is eigenvalue problem, but deg $\geqslant 3$ is highly nonconvex.

Random optimization problems

How well can an efficient algorithm optimize a random objective?
Extensively studied models of disordered objectives:

- **Constraint satisfaction**: max cut / max independent set on $G(N, p)$, random instances of (max)-k-SAT, ...

- **Mean-field spin glass**: polynomial with IID gaussian coefs, e.g. cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k} \sigma_i \sigma_j \sigma_k , \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

Maximize over domain $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising)
Degree 2 on S_N is eigenvalue problem, but deg $\geqslant 3$ is highly nonconvex.

- **Random perceptron**: for IID $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^M \sim \mathcal{N}(0, I_N)$, $\underset{\text{activation}}{\underset{\uparrow}{\varphi}} : \mathbb{R} \to \mathbb{R}$,

$$H(\boldsymbol{\sigma}) = \sum_{a=1}^{M} \varphi\Big(\frac{(\boldsymbol{\sigma}, \boldsymbol{g}^a)}{\sqrt{N}}\Big)$$

Random optimization problems: motivation

- MLE in statistical tasks, e.g. **tensor PCA**: estimate $\boldsymbol{x}_0 \sim \text{unif}(\mathbb{S}^{N-1})$ from

$$\boldsymbol{T} = \lambda \boldsymbol{x}_0^{\otimes p} + \boldsymbol{G}^{(p)}, \qquad \boldsymbol{G}^{(p)} \in (\mathbb{R}^N)^{\otimes p} \text{ has i.i.d. } \mathcal{N}(0,1) \text{ entries}$$

Random optimization problems: motivation

• MLE in statistical tasks, e.g. **tensor PCA**: estimate $\boldsymbol{x}_0 \sim \text{unif}(\mathbb{S}^{N-1})$ from

$$\boldsymbol{T} = \lambda \boldsymbol{x}_0^{\otimes p} + \boldsymbol{G}^{(p)}, \qquad \boldsymbol{G}^{(p)} \in (\mathbb{R}^N)^{\otimes p} \text{ has i.i.d. } \mathcal{N}(0,1) \text{ entries}$$

Max-likelihood estimator is **random, non-convex** optimization:

$$\boldsymbol{x}_{\text{MLE}} = \arg\max_{\boldsymbol{x} \in \mathbb{S}^{N-1}} \langle \boldsymbol{T}, \boldsymbol{x}^{\otimes p} \rangle.$$

Random optimization problems: motivation

- MLE in statistical tasks, e.g. **tensor PCA**: estimate $\boldsymbol{x}_0 \sim \text{unif}(\mathbb{S}^{N-1})$ from

$$\boldsymbol{T} = \lambda \boldsymbol{x}_0^{\otimes p} + \boldsymbol{G}^{(p)}, \qquad \boldsymbol{G}^{(p)} \in (\mathbb{R}^N)^{\otimes p} \text{ has i.i.d. } \mathcal{N}(0,1) \text{ entries}$$

Max-likelihood estimator is **random, non-convex** optimization:

$$\boldsymbol{x}_{\text{MLE}} = \arg\max_{\boldsymbol{x} \in \mathbb{S}^{N-1}} \langle \boldsymbol{T}, \boldsymbol{x}^{\otimes p} \rangle.$$

Mean-field spin glass is **null model** $\lambda = 0$.

Random optimization problems: motivation

• MLE in statistical tasks, e.g. **tensor PCA**: estimate $\boldsymbol{x}_0 \sim \mathrm{unif}(\mathbb{S}^{N-1})$ from

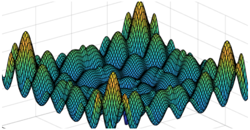
$$\boldsymbol{T} = \lambda \boldsymbol{x}_0^{\otimes p} + \boldsymbol{G}^{(p)}, \qquad \boldsymbol{G}^{(p)} \in (\mathbb{R}^N)^{\otimes p} \text{ has i.i.d. } \mathcal{N}(0,1) \text{ entries}$$

Max-likelihood estimator is **random, non-convex** optimization:

$$\boldsymbol{x}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{x} \in \mathbb{S}^{N-1}} \langle \boldsymbol{T}, \boldsymbol{x}^{\otimes p} \rangle.$$

Mean-field spin glass is **null model** $\lambda = 0$.

• Model of **neural network loss landscapes** (Choromanska Henaff Mathieu Ben Arous LeCun 15)



Random optimization problems: motivation

• MLE in statistical tasks, e.g. **tensor PCA**: estimate $\boldsymbol{x}_0 \sim \text{unif}(\mathbb{S}^{N-1})$ from

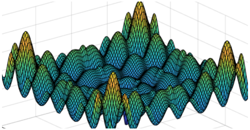
$$\boldsymbol{T} = \lambda \boldsymbol{x}_0^{\otimes p} + \boldsymbol{G}^{(p)}, \qquad \boldsymbol{G}^{(p)} \in (\mathbb{R}^N)^{\otimes p} \text{ has i.i.d. } \mathcal{N}(0,1) \text{ entries}$$

Max-likelihood estimator is **random, non-convex** optimization:

$$\boldsymbol{x}_{\text{MLE}} = \arg\max_{\boldsymbol{x} \in \mathbb{S}^{N-1}} \langle \boldsymbol{T}, \boldsymbol{x}^{\otimes p} \rangle.$$

Mean-field spin glass is **null model** $\lambda = 0$.

• Model of **neural network loss landscapes** (Choromanska Henaff Mathieu Ben Arous LeCun 15)



• Random perceptron ↔ loss landscape of neural net on **random data**

Random optimization problems: goals

Ground state: what is the max value of H that **exists** whp? (OPT)

Random optimization problems: goals

Ground state: what is the max value of H that **exists** whp? (OPT)

Optimization: given random realization of H, **algorithmically** find $\boldsymbol{\sigma}^{\mathsf{alg}}$ with $H(\boldsymbol{\sigma}^{\mathsf{alg}})$ as large as possible.

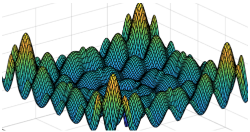
What value ALG can an efficient algorithm achieve? Does ALG = OPT?

Random optimization problems: goals

Ground state: what is the max value of H that **exists** whp? (OPT)

Optimization: given random realization of H, **algorithmically** find $\boldsymbol{\sigma}^{\mathrm{alg}}$ with $H(\boldsymbol{\sigma}^{\mathrm{alg}})$ as large as possible.

What value ALG can an efficient algorithm achieve? Does ALG = OPT?



Challenges:

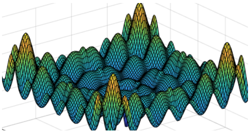
- Highly **nonconvex** landscape with $e^{\Omega(N)}$ maxima — what are good algorithms?

Random optimization problems: goals

Ground state: what is the max value of H that **exists** whp? (OPT)

Optimization: given random realization of H, **algorithmically** find $\boldsymbol{\sigma}^{\mathrm{alg}}$ with $H(\boldsymbol{\sigma}^{\mathrm{alg}})$ as large as possible.

What value ALG can an efficient algorithm achieve? Does ALG = OPT?



Challenges:

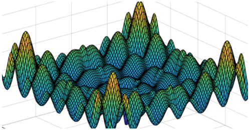
- Highly **nonconvex** landscape with $e^{\Omega(N)}$ maxima — what are good algorithms?
- **Average case** setting — how to reason about algorithmic hardness?

Random optimization problems: goals

Ground state: what is the max value of H that **exists** whp? (OPT)

Optimization: given random realization of H, **algorithmically** find $\boldsymbol{\sigma}^{\text{alg}}$ with $H(\boldsymbol{\sigma}^{\text{alg}})$ as large as possible.

What value ALG can an efficient algorithm achieve? Does ALG = OPT?



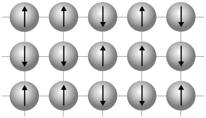
Challenges:

- Highly **nonconvex** landscape with $e^{\Omega(N)}$ maxima — what are good algorithms?
- **Average case** setting — how to reason about algorithmic hardness?

Sampling: algorithmically sample from **Gibbs measure** $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$. For which β can an efficient algorithm succeed?

Comparison with ferromagnetic Ising model

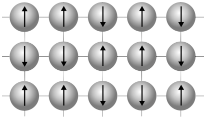
Ferromagnetic Ising: positive couplings on edges of a graph G



$$H^{\mathsf{Fer}}(\boldsymbol{\sigma}) = \sum_{(i,j)\in E(G)} \sigma_i \sigma_j$$

Comparison with ferromagnetic Ising model

Ferromagnetic Ising: positive couplings on edges of a graph G



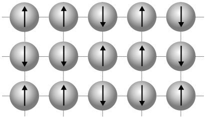
$$H^{\mathrm{Fer}}(\boldsymbol{\sigma}) = \sum_{(i,j)\in E(G)} \sigma_i \sigma_j$$

Main tension between **entropy** and **energy**. For $\mu_\beta(\boldsymbol{\sigma}) = \frac{1}{Z} e^{\beta H^{\mathrm{Fer}}(\boldsymbol{\sigma})}$

- β small $\Rightarrow$ entropy wins, coordinates of $\boldsymbol{\sigma} \sim \mu_\beta$ not aligned
- β large $\Rightarrow$ energy wins, $\boldsymbol{\sigma} \sim \mu_\beta$ aligns with $+\vec{1}$ or $-\vec{1}$

Comparison with ferromagnetic Ising model

Ferromagnetic Ising: positive couplings on edges of a graph G

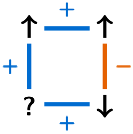


$$H^{\text{Fer}}(\boldsymbol{\sigma}) = \sum_{(i,j)\in E(G)} \sigma_i \sigma_j$$

Main tension between **entropy** and **energy**. For $\mu_\beta(\boldsymbol{\sigma}) = \frac{1}{Z} e^{\beta H^{\text{Fer}}(\boldsymbol{\sigma})}$

- β small $\Rightarrow$ entropy wins, coordinates of $\boldsymbol{\sigma} \sim \mu_\beta$ not aligned
- β large $\Rightarrow$ energy wins, $\boldsymbol{\sigma} \sim \mu_\beta$ aligns with $+\vec{1}$ or $-\vec{1}$

In spin glasses, random $g_{i,j}$ yield **frustration**: can't satisfy all couplings. A priori unclear what ground state looks like.



Comparison with signal recovery

Many similar problems about detecting / recovering a planted signal:

- **Planted clique**: find a k-clique planted in $G(N, 1/2)$
- **Tensor PCA**: recover rank 1 spike planted in gaussian p-tensor
- **Single/multi-index models**: recover $\boldsymbol{W}^*$ from $y_i = f(\boldsymbol{W}^* \boldsymbol{x}_i)$

Comparison with signal recovery

Many similar problems about detecting / recovering a planted signal:

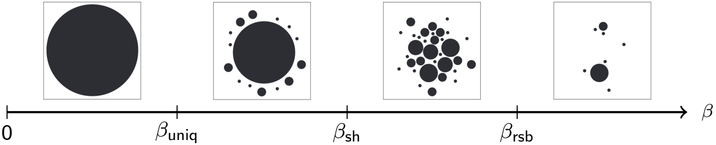
- **Planted clique**: find a k-clique planted in $G(N, 1/2)$
- **Tensor PCA**: recover rank 1 spike planted in gaussian p-tensor
- **Single/multi-index models**: recover $\boldsymbol{W}^*$ from $y_i = f(\boldsymbol{W}^* \boldsymbol{x}_i)$

The models we focus on are "pure noise," no planted signal

- **Null models** for signal recovery problems
- Progress can be made "in many directions"
- No notion of sample complexity / SNR

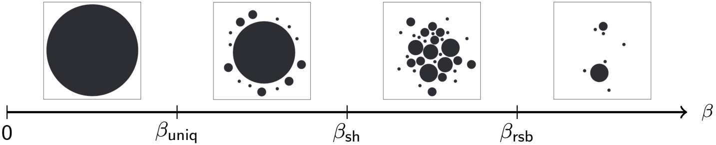
Some possible phases for disordered systems

Predictions of geometric phase transitions + algorithmic implications:



Some possible phases for disordered systems

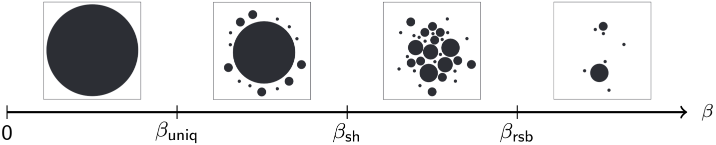
Predictions of geometric phase transitions + algorithmic implications:



- $\beta \in (0, \beta_{\mathrm{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality
- $\beta \in (\beta_{\mathrm{uniq}}, \beta_{\mathrm{sh}})$: rapid mixing from **random** but not worst-case start
- $\beta \in (\beta_{\mathrm{sh}}, \beta_{\mathrm{rsb}})$: μ_β **shatters** into $e^{\Omega(N)}$ clusters of mass $e^{-\Omega(N)}$
- $\beta \in (\beta_{\mathrm{rsb}}, \infty)$: $e^{\Omega(N)}$ clusters but largest $O(1)$ dominate the mass (**condensation / RSB**)

Some possible phases for disordered systems

Predictions of geometric phase transitions + algorithmic implications:



- $\beta \in (0, \beta_{\text{uniq}})$: dynamics exhibit **rapid mixing** & Poincaré inequality
- $\beta \in (\beta_{\text{uniq}}, \beta_{\text{sh}})$: rapid mixing from **random** but not worst-case start
- $\beta \in (\beta_{\text{sh}}, \beta_{\text{rsb}})$: μ_β **shatters** into $e^{\Omega(N)}$ clusters of mass $e^{-\Omega(N)}$
- $\beta \in (\beta_{\text{rsb}}, \infty)$: $e^{\Omega(N)}$ clusters but largest $O(1)$ dominate the mass (**condensation / RSB**)

Does solution geometry have rigorous implications for algorithms?

This talk: overlap gap property

Gamarnik Sudan 14: **solution landscape** properties → rigorous hardness for **stable** algorithms in random optimization / search problems

This talk: overlap gap property

Gamarnik Sudan 14: **solution landscape** properties → rigorous hardness for **stable** algorithms in random optimization / search problems

- **Stability**: view alg $\mathcal{A} : H \to \boldsymbol{\sigma}$ as function of problem instance H

This talk: overlap gap property

Gamarnik Sudan 14: **solution landscape** properties → rigorous hardness for **stable** algorithms in random optimization / search problems

- **Stability**: view alg $\mathcal{A} : H \to \boldsymbol{\sigma}$ as function of problem instance H

- Motivated by "**shattering → hardness**" intuition but formally independent (Mézard Mora Zecchina 05, Achlioptas Coja-Oghlan 08, Coja-Oghlan Efthymiou 10)

This talk: overlap gap property

Gamarnik Sudan 14: **solution landscape** properties → rigorous hardness for **stable** algorithms in random optimization / search problems

- **Stability**: view alg $\mathcal{A} : H \to \boldsymbol{\sigma}$ as function of problem instance H

- Motivated by "**shattering → hardness**" intuition but formally independent (Mézard Mora Zecchina 05, Achlioptas Coja-Oghlan 08, Coja-Oghlan Efthymiou 10)

- In hindsight: algorithmic threshold for **optimization** given by a **geometric phase transition**, but this threshold **cannot be sharply understood in terms of Gibbs measures**

This talk: overlap gap property

Gamarnik Sudan 14: **solution landscape** properties → rigorous hardness for **stable** algorithms in random optimization / search problems

• **Stability**: view alg $\mathcal{A} : H \to \boldsymbol{\sigma}$ as function of problem instance H

• Motivated by "**shattering → hardness**" intuition but formally independent (Mézard Mora Zecchina 05, Achlioptas Coja-Oghlan 08, Coja-Oghlan Efthymiou 10)

• In hindsight: algorithmic threshold for **optimization** given by a **geometric phase transition**, but this threshold **cannot be sharply understood in terms of Gibbs measures**

(Contrast: for **sampling**, shattering threshold β_{sh} appears to be the fundamental barrier; much recent progress)

Outline of part II: a survey on the overlap gap property

Introduction and motivating problems

Overlap gap property: the basics

More OGPs and algorithm classes

Further enhancements

Hardness of finding strict local maxima

Strong low degree hardness

Where it all started

Max independent set: find a large ind set of Erdős–Rényi $G(N, d/N)$
(d large but fixed, $N \to \infty$)

Where it all started

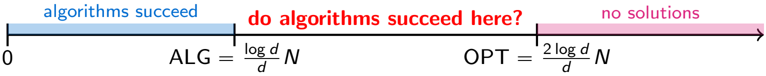
Max independent set: find a large ind set of Erdős–Rényi $G(N, d/N)$
(d large but fixed, $N \to \infty$)

- Largest ind set that **exists**: $(\mathbf{2} + o_d(1))\frac{\log d}{d}N$ (Frieze 90)
- Best known **algorithm** finds: $(\mathbf{1} + o_d(1))\frac{\log d}{d}N$ (trivial, greedy)

Where it all started

Max independent set: find a large ind set of Erdős–Rényi $G(N, d/N)$
(d large but fixed, $N \to \infty$)

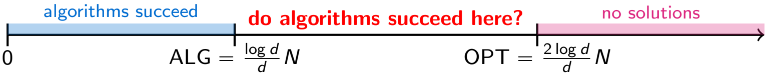
- Largest ind set that **exists**: $(\mathbf{2} + o_d(1))\frac{\log d}{d}N$ (Frieze 90)
- Best known **algorithm** finds: $(\mathbf{1} + o_d(1))\frac{\log d}{d}N$ (trivial, greedy)



Where it all started

Max independent set: find a large ind set of Erdős–Rényi $G(N, d/N)$
(d large but fixed, $N \to \infty$)

- Largest ind set that **exists**: $(\mathbf{2} + o_d(1))\frac{\log d}{d}N$ (Frieze 90)
- Best known **algorithm** finds: $(\mathbf{1} + o_d(1))\frac{\log d}{d}N$ (trivial, greedy)

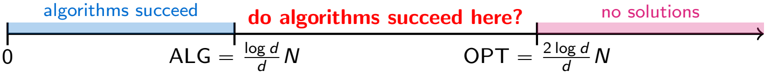


Hatami Lovász Szegedy 12 conjecture: **local algorithms** can $(1 - o(1))$-approximate OPT

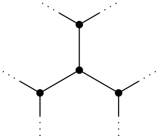
Where it all started

Max independent set: find a large ind set of Erdős–Rényi $G(N, d/N)$
(d large but fixed, $N \to \infty$)

- Largest ind set that **exists**: $(\mathbf{2} + o_d(1))\frac{\log d}{d}N$ (Frieze 90)
- Best known **algorithm** finds: $(\mathbf{1} + o_d(1))\frac{\log d}{d}N$ (trivial, greedy)



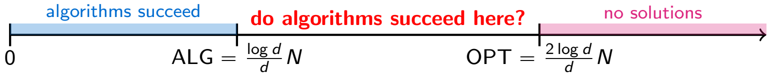
Hatami Lovász Szegedy 12 conjecture: **local algorithms** can $(1 - o(1))$-approximate OPT



Where it all started

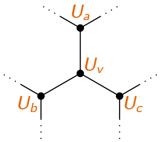
Max independent set: find a large ind set of Erdős–Rényi $G(N, d/N)$
(d large but fixed, $N \to \infty$)

- Largest ind set that **exists**: $(\mathbf{2} + o_d(1))\frac{\log d}{d}N$ (Frieze 90)
- Best known **algorithm** finds: $(\mathbf{1} + o_d(1))\frac{\log d}{d}N$ (trivial, greedy)



Hatami Lovász Szegedy 12 conjecture: **local algorithms** can $(1 - o(1))$-approximate OPT

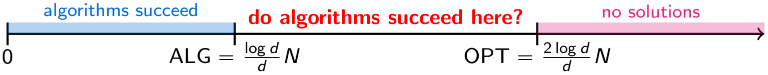
Local algorithm: generate $U_v \sim \text{unif}([0,1])$ at each $v \in G$ (shared randomness)



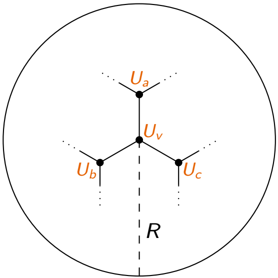
Where it all started

Max independent set: find a large ind set of Erdős–Rényi $G(N, d/N)$
(d large but fixed, $N \to \infty$)

- Largest ind set that **exists**: $(\mathbf{2} + o_d(1))\frac{\log d}{d}N$ (Frieze 90)
- Best known **algorithm** finds: $(\mathbf{1} + o_d(1))\frac{\log d}{d}N$ (trivial, greedy)



Hatami Lovász Szegedy 12 conjecture: **local algorithms** can $(1 - o(1))$-approximate OPT



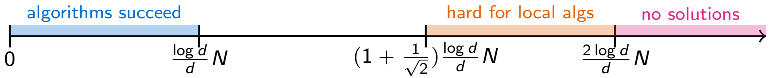
Local algorithm: generate $U_v \sim \text{unif}([0, 1])$ at each $v \in G$ (shared randomness)

At each $v \in G$, decide output $\sigma_v \in \{0, 1\}$ based on only data within R-**neighborhood** of v ($R = O(1)$)

Local algorithms do not reach OPT

Theorem (Gamarnik Sudan 14)

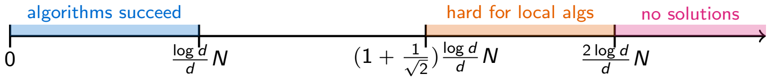
Any $O(1)$-local algorithm does **not** *find an independent set of size $\geqslant (\mathbf{1 + \frac{1}{\sqrt{2}}} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.*



Local algorithms do not reach OPT

Theorem (Gamarnik Sudan 14)

Any $O(1)$-local algorithm does **not** *find an independent set of size $\geqslant (\mathbf{1 + \frac{1}{\sqrt{2}}} + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$, whp.*

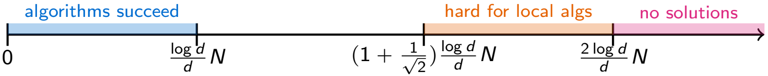


Consider correlated family of $G(N, d/N)$ indexed by $0 \leqslant t \leqslant T = \binom{N}{2}$:

Local algorithms do not reach OPT

Theorem (Gamarnik Sudan 14)

Any $O(1)$-local algorithm does **not** *find an independent set of size $\geqslant (\mathbf{1 + \frac{1}{\sqrt{2}}} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.*



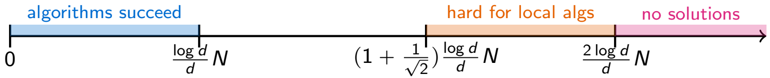
Consider correlated family of $G(N, d/N)$ indexed by $0 \leqslant t \leqslant T = \binom{N}{2}$:

Each $\boldsymbol{G}^t$ resamples one edge of $\boldsymbol{G}^{t-1}$ (so $\boldsymbol{G}^0, \boldsymbol{G}^T$ independent)

Local algorithms do not reach OPT

Theorem (Gamarnik Sudan 14)

Any $O(1)$-local algorithm does **not** *find an independent set of size $\geqslant (\mathbf{1 + \frac{1}{\sqrt{2}}} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.*



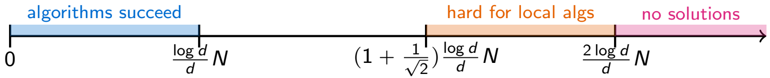
Consider correlated family of $G(N, d/N)$ indexed by $0 \leqslant t \leqslant T = \binom{N}{2}$:

Each $\boldsymbol{G}^t$ resamples one edge of $\boldsymbol{G}^{t-1}$ (so $\boldsymbol{G}^0, \boldsymbol{G}^T$ independent)

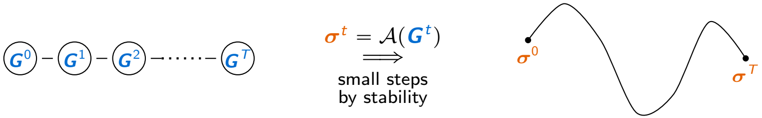
Local algorithms do not reach OPT

Theorem (Gamarnik Sudan 14)

Any $O(1)$-local algorithm does **not** *find an independent set of size $\geqslant (\mathbf{1 + \frac{1}{\sqrt{2}}} + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$, whp.*



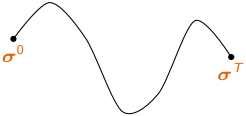
Consider correlated family of $G(N, d/N)$ indexed by $0 \leqslant t \leqslant T = \binom{N}{2}$:



Each $\boldsymbol{G}^t$ resamples one edge of $\boldsymbol{G}^{t-1}$ (so $\boldsymbol{G}^0, \boldsymbol{G}^T$ independent)

Local algorithms do not reach OPT

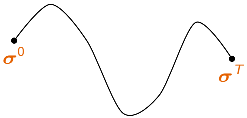
$$\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$$
$$\Longrightarrow$$
small steps
by stability



Local algorithms do not reach OPT

$\boldsymbol{G}^0 - \boldsymbol{G}^1 - \boldsymbol{G}^2 - \cdots\cdots - \boldsymbol{G}^T$

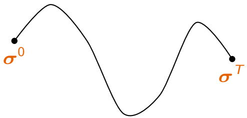
$\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$
$\Longrightarrow$
small steps
by stability



Suppose local $\mathcal{A}$ beats $(1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d} N$

Local algorithms do not reach OPT

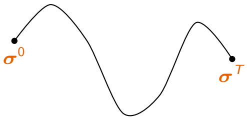
$G^0 - G^1 - G^2 - \cdots\cdots - G^T$ $\overset{\sigma^t = \mathcal{A}(G^t)}{\Longrightarrow}$ small steps by stability



Suppose local $\mathcal{A}$ beats $(1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N \Rightarrow$ each σ^t solves G^t
(i.e. is ind set of size $\geqslant (1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N$)

Local algorithms do not reach OPT

$\sigma^t = \mathcal{A}(G^t)$ $\Longrightarrow$ small steps by stability

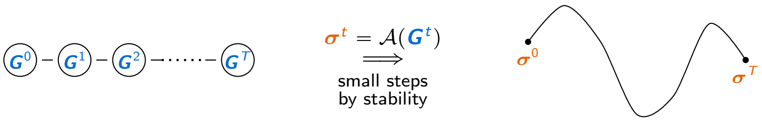


Suppose local $\mathcal{A}$ beats $(1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N \Rightarrow$ each σ^t solves G^t
(i.e. is ind set of size $\geqslant (1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N$)

Landscape obstruction : for "medium" q_{ogp}, there do not exist σ, ρ such that $\|\sigma - \rho\| = q_{\text{ogp}}$ and

σ solves G^{t_1} and ρ solves G^{t_2} for some t_1, t_2

Local algorithms do not reach OPT



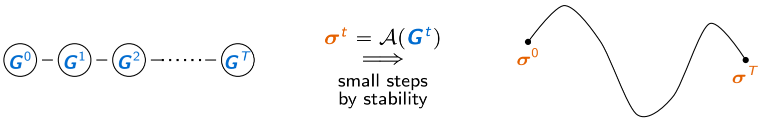
Suppose local $\mathcal{A}$ beats $(1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N \Rightarrow$ each $\boldsymbol{\sigma}^t$ solves $\boldsymbol{G}^t$
(i.e. is ind set of size $\geqslant (1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N$)

Landscape obstruction : for "medium" q_{ogp}, there do not exist $\boldsymbol{\sigma}, \boldsymbol{\rho}$ such that $\|\boldsymbol{\sigma} - \boldsymbol{\rho}\| = q_{\text{ogp}}$ and

$\boldsymbol{\sigma}$ solves $\boldsymbol{G}^{t_1}$ and $\boldsymbol{\rho}$ solves $\boldsymbol{G}^{t_2}$ for some t_1, t_2

Chaos property : $\boldsymbol{G}^0$, $\boldsymbol{G}^T$ don't have solutions $\boldsymbol{\sigma}, \boldsymbol{\rho}$ with dist $\leqslant q_{\text{ogp}}$

Local algorithms do not reach OPT



Suppose local $\mathcal{A}$ beats $(1+\frac{1}{\sqrt{2}}+\varepsilon)\frac{\log d}{d}N \Rightarrow$ each σ^t solves G^t
(i.e. is ind set of size $\geqslant (1+\frac{1}{\sqrt{2}}+\varepsilon)\frac{\log d}{d}N$)

Landscape obstruction : for "medium" q_{ogp}, there do not exist σ, ρ such that $\|\sigma - \rho\| = q_{\text{ogp}}$ and

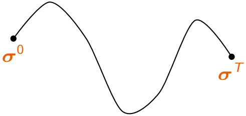
σ solves G^{t_1} and ρ solves G^{t_2} for some t_1, t_2

Chaos property : G^0, G^T don't have solutions σ, ρ with dist $\leqslant q_{\text{ogp}}$

Proof for both: calculate $\mathbb{E}\#\{\text{such } (\sigma, \rho)\} \ll 1$

Local algorithms do not reach OPT

$\sigma^t = \mathcal{A}(G^t)$
$\Longrightarrow$
small steps
by stability



Suppose local $\mathcal{A}$ beats $(1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N \Rightarrow$ each $\boldsymbol{\sigma}^t$ solves $\boldsymbol{G}^t$
(i.e. is ind set of size $\geqslant (1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N$)

Landscape obstruction : for "medium" q_{ogp}, there do not exist $\boldsymbol{\sigma}, \boldsymbol{\rho}$ such that $\|\boldsymbol{\sigma} - \boldsymbol{\rho}\| = q_{\text{ogp}}$ and

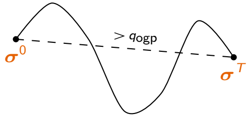
$\boldsymbol{\sigma}$ solves $\boldsymbol{G}^{t_1}$ and $\boldsymbol{\rho}$ solves $\boldsymbol{G}^{t_2}$ for some t_1, t_2

Chaos property: $\boldsymbol{G}^0$, $\boldsymbol{G}^T$ don't have solutions $\boldsymbol{\sigma}, \boldsymbol{\rho}$ with dist $\leqslant q_{\text{ogp}}$

Proof for both: calculate $\mathbb{E}\#\{\text{such } (\boldsymbol{\sigma}, \boldsymbol{\rho})\} \ll 1$

Local algorithms do not reach OPT

$\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$ $\Longrightarrow$ small steps by stability



Suppose local $\mathcal{A}$ beats $(1+\frac{1}{\sqrt{2}}+\varepsilon)\frac{\log d}{d}N \Rightarrow$ each $\boldsymbol{\sigma}^t$ solves $\boldsymbol{G}^t$
(i.e. is ind set of size $\geqslant (1+\frac{1}{\sqrt{2}}+\varepsilon)\frac{\log d}{d}N$)

Landscape obstruction : for "medium" q_{ogp}, there do not exist $\boldsymbol{\sigma}, \boldsymbol{\rho}$ such that $\|\boldsymbol{\sigma}-\boldsymbol{\rho}\| = q_{\text{ogp}}$ and

$\boldsymbol{\sigma}$ solves $\boldsymbol{G}^{t_1}$ and $\boldsymbol{\rho}$ solves $\boldsymbol{G}^{t_2}$ for some t_1, t_2

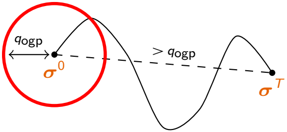
Chaos property : $\boldsymbol{G}^0$, $\boldsymbol{G}^T$ don't have solutions $\boldsymbol{\sigma}, \boldsymbol{\rho}$ with dist $\leqslant q_{\text{ogp}}$

Proof for both: calculate $\mathbb{E}\#\{\text{such } (\boldsymbol{\sigma}, \boldsymbol{\rho})\} \ll 1$

Local algorithms do not reach OPT

$G^0 - G^1 - G^2 - \cdots - G^T$

$\sigma^t = \mathcal{A}(G^t)$
$\Longrightarrow$
small steps
by stability



Suppose local $\mathcal{A}$ beats $(1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N \Rightarrow$ each σ^t solves G^t
(i.e. is ind set of size $\geqslant (1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N$)

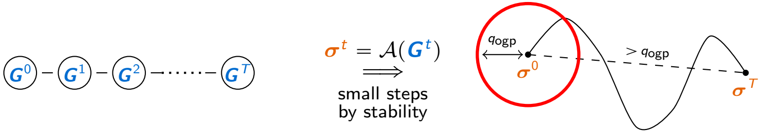
Landscape obstruction : for "medium" q_{ogp}, there do not exist σ, ρ such that $\|\sigma - \rho\| = q_{\text{ogp}}$ and

σ solves G^{t_1} and ρ solves G^{t_2} for some t_1, t_2

Chaos property : G^0, G^T don't have solutions σ, ρ with dist $\leqslant q_{\text{ogp}}$

Proof for both: calculate $\mathbb{E}\#\{\text{such } (\sigma, \rho)\} \ll 1$

Local algorithms do not reach OPT



Suppose local $\mathcal{A}$ beats $(1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N \Rightarrow$ each σ^t solves G^t
(i.e. is ind set of size $\geqslant (1 + \frac{1}{\sqrt{2}} + \varepsilon)\frac{\log d}{d}N$)

Landscape obstruction: for "medium" q_{ogp}, there do not exist σ, ρ such that $\|\sigma - \rho\| = q_{\text{ogp}}$ and

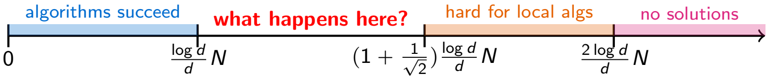
σ solves G^{t_1} and ρ solves G^{t_2} for some t_1, t_2

Chaos property: G^0, G^T don't have solutions σ, ρ with dist $\leqslant q_{\text{ogp}}$

Proof for both: calculate $\mathbb{E}\#\{\text{such } (\sigma, \rho)\} \ll 1$

Questions

- Can we show a tighter bound?



- Problems beyond max independent set?

- Algorithm classes beyond local algorithms?

- Finer-grained runtime bounds?

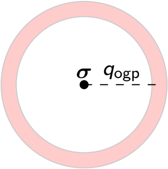
Remarks

Instead of distance $\|\boldsymbol{\sigma} - \boldsymbol{\rho}\|$, equivalent to consider **overlap** $(\boldsymbol{\sigma}, \boldsymbol{\rho})/N$

Remarks

Instead of distance $\|\boldsymbol{\sigma} - \boldsymbol{\rho}\|$, equivalent to consider **overlap** $(\boldsymbol{\sigma}, \boldsymbol{\rho})/N$

"No solutions $\boldsymbol{\sigma}, \boldsymbol{\rho}$ with medium overlap" is notion of **clustering**:

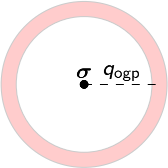


no solutions in red zone

Remarks

Instead of distance $\|\sigma - \rho\|$, equivalent to consider **overlap** $(\sigma, \rho)/N$

"No solutions σ, ρ with medium overlap" is notion of **clustering**:



no solutions in red zone

Key distinction: clustering of **most** vs **all** solutions

• Shattering, RSB, etc. concern when **most** solutions cluster/isolated. Algorithms may succeed by finding atypical solutions (Baldassi Ingrosso Lucibello Saglietti Zecchina 15, Abbe Li Sly 21)

• OGP: **all** solutions cluster (even across correlated instances), which implies hardness rigorously

Remarks

OGP uses **geometry** to rule out **stable algorithms**. We hope this is indicative of hardness for all **polynomial time** algorithms.

Known exceptions:

- Random k-XOR-SAT exhibits OGP, but solved by gaussian elimination
- Lattice methods use algebraic structure (Zadik Song Wein Bruna 21)
- Shortest path exhibits OGP but easy (Li Schramm 24)

Outline of part II: a survey on the overlap gap property

Introduction and motivating problems

Overlap gap property: the basics

More OGPs and algorithm classes

Further enhancements

Hardness of finding strict local maxima

Strong low degree hardness

Beyond the classic OGP

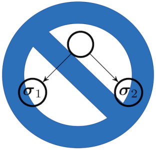
Many developments after the classic OGP, following same principle:

- If algorithm succeeds, it can build some constellation of solutions
- But we can show this constellation doesn't exist

Beyond the classic OGP

Many developments after the classic OGP, following same principle:

- If algorithm succeeds, it can build some constellation of solutions
- But we can show this constellation doesn't exist

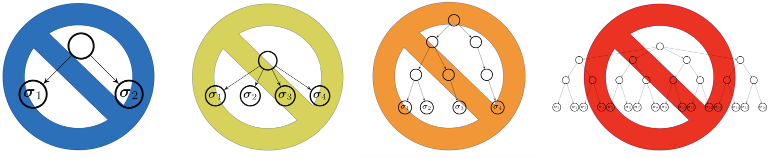


- **Classic** OGP: two points with distance q (Gamarnik Sudan 14)

Beyond the classic OGP

Many developments after the classic OGP, following same principle:

- If algorithm succeeds, it can build some constellation of solutions
- But we can show this constellation doesn't exist

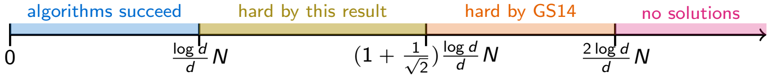


- **Classic** OGP: two points with distance q (Gamarnik Sudan 14)
- **Star** OGP: several points with pairwise distance q (Rahman Virág 17)
- **Ladder** OGP: $\boldsymbol{\sigma}^i$ has distance q to span$(\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^{i-1})$ (Wein 21)
- **Branching** OGP: densely branching tree (H Sellke 21)

Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

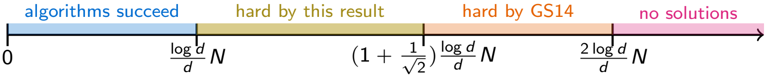
Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.



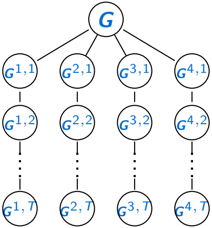
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.



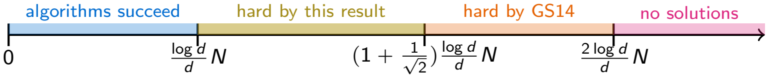
Construct $m = O(1)$ independent resample paths:



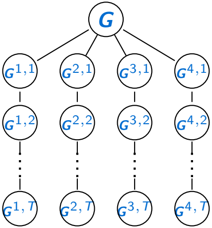
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$, whp.



Construct $m = O(1)$ independent resample paths:

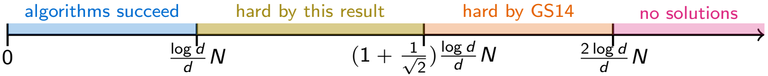


$$\boldsymbol{\sigma}^{i,t} \underset{\Longrightarrow}{=} \mathcal{A}(\boldsymbol{G}^{i,t})$$

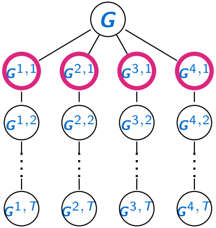
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

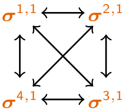
Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.



Construct $m = O(1)$ independent resample paths:



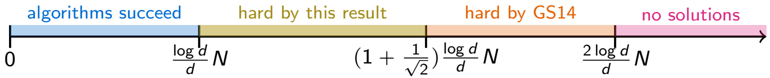
$\sigma^{i,t} = \mathcal{A}(G^{i,t}) \implies$



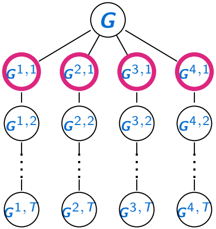
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$, whp.

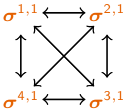


Construct $m = O(1)$ independent resample paths:



$$\boldsymbol{\sigma}^{i,t} = \mathcal{A}(\boldsymbol{G}^{i,t}) \Longrightarrow$$

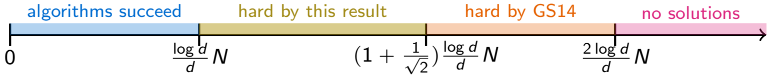
overlaps concentrate
and thus pairwise ≈equal



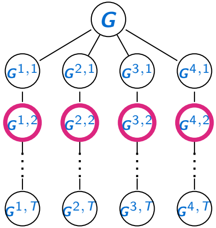
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.

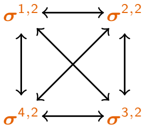


Construct $m = O(1)$ independent resample paths:



$\boldsymbol{\sigma}^{i,t} = \mathcal{A}(\boldsymbol{G}^{i,t}) \Longrightarrow$

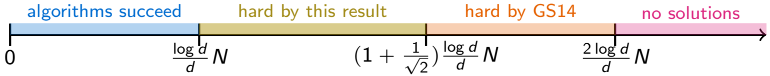
overlaps concentrate
and thus pairwise ≈equal



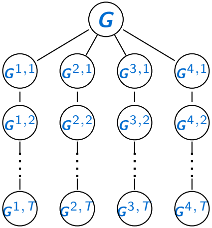
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.

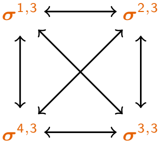


Construct $m = O(1)$ independent resample paths:



$\sigma^{i,t} = \mathcal{A}(G^{i,t})$ $\Longrightarrow$

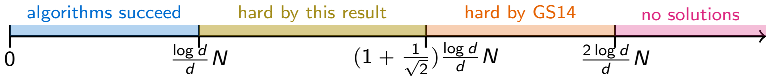
overlaps concentrate
and thus pairwise ≈equal



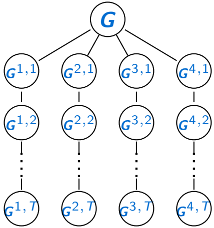
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$, whp.

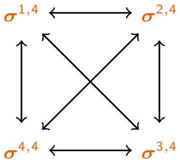


Construct $m = O(1)$ independent resample paths:



$\sigma^{i,t} = \mathcal{A}(G^{i,t})$ $\Longrightarrow$

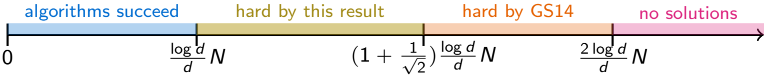
overlaps concentrate
and thus pairwise ≈equal



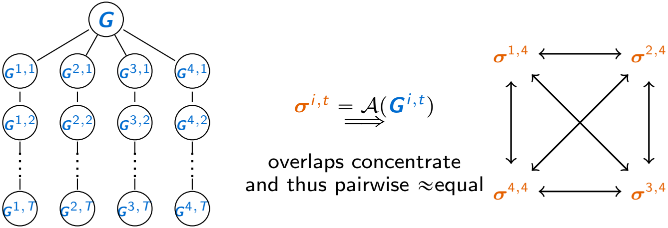
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$, whp.



Construct $m = O(1)$ independent resample paths:

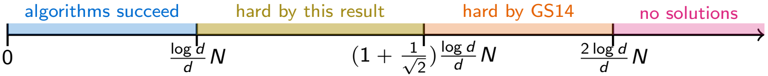


Landscape obstruction: there don't exist m ind sets of size $\geqslant (1 + \varepsilon)\frac{\log d}{d} N$ (in possibly different $G^{i,t}$) with pairwise overlap q_{ogp}

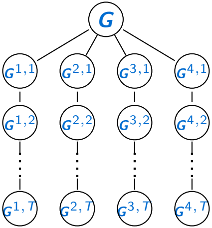
Star OGP: tight hardness for max independent set

Theorem (Rahman Virág 17)

Any $O(1)$-local algorithm does not find an independent set of size $\geqslant (\mathbf{1} + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$, whp.

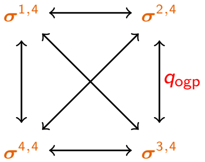


Construct $m = O(1)$ independent resample paths:



$\boldsymbol{\sigma}^{i,t} = \mathcal{A}(\boldsymbol{G}^{i,t})$ $\Longrightarrow$

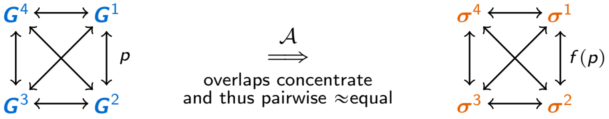
overlaps concentrate
and thus pairwise ≈equal



Landscape obstruction: there don't exist m ind sets of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (in possibly different $\boldsymbol{G}^{i,t}$) with pairwise overlap q_{ogp}

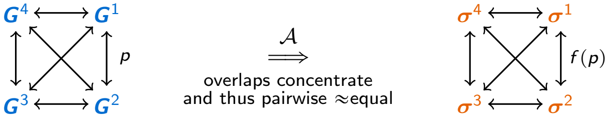
Stable vs concentrated algorithms

This argument uses strongly that overlaps $(\mathcal{A}(G), \mathcal{A}(G'))$ concentrate



Stable vs concentrated algorithms

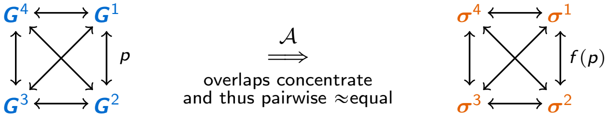
This argument uses strongly that overlaps $(\mathcal{A}(G), \mathcal{A}(G'))$ concentrate



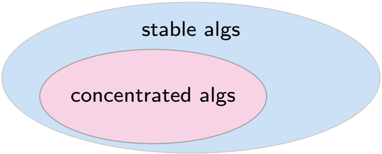
- **Stable** alg: $\mathbb{E}\|\mathcal{A}(G) - \mathcal{A}(G')\|_2^2 \lesssim \varepsilon N$ for $(1-\varepsilon)$-correlated G, G'
- Stable alg is **concentrated** if: $(\mathcal{A}(G), \mathcal{A}(G'))$ concentrates for q-correlated G, G', $\forall q \in [0, 1]$

Stable vs concentrated algorithms

This argument uses strongly that overlaps $(\mathcal{A}(G), \mathcal{A}(G'))$ concentrate

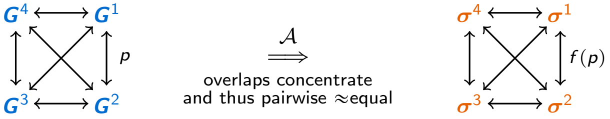


- **Stable** alg: $\mathbb{E}\|\mathcal{A}(G) - \mathcal{A}(G')\|_2^2 \lesssim \varepsilon N$ for $(1-\varepsilon)$-correlated G, G'
- Stable alg is **concentrated** if: $(\mathcal{A}(G), \mathcal{A}(G'))$ concentrates for q-correlated G, G', $\forall q \in [0,1]$

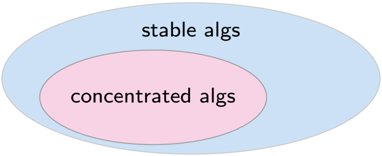


Stable vs concentrated algorithms

This argument uses strongly that overlaps $(\mathcal{A}(G), \mathcal{A}(G'))$ concentrate



- **Stable** alg: $\mathbb{E}\|\mathcal{A}(G) - \mathcal{A}(G')\|_2^2 \lesssim \varepsilon N$ for $(1-\varepsilon)$-correlated G, G'
- Stable alg is **concentrated** if: $(\mathcal{A}(G), \mathcal{A}(G'))$ concentrates for q-correlated G, G', $\forall q \in [0, 1]$



- Concentration $\Rightarrow$ control **all $\binom{m}{2}$ overlaps** among $\mathcal{A}(G^1), \ldots, \mathcal{A}(G^m)$
- Stability $\Rightarrow$ can only use **IVT** considerations to control $\approx m$ **overlaps**.

Stable vs concentrated algorithms

Stable algorithms:
$\mathbb{E}\|\mathcal{A}(G) - \mathcal{A}(G')\|_2^2 \lesssim \varepsilon N$ for $(1-\varepsilon)$-correlated G, G'

- Gradient descent, Langevin dynamics for $O(1)$ time
- AMP and general $O(1)$ order algorithms for $O(1)$ time
- $O(1)$-local algorithms
- Low degree polynomials
- Low depth circuits

Stable vs concentrated algorithms

Stable algorithms:
$\mathbb{E}\|\mathcal{A}(G) - \mathcal{A}(G')\|_2^2 \lesssim \varepsilon N$ for $(1-\varepsilon)$-correlated G, G'

- Gradient descent, Langevin dynamics for $O(1)$ time
- AMP and general $O(1)$ order algorithms for $O(1)$ time
- $O(1)$-local algorithms
- Low degree polynomials
- Low depth circuits

Concentrated algorithms:
$(\mathcal{A}(G), \mathcal{A}(G'))$ concentrates for q-correlated G, G', $\forall q \in [0,1]$

- Gradient descent, Langevin dynamics for $O(1)$ time
- AMP and general $O(1)$ order algorithms for $O(1)$ time
- $O(1)$-local algorithms
- ~~Low degree polynomials~~
- ~~Low depth circuits~~

Stable vs concentrated algorithms

Stable algorithms:
$\mathbb{E}\|\mathcal{A}(G) - \mathcal{A}(G')\|_2^2 \lesssim \varepsilon N$ for $(1-\varepsilon)$-correlated G, G'

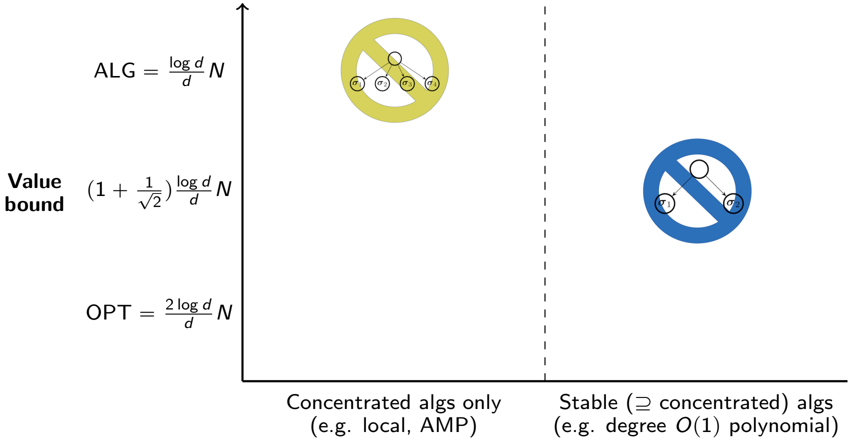
- Gradient descent, Langevin dynamics for $O(1)$ time
- AMP and general $O(1)$ order algorithms for $O(1)$ time
- $O(1)$-local algorithms
- Low degree polynomials
- Low depth circuits

Concentrated algorithms:
$(\mathcal{A}(G), \mathcal{A}(G'))$ concentrates for q-correlated G, G', $\forall q \in [0,1]$

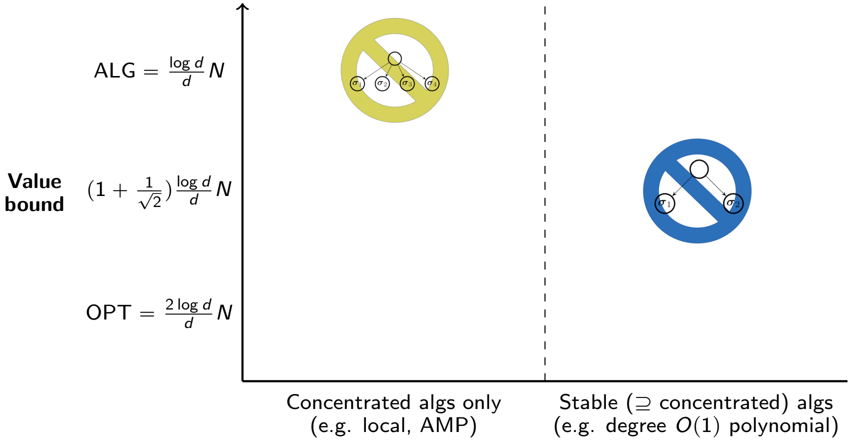
- Gradient descent, Langevin dynamics for $O(1)$ time
- AMP and general $O(1)$ order algorithms for $O(1)$ time
- $O(1)$-local algorithms
- ~~Low degree polynomials~~
- ~~Low depth circuits~~

Two classes of OGP hardness proofs: those where **stability is enough**, and those that **only work on concentrated algorithms**

Comparison of OGPs for max independent set

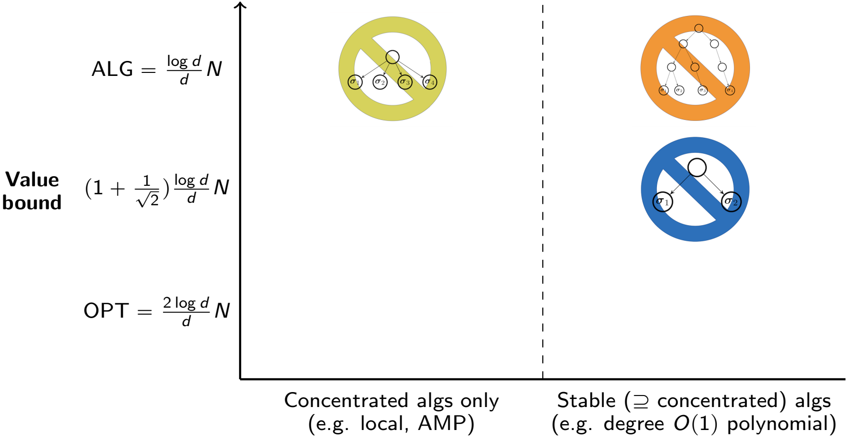


Comparison of OGPs for max independent set



Can we get optimal hardness in this problem for **stable** algorithms, like **low degree polynomials?**

Comparison of OGPs for max independent set



Can we get optimal hardness in this problem for **stable** algorithms, like **low degree polynomials?**

Low degree polynomials

Encode $\boldsymbol{G} \in \{0,1\}^{\binom{N}{2}}$ by edge indicators.

Low degree polynomials

Encode $\boldsymbol{G} \in \{0,1\}^{\binom{N}{2}}$ by edge indicators. Class of algs $\mathcal{A} : \mathbb{R}^{\binom{N}{2}} \to \mathbb{R}^{N}$

$$\mathcal{A}(\boldsymbol{G}) = (\mathcal{A}_1(\boldsymbol{G}), \ldots, \mathcal{A}_N(\boldsymbol{G}))\,.$$

where each $\mathcal{A}_i$ is a (possibly random) degree $\leqslant \boldsymbol{D}$ polynomial.

Low degree polynomials

Encode $\boldsymbol{G} \in \{0,1\}^{\binom{N}{2}}$ by edge indicators. Class of algs $\mathcal{A} : \mathbb{R}^{\binom{N}{2}} \to \mathbb{R}^{N}$

$$\mathcal{A}(\boldsymbol{G}) = (\mathcal{A}_1(\boldsymbol{G}), \ldots, \mathcal{A}_N(\boldsymbol{G}))\,.$$

where each $\mathcal{A}_i$ is a (possibly random) degree $\leqslant \boldsymbol{D}$ polynomial.
Round to get output in $\{\pm 1\}^N$.

Low degree polynomials

Encode $\boldsymbol{G} \in \{0,1\}^{\binom{N}{2}}$ by edge indicators. Class of algs $\mathcal{A} : \mathbb{R}^{\binom{N}{2}} \to \mathbb{R}^N$

$$\mathcal{A}(\boldsymbol{G}) = (\mathcal{A}_1(\boldsymbol{G}), \dots, \mathcal{A}_N(\boldsymbol{G}))\,.$$

where each $\mathcal{A}_i$ is a (possibly random) degree $\leqslant \boldsymbol{D}$ polynomial.
Round to get output in $\{\pm 1\}^N$. Deg $O(1)$ polys include:

- Gradient descent, Langevin dynamics
- Approximate message passing, first-order iterations
- Message passing algorithms on factor graphs

} for $O(1)$ time

Low degree polynomials

Encode $\boldsymbol{G} \in \{0,1\}^{\binom{N}{2}}$ by edge indicators. Class of algs $\mathcal{A} : \mathbb{R}^{\binom{N}{2}} \to \mathbb{R}^N$

$$\mathcal{A}(\boldsymbol{G}) = (\mathcal{A}_1(\boldsymbol{G}), \ldots, \mathcal{A}_N(\boldsymbol{G}))\,.$$

where each $\mathcal{A}_i$ is a (possibly random) degree $\leqslant \boldsymbol{D}$ polynomial.
Round to get output in $\{\pm 1\}^N$. Deg $O(1)$ polys include:

- Gradient descent, Langevin dynamics
- Approximate message passing, first-order iterations
- Message passing algorithms on factor graphs

} for $O(1)$ time

Low degree heuristic: deg $\leqslant \boldsymbol{D}$ polynomials $\approx e^{\tilde{O}(\boldsymbol{D})}$ time algorithms in many statistical problems (Hopkins 18, Kunisky Wein Bandeira 19, . . .)

Low degree polynomials

Encode $\boldsymbol{G} \in \{0,1\}^{\binom{N}{2}}$ by edge indicators. Class of algs $\mathcal{A} : \mathbb{R}^{\binom{N}{2}} \to \mathbb{R}^N$

$$\mathcal{A}(\boldsymbol{G}) = (\mathcal{A}_1(\boldsymbol{G}), \ldots, \mathcal{A}_N(\boldsymbol{G}))\,.$$

where each $\mathcal{A}_i$ is a (possibly random) degree $\leqslant \boldsymbol{D}$ polynomial.
Round to get output in $\{\pm 1\}^N$. Deg $O(1)$ polys include:

- Gradient descent, Langevin dynamics
- Approximate message passing, first-order iterations
- Message passing algorithms on factor graphs

} for $O(1)$ time

Low degree heuristic: deg $\leqslant \boldsymbol{D}$ polynomials $\approx e^{\widetilde{O}(\boldsymbol{D})}$ time algorithms in many statistical problems (Hopkins 18, Kunisky Wein Bandeira 19, . . .)

(But see Buhai Hsieh Jain Kothari 25 for counterexample)

Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1 + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$.

Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1 + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1 + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \cdots - G^{mT}$$

Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1 + o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\rho^1, \ldots, \rho^m$ of size $\geqslant (1 + \varepsilon)\frac{\log d}{d} N$ (to possibly different G^t) where

Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1 + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$\boldsymbol{G}^0 - \boldsymbol{G}^1 - \boldsymbol{G}^2 - \boldsymbol{G}^3 - \boldsymbol{G}^4 - \boldsymbol{G}^5 - \quad \cdots \quad - \boldsymbol{G}^{mT}$$

Forbidden structure: m ind sets $\boldsymbol{\rho}^1, \ldots, \boldsymbol{\rho}^m$ of size $\geqslant (1 + \varepsilon)\frac{\log d}{d}N$ (to possibly different $\boldsymbol{G}^t$) where

$$\frac{1}{N}|\boldsymbol{\rho}^t \backslash (\boldsymbol{\rho}^1 \cup \cdots \cup \boldsymbol{\rho}^{t-1})| \in [q_{\mathrm{ogp}}, q_{\mathrm{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1 + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \cdots - G^{mT}$$

Forbidden structure: m ind sets $\rho^1, \ldots, \rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$

Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

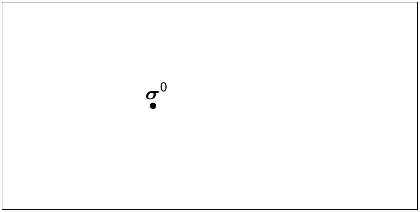
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N,d/N)$. (Later: we improve this to deg $o(N)$)

$$\boldsymbol{G}^0 - \boldsymbol{G}^1 - \boldsymbol{G}^2 - \boldsymbol{G}^3 - \boldsymbol{G}^4 - \boldsymbol{G}^5 - \quad \cdots \quad - \boldsymbol{G}^{mT}$$

Forbidden structure: m ind sets $\boldsymbol{\rho}^1, \dots, \boldsymbol{\rho}^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different $\boldsymbol{G}^t$) where

$$\frac{1}{N}|\boldsymbol{\rho}^t \backslash (\boldsymbol{\rho}^1 \cup \cdots \cup \boldsymbol{\rho}^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$ is large ind set in $\boldsymbol{G}^t$, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

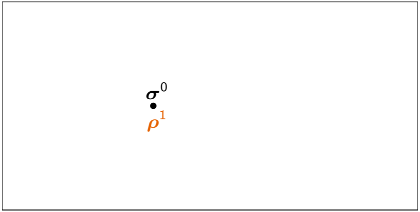
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N,d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\boldsymbol{\rho}^1, \ldots, \boldsymbol{\rho}^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different $\boldsymbol{G}^t$) where

$$\frac{1}{N}|\boldsymbol{\rho}^t \backslash (\boldsymbol{\rho}^1 \cup \cdots \cup \boldsymbol{\rho}^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$ is large ind set in $\boldsymbol{G}^t$, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

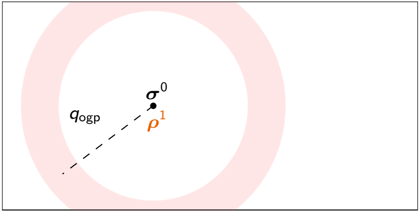
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\rho^1, \dots, \rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

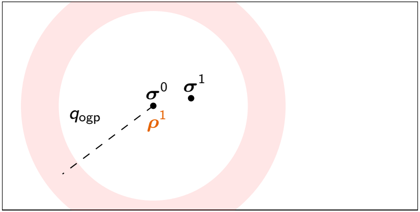
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N,d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\rho^1,\dots,\rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}}+\delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

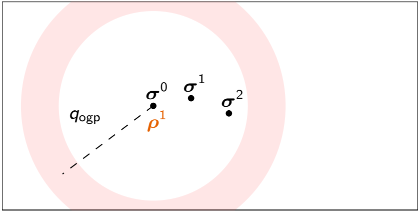
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\rho^1, \ldots, \rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

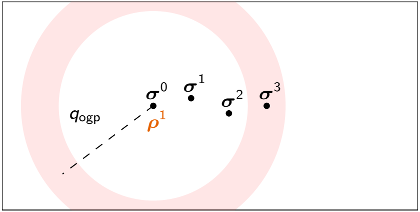
Theorem (Wein 21)

Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

Forbidden structure: m ind sets $\rho^1, \dots, \rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \dots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

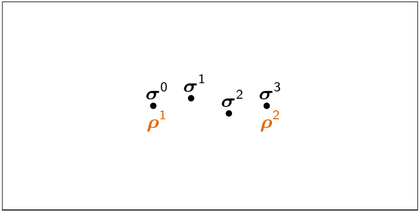
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N,d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\rho^1,\ldots,\rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \setminus (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}}+\delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

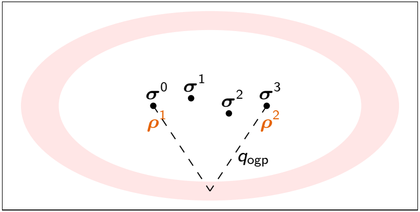
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N,d/N)$. (Later: we improve this to deg $o(N)$)

$$\boldsymbol{G}^0 - \boldsymbol{G}^1 - \boldsymbol{G}^2 - \boldsymbol{G}^3 - \boldsymbol{G}^4 - \boldsymbol{G}^5 - \quad \cdots \quad - \boldsymbol{G}^{mT}$$

Forbidden structure: m ind sets $\boldsymbol{\rho}^1,\dots,\boldsymbol{\rho}^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different $\boldsymbol{G}^t$) where

$$\frac{1}{N}|\boldsymbol{\rho}^t \backslash (\boldsymbol{\rho}^1 \cup \cdots \cup \boldsymbol{\rho}^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}}+\delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$ is large ind set in $\boldsymbol{G}^t$, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

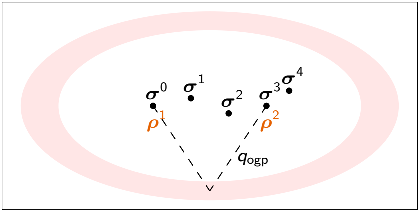
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\rho^1, \dots, \rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

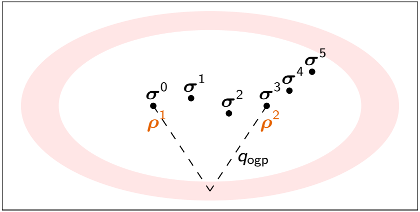
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1 + o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\rho^1, \ldots, \rho^m$ of size $\geqslant (1 + \varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\boldsymbol{\sigma}^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

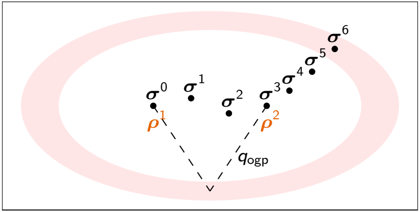
Theorem (Wein 21)

Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d} N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

Forbidden structure: m ind sets $\rho^1, \ldots, \rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d} N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

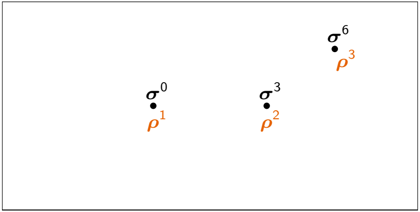
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N, d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\rho^1, \ldots, \rho^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different G^t) where

$$\frac{1}{N}|\rho^t \backslash (\rho^1 \cup \cdots \cup \rho^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\sigma^t = \mathcal{A}(G^t)$ is large ind set in G^t, $\forall t$



Ladder OGP: hardness for low degree polynomials

Theorem (Wein 21)

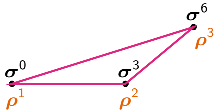
Any $O(1)$-deg polynomial does not find an independent set of size $\geqslant (1+o_d(1))\frac{\log d}{d}N$ in $G(N,d/N)$. (Later: we improve this to deg $o(N)$)

$$G^0 - G^1 - G^2 - G^3 - G^4 - G^5 - \quad \cdots \quad - G^{mT}$$

Forbidden structure: m ind sets $\boldsymbol{\rho}^1,\ldots,\boldsymbol{\rho}^m$ of size $\geqslant (1+\varepsilon)\frac{\log d}{d}N$ (to possibly different $\boldsymbol{G}^t$) where

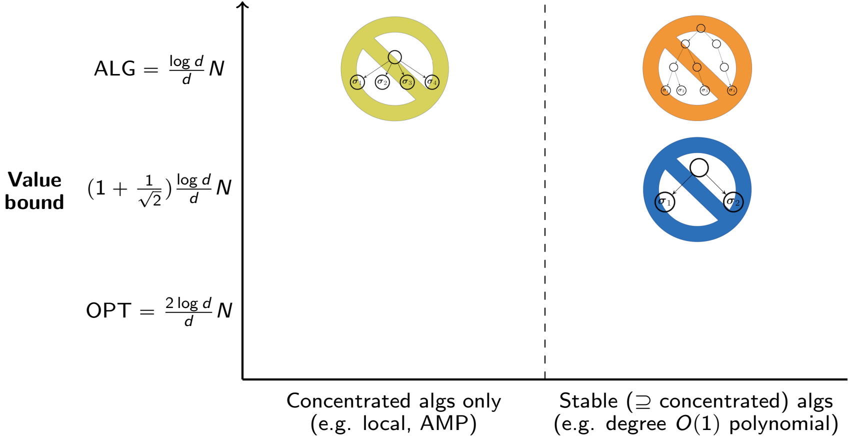
$$\frac{1}{N}|\boldsymbol{\rho}^t \backslash (\boldsymbol{\rho}^1 \cup \cdots \cup \boldsymbol{\rho}^{t-1})| \in [q_{\text{ogp}}, q_{\text{ogp}}+\delta] \qquad \text{for all } 2 \leqslant t \leqslant m$$

Suppose $O(1)$-deg $\mathcal{A}$ succeeds: $\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$ is large ind set in $\boldsymbol{G}^t$, $\forall t$



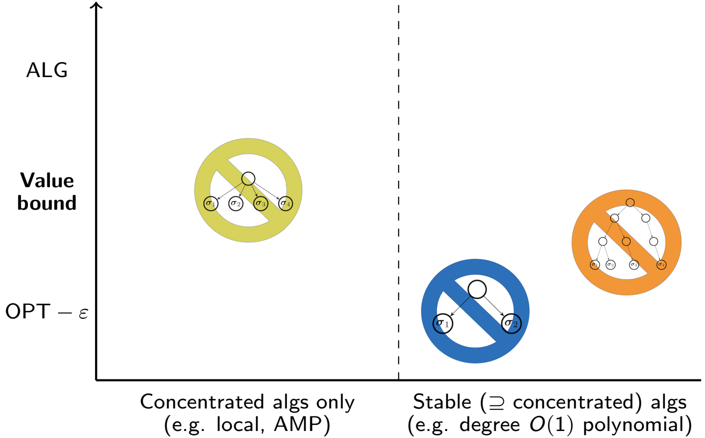
This is the forbidden structure!

Comparison of OGPs for max independent set



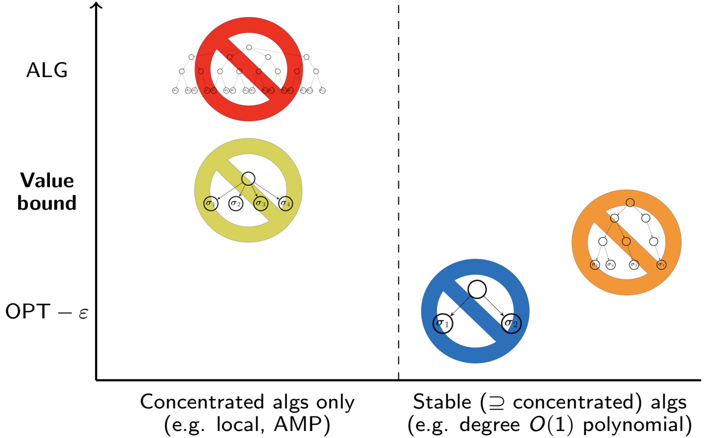
What about other problems?

Comparison of OGPs in general



These don't generally match ALG. **Can we get optimal hardness?**

Comparison of OGPs in general



These don't generally match ALG. **Can we get optimal hardness?**

Another model: mean-field spin glass

Polynomials with IID gaussian coefficients, e.g. random cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k}\, \sigma_i \sigma_j \sigma_k\,, \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

More generally, linear combinations of different degrees.

Another model: mean-field spin glass

Polynomials with IID gaussian coefficients, e.g. random cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k}\, \sigma_i \sigma_j \sigma_k\,, \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

More generally, linear combinations of different degrees.
Domain: $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising).

Another model: mean-field spin glass

Polynomials with IID gaussian coefficients, e.g. random cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k}\, \sigma_i \sigma_j \sigma_k\,, \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

More generally, linear combinations of different degrees.
Domain: $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising).

Q: given H, algorithmically find $\boldsymbol{\sigma}^{\text{alg}}$ with $H(\boldsymbol{\sigma}^{\text{alg}})$ as large as possible.

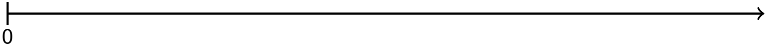
Another model: mean-field spin glass

Polynomials with IID gaussian coefficients, e.g. random cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k}\, \sigma_i \sigma_j \sigma_k\,, \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

More generally, linear combinations of different degrees.
Domain: $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising).

Q: given H, algorithmically find $\boldsymbol{\sigma}^{\mathrm{alg}}$ with $H(\boldsymbol{\sigma}^{\mathrm{alg}})$ as large as possible.



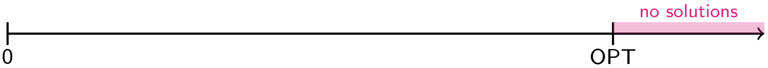
Another model: mean-field spin glass

Polynomials with IID gaussian coefficients, e.g. random cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k}\, \sigma_i \sigma_j \sigma_k\,, \qquad g_{i,j,k} \stackrel{IID}{\sim} \mathcal{N}(0,1)$$

More generally, linear combinations of different degrees.
Domain: $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising).

Q: given H, algorithmically find $\boldsymbol{\sigma}^{\mathrm{alg}}$ with $H(\boldsymbol{\sigma}^{\mathrm{alg}})$ as large as possible.



- OPT given by **Parisi formula** (Parisi 79, Talagrand 06, Panchenko 13)

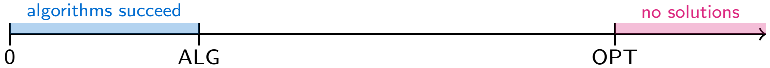
Another model: mean-field spin glass

Polynomials with IID gaussian coefficients, e.g. random cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k}\, \sigma_i \sigma_j \sigma_k\,, \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

More generally, linear combinations of different degrees.
Domain: $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising).

Q: given H, algorithmically find $\boldsymbol{\sigma}^{\mathrm{alg}}$ with $H(\boldsymbol{\sigma}^{\mathrm{alg}})$ as large as possible.



- OPT given by **Parisi formula** (Parisi 79, Talagrand 06, Panchenko 13)
- Algorithms reach explicit ALG (Subag 18, Alaoui Montanari Sellke 20)

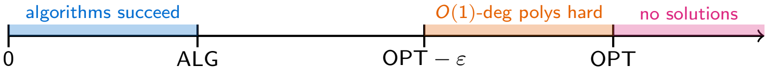
Another model: mean-field spin glass

Polynomials with IID gaussian coefficients, e.g. random cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k}\, \sigma_i \sigma_j \sigma_k\,, \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

More generally, linear combinations of different degrees.
Domain: $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising).

Q: given H, algorithmically find $\boldsymbol{\sigma}^{\mathrm{alg}}$ with $H(\boldsymbol{\sigma}^{\mathrm{alg}})$ as large as possible.



- OPT given by **Parisi formula** (Parisi 79, Talagrand 06, Panchenko 13)
- Algorithms reach explicit ALG (Subag 18, Alaoui Montanari Sellke 20)
- Classic OGP $\Rightarrow$ hardness at OPT $-\, \varepsilon$ (Gamarnik Jagannath Wein 20)

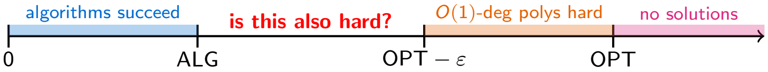
Another model: mean-field spin glass

Polynomials with IID gaussian coefficients, e.g. random cubic

$$H(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i,j,k=1}^{N} g_{i,j,k}\, \sigma_i \sigma_j \sigma_k \,, \qquad g_{i,j,k} \overset{IID}{\sim} \mathcal{N}(0,1)$$

More generally, linear combinations of different degrees.
Domain: $S_N = \sqrt{N}\mathbb{S}^{N-1}$ (spherical) or $\Sigma_N = \{\pm 1\}^N$ (Ising).

Q: given H, algorithmically find $\boldsymbol{\sigma}^{\mathrm{alg}}$ with $H(\boldsymbol{\sigma}^{\mathrm{alg}})$ as large as possible.



- OPT given by **Parisi formula** (Parisi 79, Talagrand 06, Panchenko 13)
- Algorithms reach explicit ALG (Subag 18, Alaoui Montanari Sellke 20)
- Classic OGP $\Rightarrow$ hardness at OPT $- \varepsilon$ (Gamarnik Jagannath Wein 20)

Branching OGP: exact threshold

Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG*.*

Branching OGP: exact threshold

Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG*.*

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Branching OGP: exact threshold

Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys

Branching OGP: exact threshold

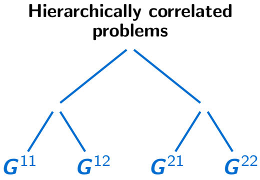
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG*.*

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

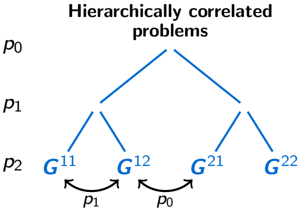
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

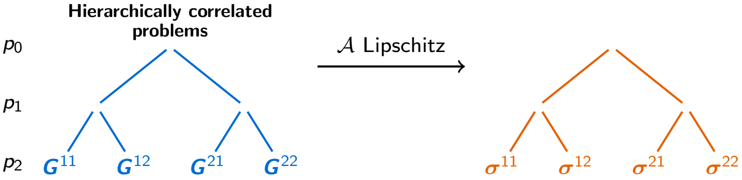
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

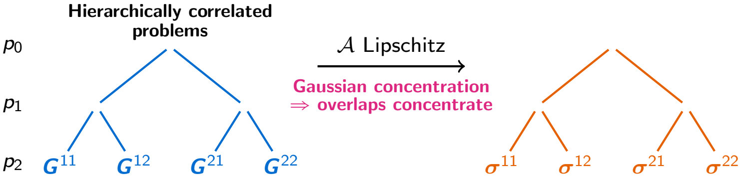
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

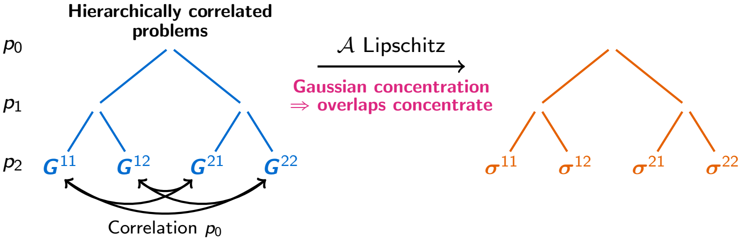
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

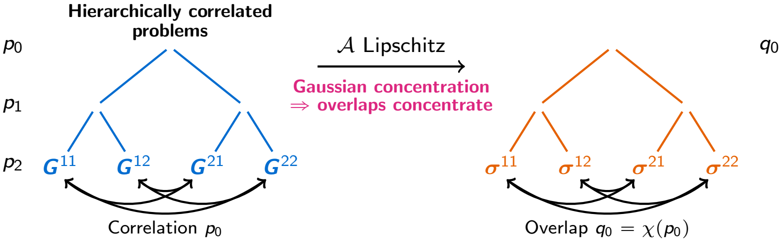
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

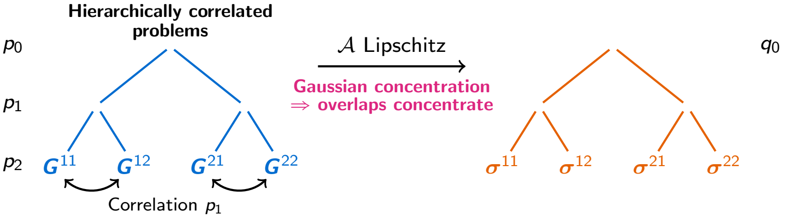
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

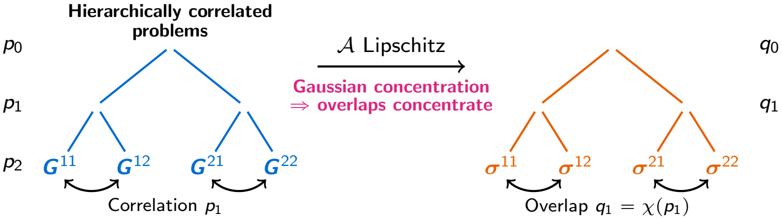
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

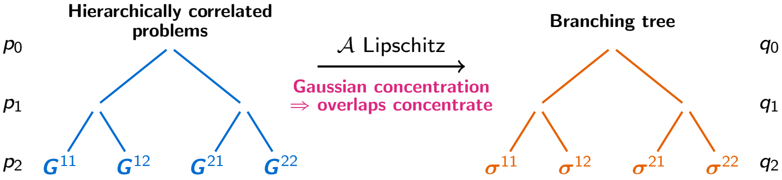
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats* ALG.

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Branching OGP: exact threshold

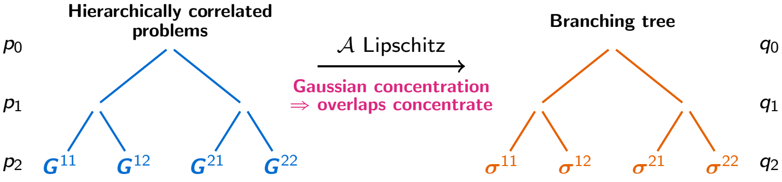
Theorem (H Sellke 21+23)

No concentrated (e.g. **Lipschitz***) algorithm beats ALG.*

Lipschitz algorithm: $\mathcal{A} : \mathbb{R}^M \to \mathbb{R}^N$ where

$$\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2 \leqslant O(1) \cdot \|\boldsymbol{G} - \boldsymbol{G}'\|_2$$

Includes grad desc, Langevin, AMP for $O(1)$ time, but not low deg polys



Forbidden structure: branching tree of $\boldsymbol{\sigma}^i$ each with value $\geqslant$ ALG $+ \varepsilon$

Geometric description of algorithmic threshold

(Lipschitz) algorithmic threshold is the supremal $\boldsymbol{E}$ whose super-level set

$$\left\{\boldsymbol{\sigma} : H(\boldsymbol{\sigma})/N \geqslant \boldsymbol{E}\right\}$$

contains the leaves of a densely branching tree (with high probability)

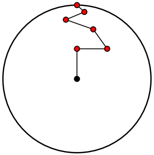
Geometric description of algorithmic threshold

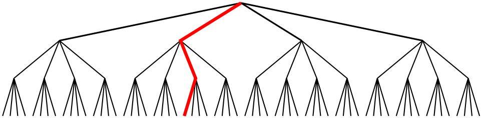
(Lipschitz) algorithmic threshold is the supremal $\boldsymbol{E}$ whose super-level set

$$\left\{\boldsymbol{\sigma} : H(\boldsymbol{\sigma})/N \geqslant \boldsymbol{E}\right\}$$

contains the leaves of a densely branching tree (with high probability)

• Achievability: efficient algorithms following approach of Subag 18, Montanari 18, El Alaoui Montanari Sellke 20 can descend this tree.





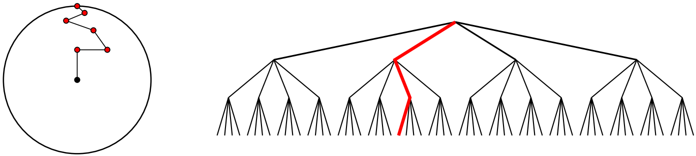
Geometric description of algorithmic threshold

(Lipschitz) algorithmic threshold is the supremal $\boldsymbol{E}$ whose super-level set

$$\left\{\boldsymbol{\sigma} : H(\boldsymbol{\sigma})/N \geqslant \boldsymbol{E}\right\}$$

contains the leaves of a densely branching tree (with high probability)

- Achievability: efficient algorithms following approach of Subag 18, Montanari 18, El Alaoui Montanari Sellke 20 can descend this tree.



- Hardness: any Lipschitz algorithm can be made to output such a tree.

Exact algorithmic thresholds in many problems

Branching OGP locates ALG for **Lipschitz** / **online** algorithms in:

Exact algorithmic thresholds in many problems

Branching OGP locates ALG for **Lipschitz** / **online** algorithms in:

- Optimizing mean-field spin glasses (Subag 18, Montanari 18, El Alaoui Montanari Sellke 20, H Sellke 21)
- Multi-species spherical spin glasses (H Sellke 23)

Exact algorithmic thresholds in many problems

Branching OGP locates ALG for **Lipschitz** / **online** algorithms in:

- Optimizing mean-field spin glasses (Subag 18, Montanari 18, El Alaoui Montanari Sellke 20, H Sellke 21)
- Multi-species spherical spin glasses (H Sellke 23)
- Random CSPs with average degree $\gg 1$ (Jones Marwaha Sandhu Shi 22, Chen Huang Marwaha 23)
- Random graph alignment (Du Gong Huang 25)

Exact algorithmic thresholds in many problems

Branching OGP locates ALG for **Lipschitz** / **online** algorithms in:

- Optimizing mean-field spin glasses (Subag 18, Montanari 18, El Alaoui Montanari Sellke 20, H Sellke 21)
- Multi-species spherical spin glasses (H Sellke 23)
- Random CSPs with average degree $\gg 1$ (Jones Marwaha Sandhu Shi 22, Chen Huang Marwaha 23)
- Random graph alignment (Du Gong Huang 25)
- Random systems of polynomial equations (Montanari Subag 24)
- Largest average submatrix / subtensor (Gamarnik Li 16, Bhamidi Gamarnik Gong 25)
- Random perceptron (Montanari Zhou 24, H Sellke Sun 25^{+})

Outline of part II: a survey on the overlap gap property

Introduction and motivating problems

Overlap gap property: the basics

More OGPs and algorithm classes

Further enhancements

Hardness of finding strict local maxima

Strong low degree hardness

Further enhancements

OGP methodology:

1. If algorithm succeeds, it can build some constellation of solutions
2. But we can show this constellation doesn't exist

Further enhancements

OGP methodology:

1. If algorithm succeeds, it can build some constellation of solutions
2. But we can show this constellation doesn't exist

So far: optimizations to step 2

- Make algorithm build **more complex** constellation
- Work harder to show this constellation doesn't exist

Further enhancements

OGP methodology:

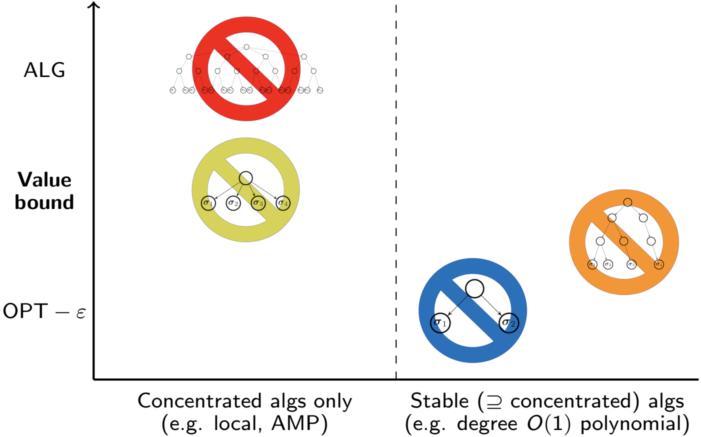
1. If algorithm succeeds, it can build some constellation of solutions
2. But we can show this constellation doesn't exist

So far: optimizations to step 2

- Make algorithm build **more complex** constellation
- Work harder to show this constellation doesn't exist

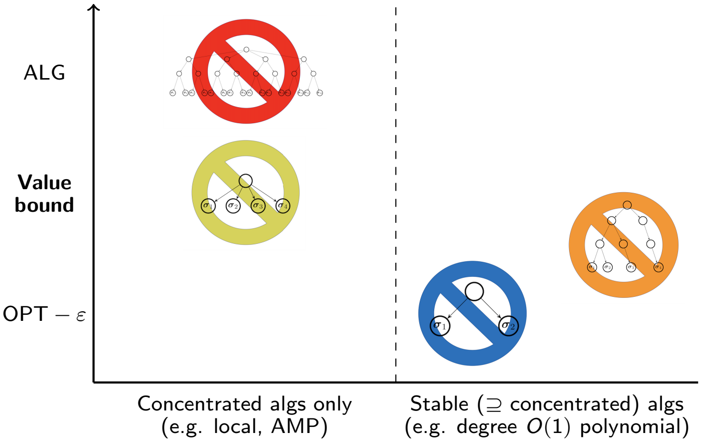
Next few slides: enhancements to step 1. More clever ways to force algorithm to **build a simple constellation**

Ramsey trick



Q: if we know our problem satisfies a **star** OGP, can we show hardness for **stable but not concentrated** algorithms?

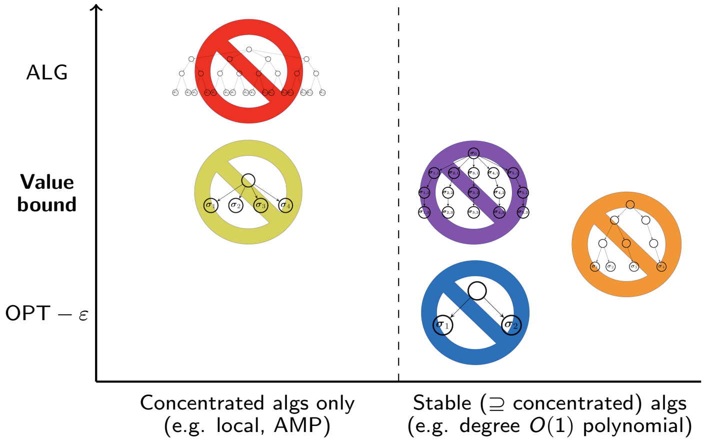
Ramsey trick



Q: if we know our problem satisfies a **star** OGP, can we show hardness for **stable but not concentrated** algorithms?

Gamarnik Kızıldağ 21: yes, via Ramsey theoretic argument

Ramsey trick



Q: if we know our problem satisfies a **star** OGP, can we show hardness for **stable but not concentrated** algorithms?

Gamarnik Kızıldağ 21: yes, via Ramsey theoretic argument

Symmetric Ising perceptron

Average case setting of **discrepancy minimization**: for $M/N = \alpha$, $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N \overset{\text{IID}}{\sim} \mathcal{N}(0, I_M)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ such that

$$\Big\| \sum_{i=1}^{N} \sigma_i \boldsymbol{g}^i \Big\|_\infty \leqslant \kappa \sqrt{N} \qquad (\kappa \ll 1)$$

Symmetric Ising perceptron

Average case setting of **discrepancy minimization**: for $M/N = \alpha$, $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N \overset{\text{IID}}{\sim} \mathcal{N}(0, I_M)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ such that

$$\Big\| \sum_{i=1}^{N} \sigma_i \boldsymbol{g}^i \Big\|_\infty \leqslant \kappa \sqrt{N} \qquad (\kappa \ll 1)$$

- Solutions exist up to $\alpha \approx \frac{1}{\log_2(1/\kappa)}$ (Abbe Li Sly 21)
- Algorithm finds solution up to $\alpha \asymp \kappa^2$ (Bansal Spencer 20)

Symmetric Ising perceptron

Average case setting of **discrepancy minimization**: for $M/N = \alpha$, $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N \overset{\text{IID}}{\sim} \mathcal{N}(0, I_M)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ such that

$$\Big\| \sum_{i=1}^{N} \sigma_i \boldsymbol{g}^i \Big\|_\infty \leqslant \kappa\sqrt{N} \qquad (\kappa \ll 1)$$

- Solutions exist up to $\alpha \approx \frac{1}{\log_2(1/\kappa)}$ (Abbe Li Sly 21)
- Algorithm finds solution up to $\alpha \asymp \kappa^2$ (Bansal Spencer 20)

Recall $\mathcal{A} : \boldsymbol{G} = (\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N) \to \boldsymbol{\sigma}$ stable if for $(1-\varepsilon)$-correlated $\boldsymbol{G}, \boldsymbol{G}'$,

$$\mathbb{E}\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2^2 \lesssim \varepsilon N$$

Symmetric Ising perceptron

Average case setting of **discrepancy minimization**: for $M/N = \alpha$, $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N \overset{\text{IID}}{\sim} \mathcal{N}(0, I_M)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ such that

$$\Big\| \sum_{i=1}^{N} \sigma_i \boldsymbol{g}^i \Big\|_\infty \leqslant \kappa\sqrt{N} \qquad (\kappa \ll 1)$$

- Solutions exist up to $\alpha \approx \frac{1}{\log_2(1/\kappa)}$ (Abbe Li Sly 21)
- Algorithm finds solution up to $\alpha \asymp \kappa^2$ (Bansal Spencer 20)

Recall $\mathcal{A} : \boldsymbol{G} = (\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N) \to \boldsymbol{\sigma}$ stable if for $(1-\varepsilon)$-correlated $\boldsymbol{G}, \boldsymbol{G}'$,

$$\mathbb{E}\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2^2 \lesssim \varepsilon N$$

Theorem (Gamarnik Kızıldağ Perkins Xu 22)

Stable algorithms don't beat $\alpha_{\text{alg}} \lesssim \kappa^2 \log \frac{1}{\kappa}$.

Symmetric Ising perceptron

Average case setting of **discrepancy minimization**: for $M/N = \alpha$, $\boldsymbol{g}^1, \dots, \boldsymbol{g}^N \stackrel{\text{IID}}{\sim} \mathcal{N}(0, I_M)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ such that

$$\Big\| \sum_{i=1}^{N} \sigma_i \boldsymbol{g}^i \Big\|_\infty \leqslant \kappa \sqrt{N} \qquad (\kappa \ll 1)$$

- Solutions exist up to $\alpha \approx \frac{1}{\log_2(1/\kappa)}$ (Abbe Li Sly 21)
- Algorithm finds solution up to $\alpha \asymp \kappa^2$ (Bansal Spencer 20)

Recall $\mathcal{A} : \boldsymbol{G} = (\boldsymbol{g}^1, \dots, \boldsymbol{g}^N) \to \boldsymbol{\sigma}$ stable if for $(1-\varepsilon)$-correlated $\boldsymbol{G}, \boldsymbol{G}'$,

$$\mathbb{E}\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|_2^2 \lesssim \varepsilon N$$

Theorem (Gamarnik Kızıldağ Perkins Xu 22)

Stable algorithms don't beat $\alpha_{\text{alg}} \lesssim \kappa^2 \log \frac{1}{\kappa}$.

(This is sharp, matching algorithm of H Sellke Sun 25^+)

Symmetric Ising perceptron satisfies star OGP

Problem satisfies **star** OGP at $\alpha \gtrsim \kappa^2 \log \frac{1}{\kappa}$.

Symmetric Ising perceptron satisfies star OGP

Problem satisfies **star** OGP at $\alpha \gtrsim \kappa^2 \log \frac{1}{\kappa}$. That is, for some m, q_{ogp}

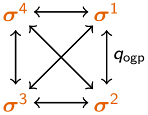
- Sample any (possibly correlated) $\boldsymbol{G}^1, \ldots, \boldsymbol{G}^C$ where $C = O(1)$

Symmetric Ising perceptron satisfies star OGP

Problem satisfies **star** OGP at $\alpha \gtrsim \kappa^2 \log \frac{1}{\kappa}$. That is, for some m, q_{ogp}

- Sample any (possibly correlated) $\boldsymbol{G}^1, \dots, \boldsymbol{G}^C$ where $C = O(1)$
- Whp there don't exist $\boldsymbol{\sigma}^1, \dots, \boldsymbol{\sigma}^m \in \{\pm 1\}^N$, each solving some $\boldsymbol{G}^i$, with

$$(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^j)/N \in [q_{\mathrm{ogp}}, q_{\mathrm{ogp}} + \delta], \qquad \text{for all } 1 \leqslant i < j \leqslant m$$

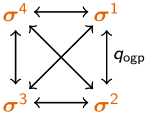


Symmetric Ising perceptron satisfies star OGP

Problem satisfies **star** OGP at $\alpha \gtrsim \kappa^2 \log \frac{1}{\kappa}$. That is, for some m, q_{ogp}

- Sample any (possibly correlated) $\boldsymbol{G}^1, \ldots, \boldsymbol{G}^C$ where $C = O(1)$
- Whp there don't exist $\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^m \in \{\pm 1\}^N$, each solving some $\boldsymbol{G}^i$, with

$$(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^j)/N \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta], \qquad \text{for all } 1 \leqslant i < j \leqslant m$$

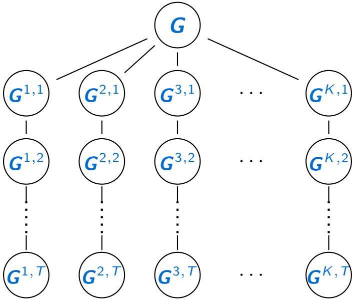


But ... how to construct this structure with a **stable** algorithm?

All we know: for $(1-\varepsilon)$-correlated $\boldsymbol{G}, \boldsymbol{G}'$, $\|\mathcal{A}(\boldsymbol{G}) - \mathcal{A}(\boldsymbol{G}')\|$ small whp

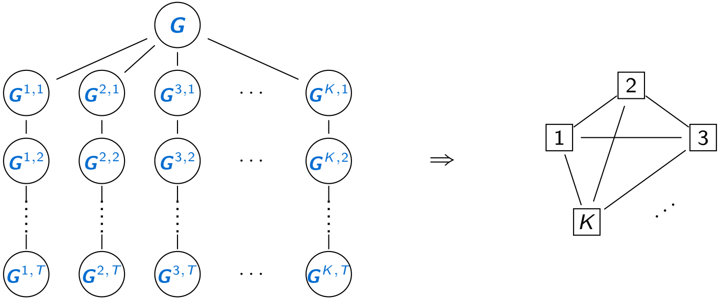
Ramsey-theoretic construction of forbidden structure

Construct K independent resample paths (K, $T = O(1)$, K large)



Ramsey-theoretic construction of forbidden structure

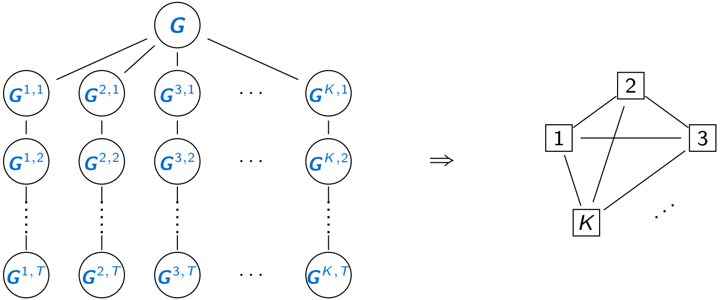
Construct K independent resample paths (K, $T = O(1)$, K large)



Color edges of complete graph on $[K]$:

Ramsey-theoretic construction of forbidden structure

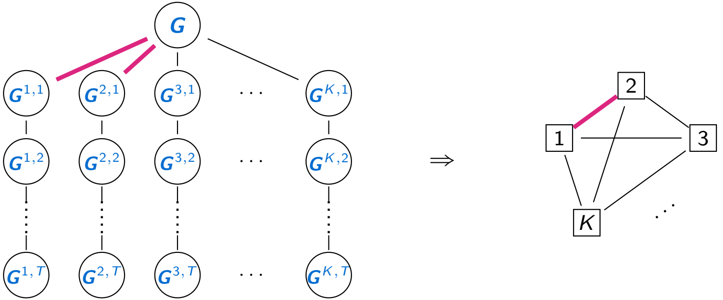
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

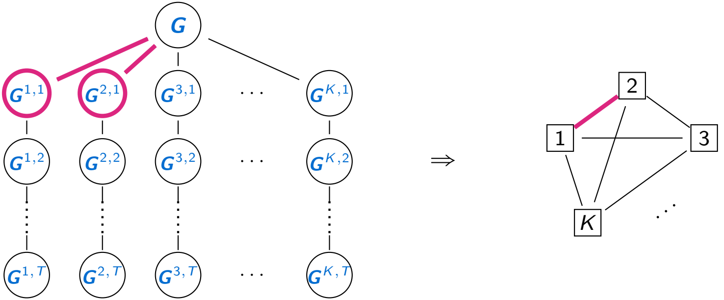
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\mathrm{ogp}}, q_{\mathrm{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

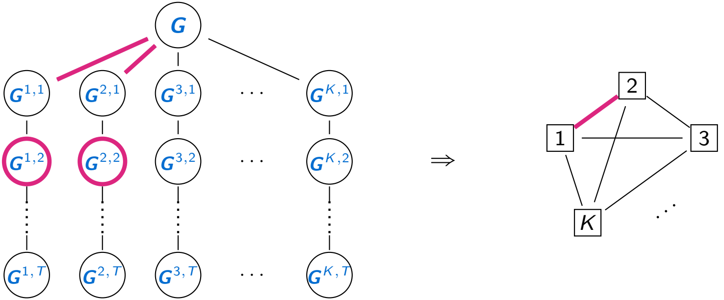
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\mathrm{ogp}}, q_{\mathrm{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

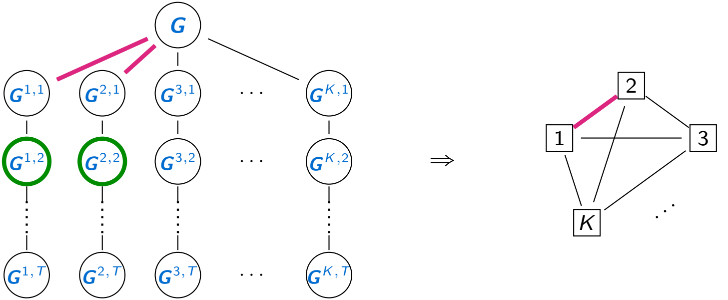
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

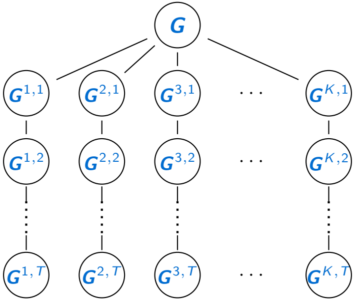
Construct K independent resample paths ($K, T = O(1)$, K large)



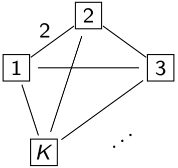
Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\rm ogp}, q_{\rm ogp} + \delta]$

Ramsey-theoretic construction of forbidden structure

Construct K independent resample paths ($K, T = O(1)$, K large)



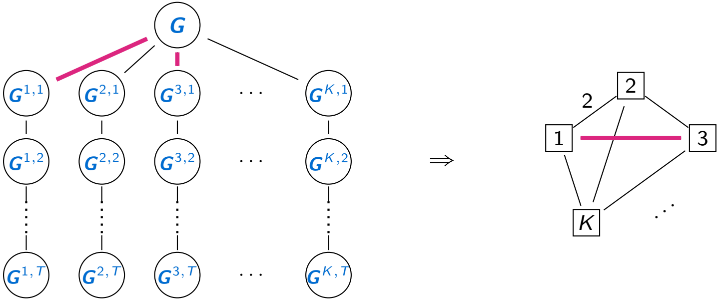
$\Rightarrow$



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

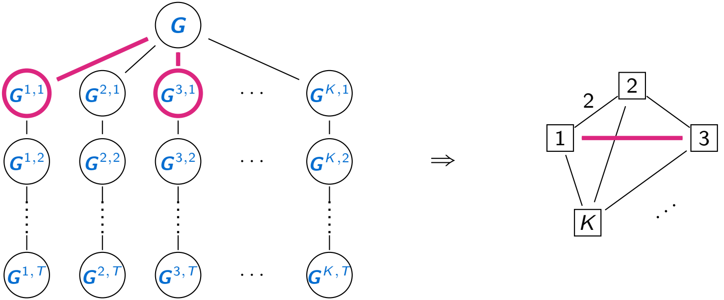
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

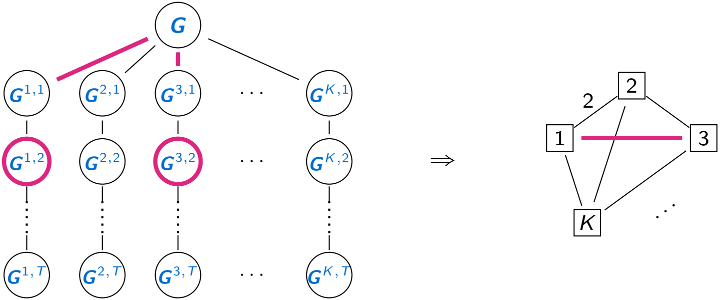
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\mathrm{ogp}}, q_{\mathrm{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

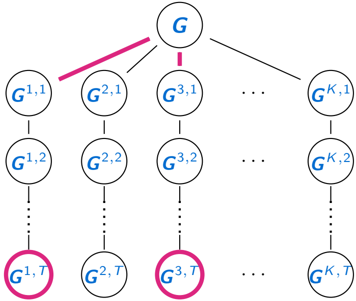
Construct K independent resample paths ($K, T = O(1)$, K large)



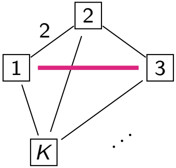
Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

Construct K independent resample paths ($K, T = O(1)$, K large)



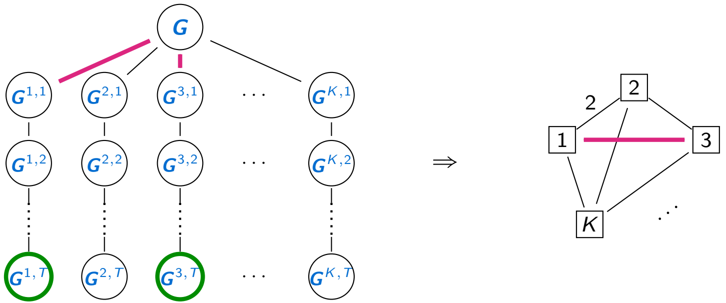
$\Rightarrow$



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

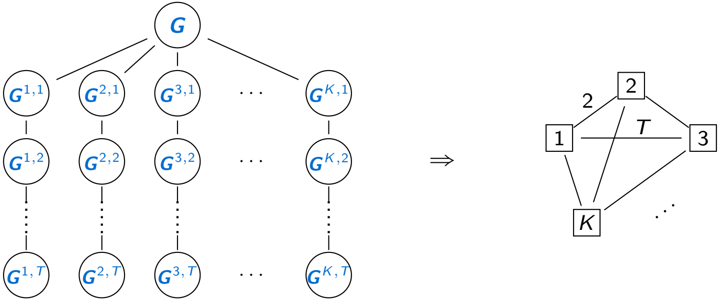
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

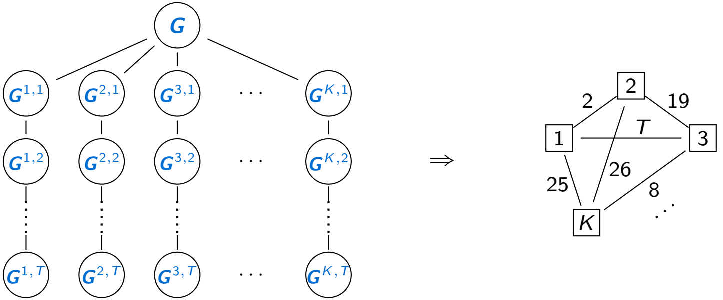
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

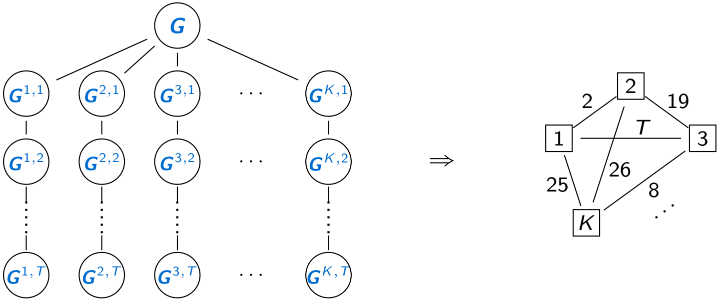
Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i,j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Ramsey-theoretic construction of forbidden structure

Construct K independent resample paths ($K, T = O(1)$, K large)

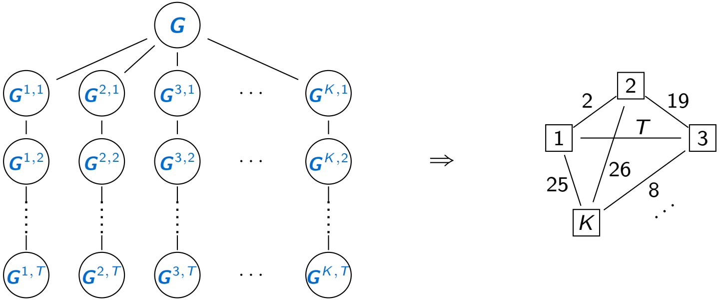


Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Erdős Szekeres 35: if $K \geqslant T^{Tm}$, exists monochromatic m-clique

Ramsey-theoretic construction of forbidden structure

Construct K independent resample paths ($K, T = O(1)$, K large)



Color edges of complete graph on $[K]$:
(i, j) colored by smallest t with $\|\mathcal{A}(\boldsymbol{G}^{i,t}) - \mathcal{A}(\boldsymbol{G}^{j,t})\| \in [q_{\text{ogp}}, q_{\text{ogp}} + \delta]$

Erdős Szekeres 35: if $K \geqslant T^{Tm}$, exists monochromatic m-clique
These form the star configuration!

Hardness for online algorithms

Q: can we show separation between **online** and **general** algorithms?

Hardness for online algorithms

Q: can we show separation between **online** and **general** algorithms?

For $M/N = \alpha$, $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N \overset{\text{IID}}{\sim} \mathcal{N}(0, I_M)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ such that

$$\Big\| \sum_{i=1}^{N} \sigma_i \boldsymbol{g}^i \Big\|_\infty \leqslant \kappa \sqrt{N} \qquad (\kappa \ll 1)$$

$\mathcal{A} : \boldsymbol{G} = (\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N) \to \boldsymbol{\sigma}$ online if σ_i depends only on $(\boldsymbol{g}^1, \ldots, \boldsymbol{g}^i)$

Hardness for online algorithms

Q: can we show separation between **online** and **general** algorithms?

For $M/N = \alpha$, $\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N \overset{\text{IID}}{\sim} \mathcal{N}(0, I_M)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ such that

$$\left\| \sum_{i=1}^N \sigma_i \boldsymbol{g}^i \right\|_\infty \leqslant \kappa\sqrt{N} \qquad (\kappa \ll 1)$$

$\mathcal{A} : \boldsymbol{G} = (\boldsymbol{g}^1, \ldots, \boldsymbol{g}^N) \to \boldsymbol{\sigma}$ online if σ_i depends only on $(\boldsymbol{g}^1, \ldots, \boldsymbol{g}^i)$

Algorithm of Bansal Spencer 20 achieving $\alpha \asymp \kappa^2$ is online

Hardness for online algorithms

Q: can we show separation between **online** and **general** algorithms?

For $M/N = \alpha$, $\boldsymbol{g}^1, \dots, \boldsymbol{g}^N \overset{\text{IID}}{\sim} \mathcal{N}(0, I_M)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ such that

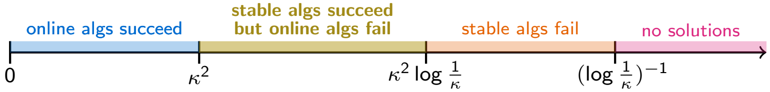
$$\Big\| \sum_{i=1}^{N} \sigma_i \boldsymbol{g}^i \Big\|_\infty \leqslant \kappa\sqrt{N} \qquad (\kappa \ll 1)$$

$\mathcal{A} : \boldsymbol{G} = (\boldsymbol{g}^1, \dots, \boldsymbol{g}^N) \to \boldsymbol{\sigma}$ online if σ_i depends only on $(\boldsymbol{g}^1, \dots, \boldsymbol{g}^i)$

Algorithm of Bansal Spencer 20 achieving $\alpha \asymp \kappa^2$ is online

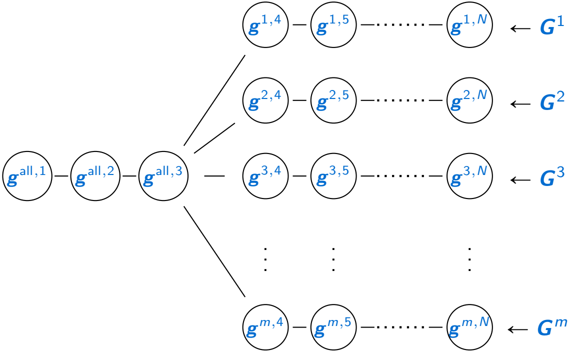
Theorem (Gamarnik Kızıldağ Perkins Xu 23)

Online algorithms cannot beat $\alpha_{\text{online}} \lesssim \kappa^2$



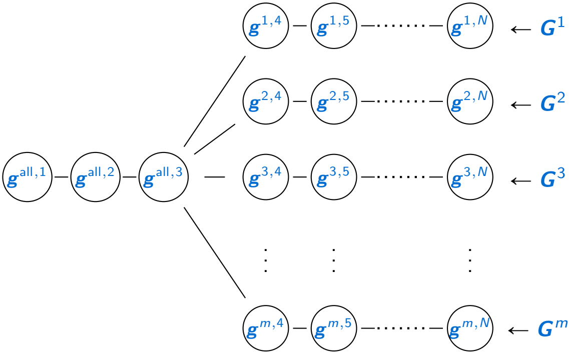
Online OGP

Consider "online-correlated" problem instances $(\boldsymbol{G}^1, \ldots, \boldsymbol{G}^m)$: identical then independent



Online OGP

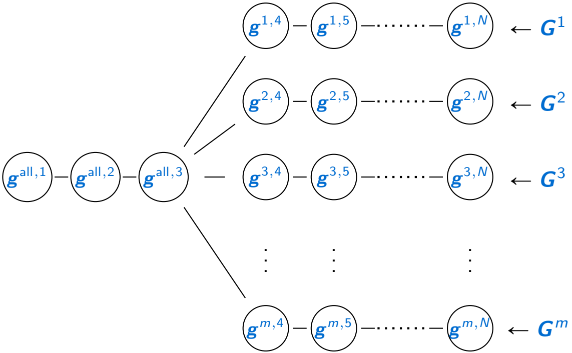
Consider "online-correlated" problem instances $(\boldsymbol{G}^1, \ldots, \boldsymbol{G}^m)$: identical then independent



More structure in outputs: $\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$ agree in entries where $\boldsymbol{g}^i$ shared

Online OGP

Consider "online-correlated" problem instances $(\boldsymbol{G}^1, \ldots, \boldsymbol{G}^m)$: identical then independent



More structure in outputs: $\boldsymbol{\sigma}^t = \mathcal{A}(\boldsymbol{G}^t)$ agree in entries where $\boldsymbol{g}^i$ shared

$\Rightarrow$ easier to show this doesn't exist in solution landscape

Outline of part II: a survey on the overlap gap property

Introduction and motivating problems

Overlap gap property: the basics

More OGPs and algorithm classes

Further enhancements

Hardness of finding strict local maxima

Strong low degree hardness

Can algorithms find strict local maxima?

Biroli 99, Müller Wyart 15, Parisi 17 prediction:

In glassy systems, low-temperature Glauber/Langevin dynamics fail to find a **stable (i.e. strongly concave) local max**

Can algorithms find strict local maxima?

Biroli 99, Müller Wyart 15, Parisi 17 prediction:

In glassy systems, low-temperature Glauber/Langevin dynamics fail to find a **stable (i.e. strongly concave) local max**

Baity-Jesi Sagun Geiger Spigler Ben Arous Cammarota LeCun Wyart Biroli 18:

during the training process [of a deep neural network] the dynamics slows down because of an increasingly **large number of flat directions**

Can algorithms find strict local maxima?

Biroli 99, Müller Wyart 15, Parisi 17 prediction:

In glassy systems, low-temperature Glauber/Langevin dynamics fail to find a **stable (i.e. strongly concave) local max**

Baity-Jesi Sagun Geiger Spigler Ben Arous Cammarota LeCun Wyart Biroli 18:

during the training process [of a deep neural network] the dynamics slows down because of an increasingly **large number of flat directions**

Behrens Arpino Kivva Zdeborová 22, Minzer Sah Sawhney 24 conjecture:
All efficient algorithms fail to find a stable local max

Notion of strict local max: gapped states

SK model: Hamiltonian $H : \{\pm 1\}^N \to \mathbb{R}$ defined by

$$H(\boldsymbol{\sigma}) = \frac{1}{2}(\boldsymbol{W}\boldsymbol{\sigma}, \boldsymbol{\sigma}), \qquad \boldsymbol{W} \sim \mathrm{GOE}(N).$$

Notion of strict local max: gapped states

SK model: Hamiltonian $H : \{\pm 1\}^N \to \mathbb{R}$ defined by

$$H(\boldsymbol{\sigma}) = \frac{1}{2}(\boldsymbol{W}\boldsymbol{\sigma}, \boldsymbol{\sigma}), \qquad \boldsymbol{W} \sim \mathrm{GOE}(N).$$

$\boldsymbol{\sigma} \in \{\pm 1\}^N$ is a γ-gapped state of H if

$$H(\boldsymbol{\sigma}) - H(\boldsymbol{\sigma} \oplus \boldsymbol{e}_i) \geqslant \gamma \qquad \forall i \in [N].$$

Notion of strict local max: gapped states

SK model: Hamiltonian $H : \{\pm 1\}^N \to \mathbb{R}$ defined by

$$H(\boldsymbol{\sigma}) = \frac{1}{2}(\boldsymbol{W}\boldsymbol{\sigma}, \boldsymbol{\sigma}), \qquad \boldsymbol{W} \sim \mathrm{GOE}(N).$$

$\boldsymbol{\sigma} \in \{\pm 1\}^N$ is a γ-gapped state of H if

$$H(\boldsymbol{\sigma}) - H(\boldsymbol{\sigma} \oplus \boldsymbol{e}_i) \geqslant \gamma \qquad \forall i \in [N].$$

Q: can an efficient algorithm find a gapped state?

Hardness of finding gapped states

$\boldsymbol{\sigma} \in \{\pm 1\}^N$ is a γ-gapped state of H if

$$H(\boldsymbol{\sigma}) - H(\boldsymbol{\sigma} \oplus \boldsymbol{e}_i) \geqslant \gamma \qquad \forall i \in [N] \,.$$

Hardness of finding gapped states

$\sigma \in \{\pm 1\}^N$ is a γ-gapped state of H if

$$H(\sigma) - H(\sigma \oplus e_i) \geqslant \gamma \qquad \forall i \in [N]\,.$$

Theorem (Dandi Gamarnik Zdeborová 23, Minzer Sah Sawhney 24)
There exists an explicit $\gamma_c > 0$ such that

$$\lim_{N\to\infty} \mathbb{P}(H \text{ has a } \gamma\text{-gapped state}) = \begin{cases} 1 & \gamma < \gamma_c\,, \\ 0 & \gamma > \gamma_c\,. \end{cases}$$

Hardness of finding gapped states

$\boldsymbol{\sigma} \in \{\pm 1\}^N$ is a γ-gapped state of H if

$$H(\boldsymbol{\sigma}) - H(\boldsymbol{\sigma} \oplus \boldsymbol{e}_i) \geqslant \gamma \qquad \forall i \in [N] \,.$$

Theorem (Dandi Gamarnik Zdeborová 23, Minzer Sah Sawhney 24)

There exists an explicit $\gamma_c > 0$ such that

$$\lim_{N \to \infty} \mathbb{P}(H \textit{ has a } \gamma\textit{-gapped state}) = \begin{cases} 1 & \gamma < \gamma_c \,, \\ 0 & \gamma > \gamma_c \,. \end{cases}$$

If $\gamma < \gamma_c$, there are $\geqslant e^{c(\gamma)N}$ γ-gapped states whp

Hardness of finding gapped states

$\sigma \in \{\pm 1\}^N$ is a γ-gapped state of H if

$$H(\sigma) - H(\sigma \oplus e_i) \geqslant \gamma \qquad \forall i \in [N] .$$

Theorem (Dandi Gamarnik Zdeborová 23, Minzer Sah Sawhney 24)

There exists an explicit $\gamma_c > 0$ such that

$$\lim_{N \to \infty} \mathbb{P}(H \text{ has a } \gamma\text{-gapped state}) = \begin{cases} 1 & \gamma < \gamma_c , \\ 0 & \gamma > \gamma_c . \end{cases}$$

If $\gamma < \gamma_c$, there are $\geqslant e^{c(\gamma)N}$ γ-gapped states whp

Theorem (H Sellke 25)

For any (constant) $\gamma > 0$ and degree $o(N)$ polynomial $\mathcal{A}$,

$$\mathbb{P}(\mathcal{A}(H) \text{ is a } \gamma\text{-gapped state of } H) = o(1)$$

Hardness of finding gapped states

$\sigma \in \{\pm 1\}^N$ is a γ-gapped state of H if

$$H(\sigma) - H(\sigma \oplus e_i) \geqslant \gamma \qquad \forall i \in [N].$$

Theorem (Dandi Gamarnik Zdeborová 23, Minzer Sah Sawhney 24)

There exists an explicit $\gamma_c > 0$ such that

$$\lim_{N\to\infty} \mathbb{P}(H \text{ has a } \gamma\text{-gapped state}) = \begin{cases} 1 & \gamma < \gamma_c\,, \\ 0 & \gamma > \gamma_c\,. \end{cases}$$

If $\gamma < \gamma_c$, there are $\geqslant e^{c(\gamma)N}$ γ-gapped states whp

Theorem (H Sellke 25)

For any (constant) $\gamma > 0$ and degree $o(N)$ polynomial $\mathcal{A}$,

$$\mathbb{P}(\mathcal{A}(H) \text{ is a } \gamma\text{-gapped state of } H) = o(1)$$

Low degree heuristic $\Rightarrow$ suggests failure of any $e^{o(N)}$ time algorithm!

Conditional OGP

Let H^0, H^1 be $(1-\varepsilon)$-correlated, i.e. $(\boldsymbol{W}^0_{i,j}, \boldsymbol{W}^1_{i,j})$ have correlation $1-\varepsilon$

Conditional OGP

Let H^0, H^1 be $(1-\varepsilon)$-correlated, i.e. $(\boldsymbol{W}^0_{i,j}, \boldsymbol{W}^1_{i,j})$ have correlation $1-\varepsilon$ (ε small depending on gappedness parameter γ)

Conditional OGP

Let H^0, H^1 be $(1-\varepsilon)$-correlated, i.e. $(\boldsymbol{W}^0_{i,j}, \boldsymbol{W}^1_{i,j})$ have correlation $1-\varepsilon$ (ε small depending on gappedness parameter γ)

Lemma (Conditional OGP)

If $\boldsymbol{\sigma}^0$ *is a* γ*-gapped state of* H^0 **depending only on** H^0, *then*

$$\mathbb{P}_{H^1}\Big(\exists\ \gamma\text{-gapped state } \boldsymbol{\sigma}^1 \text{ of } H^1 \text{ with } \Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1) \in [\varepsilon^3 N, c(\gamma) N] \Big) \leqslant e^{-cN}$$

Conditional OGP

Let H^0, H^1 be $(1-\varepsilon)$-correlated, i.e. $(\boldsymbol{W}^0_{i,j}, \boldsymbol{W}^1_{i,j})$ have correlation $1-\varepsilon$ (ε small depending on gappedness parameter γ)

Lemma (Conditional OGP)

If $\boldsymbol{\sigma}^0$ is a γ-gapped state of H^0 **depending only on** *H^0, then*

$$\mathbb{P}_{H^1}\Big(\exists\ \gamma\text{-gapped state } \boldsymbol{\sigma}^1 \text{ of } H^1 \text{ with } \Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1) \in [\varepsilon^3 N, c(\gamma)N] \Big) \leqslant e^{-cN}$$

Contrast with earlier OGPs: there **can** exist γ-gapped states $(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1)$ of (H^0, H^1) with $\Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1) \in [\varepsilon^3 N, c(\gamma)N]$.

Conditional OGP

Let H^0, H^1 be $(1-\varepsilon)$-correlated, i.e. $(\boldsymbol{W}^0_{i,j}, \boldsymbol{W}^1_{i,j})$ have correlation $1-\varepsilon$ (ε small depending on gappedness parameter γ)

Lemma (Conditional OGP)

If $\boldsymbol{\sigma}^0$ is a γ-gapped state of H^0 **depending only on** H^0*, then*

$$\mathbb{P}_{H^1}\Big(\exists\ \gamma\text{-gapped state } \boldsymbol{\sigma}^1 \text{ of } H^1 \text{ with } \Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1) \in [\varepsilon^3 N, c(\gamma)N] \Big) \leqslant e^{-cN}$$

Contrast with earlier OGPs: there **can** exist γ-gapped states $(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1)$ of (H^0, H^1) with $\Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1) \in [\varepsilon^3 N, c(\gamma)N]$.

But, **given only** H^0, can't predict which $\boldsymbol{\sigma}^0$ will be part of such pairs.

Conditional OGP

Let H^0, H^1 be $(1-\varepsilon)$-correlated, i.e. $(\boldsymbol{W}^0_{i,j}, \boldsymbol{W}^1_{i,j})$ have correlation $1-\varepsilon$ (ε small depending on gappedness parameter γ)

Lemma (Conditional OGP)

If $\boldsymbol{\sigma}^0$ is a γ-gapped state of H^0 **depending only on** H^0, *then*

$$\mathbb{P}_{H^1}\Big(\exists\ \gamma\text{-gapped state } \boldsymbol{\sigma}^1 \text{ of } H^1 \text{ with } \Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1) \in [\underset{\ll \varepsilon}{\underset{\uparrow}{\varepsilon^3}} N, c(\gamma) N] \Big) \leqslant e^{-cN}$$

Contrast with earlier OGPs: there **can** exist γ-gapped states $(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1)$ of (H^0, H^1) with $\Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1) \in [\varepsilon^3 N, c(\gamma) N]$.

But, **given only** H^0, can't predict which $\boldsymbol{\sigma}^0$ will be part of such pairs.

Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated.

Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

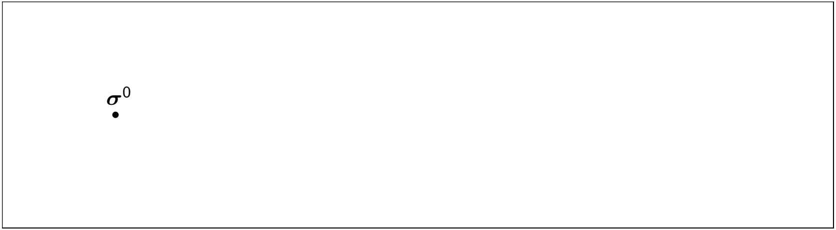
(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is $\boldsymbol{\gamma}$-gapped for H^i.

Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is γ-gapped for H^i.

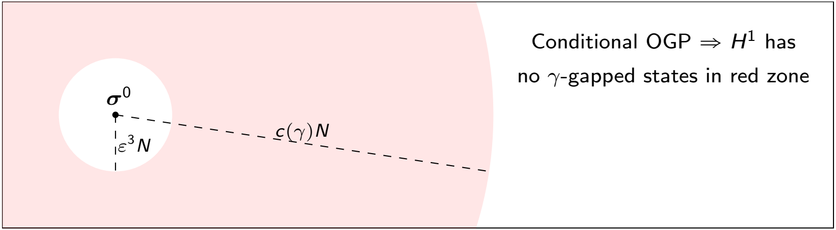


Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is $\boldsymbol{\gamma}$-gapped for H^i.

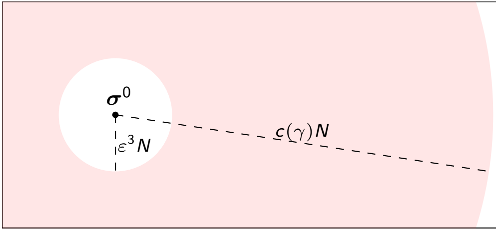


Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is $\boldsymbol{\gamma}$-gapped for H^i.



Conditional OGP $\Rightarrow H^1$ has no γ-gapped states in red zone

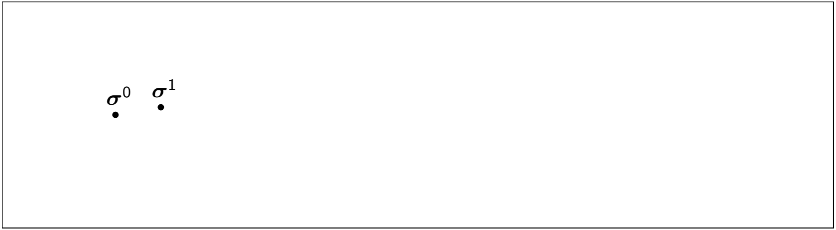
Stability of $\mathcal{A} \Rightarrow \boldsymbol{\sigma}^1$ has Hamming distance $\leqslant c(\gamma)N$ to $\boldsymbol{\sigma}^0$

Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is γ-gapped for H^i.

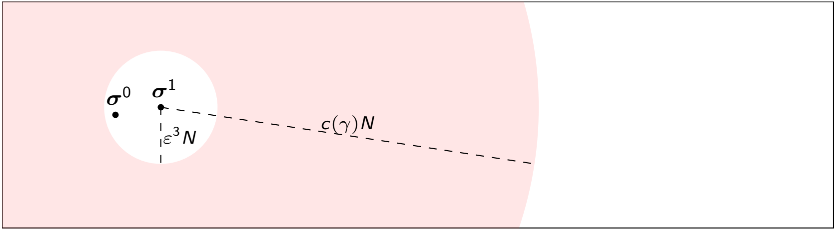


Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is γ-gapped for H^i.

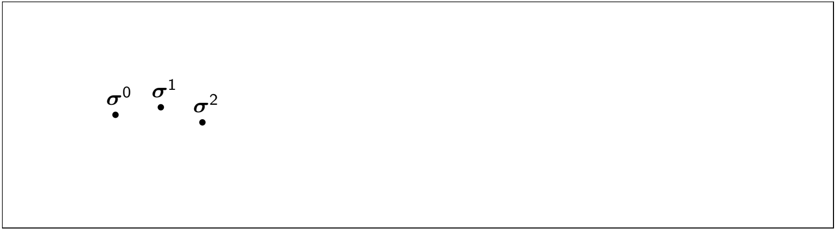


Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is γ-gapped for H^i.

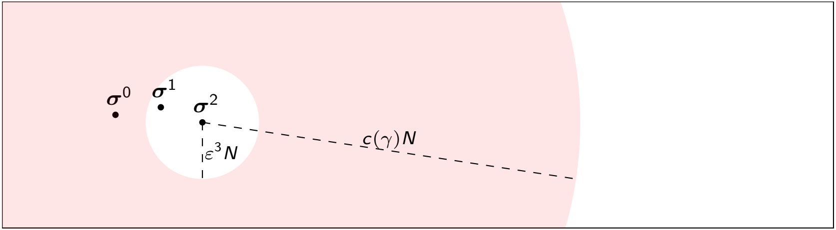


Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is γ-gapped for H^i.

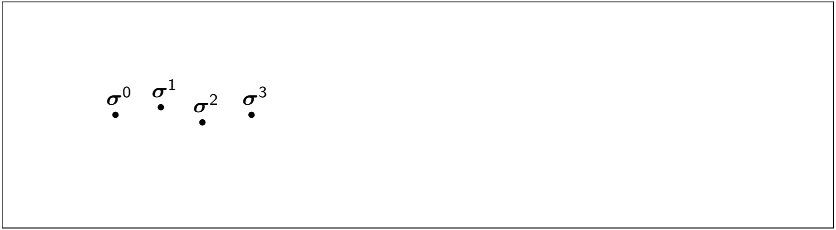


Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is γ-gapped for H^i.

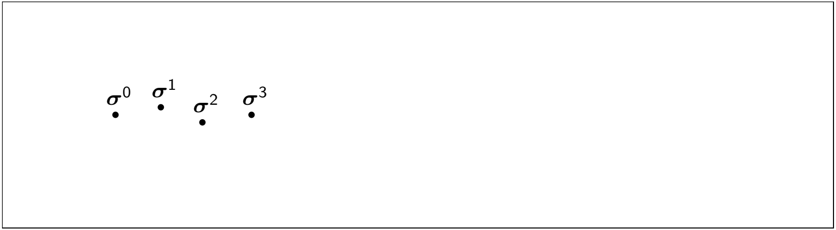


Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is γ-gapped for H^i.



So $H^0, H^{1/\varepsilon^2}$ have γ-gapped states $\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^{1/\varepsilon^2}$ with

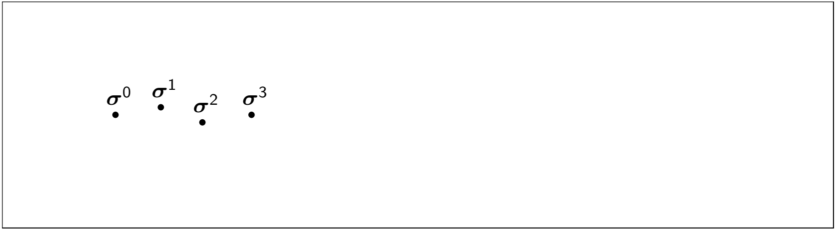
$$\Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^{1/\varepsilon^2}) \leqslant \frac{1}{\varepsilon^2} \cdot \varepsilon^3 N = \varepsilon N\,.$$

Conditional OGP to hardness

Consider Markovian sequence of Hamiltonians:

$$H^0 \to H^1 \to H^2 \to \cdots \to H^{1/\varepsilon^2}$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Suppose $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ is γ-gapped for H^i.



So $H^0, H^{1/\varepsilon^2}$ have γ-gapped states $\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^{1/\varepsilon^2}$ with

$$\Delta(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^{1/\varepsilon^2}) \leqslant \frac{1}{\varepsilon^2} \cdot \varepsilon^3 N = \varepsilon N\,.$$

Not possible because $H^0, H^{1/\varepsilon^2}$ nearly independent!

Hardness for Langevin dynamics on spherical models

Consider mixed p-spin glass

$$H(\boldsymbol{\sigma}) = \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \langle \boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \rangle , \qquad \boldsymbol{G}^{(p)}_{i_1,\ldots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$$

on spherical domain $S_N = \sqrt{N}\mathbb{S}^{N-1}$.

Hardness for Langevin dynamics on spherical models

Consider mixed p-spin glass

$$H(\boldsymbol{\sigma}) = \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \left(\boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \right), \qquad \boldsymbol{G}^{(p)}_{i_1,\dots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$$

on spherical domain $S_N = \sqrt{N}\mathbb{S}^{N-1}$.

Notion of strict local max: $\boldsymbol{\sigma}$ is a (γ, δ)-stable well if

$$\|\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma})\| \leqslant \delta \sqrt{N}, \qquad \nabla^2_{\mathrm{sp}} H(\boldsymbol{\sigma}) \preceq -\gamma I\,.$$

Hardness for Langevin dynamics on spherical models

Consider mixed p-spin glass

$$H(\boldsymbol{\sigma}) = \sum_{p \geqslant 2} \frac{\gamma_p}{N^{(p-1)/2}} \left(\boldsymbol{G}^{(p)}, \boldsymbol{\sigma}^{\otimes p} \right), \qquad \boldsymbol{G}^{(p)}_{i_1,\ldots,i_p} \overset{IID}{\sim} \mathcal{N}(0,1)$$

on spherical domain $S_N = \sqrt{N}\mathbb{S}^{N-1}$.

Notion of strict local max: $\boldsymbol{\sigma}$ is a (γ, δ)-stable well if

$$\|\nabla_{\mathrm{sp}} H(\boldsymbol{\sigma})\| \leqslant \delta\sqrt{N}, \qquad \nabla^2_{\mathrm{sp}} H(\boldsymbol{\sigma}) \preceq -\gamma I.$$

Theorem (H Sellke 25)

For any $\gamma > 0$, there exists $\delta > 0$ such that

$$\mathbb{P}(\textit{Low-temperature Langevin finds } (\gamma, \delta)\textit{-stable well in } O(1) \textit{ time}) \leqslant e^{-cN}$$

Outline of part II: a survey on the overlap gap property

Introduction and motivating problems

Overlap gap property: the basics

More OGPs and algorithm classes

Further enhancements

Hardness of finding strict local maxima

Strong low degree hardness

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \ldots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation
- This constellation does not exist in solution space of $(H^1, \ldots, H^T)$

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \ldots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation
- This constellation does not exist in solution space of $(H^1, \ldots, H^T)$

What does this actually imply for $p_{\mathsf{solve}} = \mathbb{P}(\mathcal{A}$ solves $H^1)$?

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \ldots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation
- This constellation does not exist in solution space of $(H^1, \ldots, H^T)$

(prob $1 - e^{-cN}$)

What does this actually imply for $p_{\mathsf{solve}} = \mathbb{P}(\mathcal{A}$ solves $H^1)$?

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \dots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation
- This constellation does not exist in solution space of $(H^1, \dots, H^T)$ **(prob $1 - e^{-cN}$)**

What does this actually imply for $p_{\text{solve}} = \mathbb{P}(\mathcal{A}$ solves $H^1)$?

If $\mathcal{A}$ **concentrates well** (e.g. local or Lipschitz algorithm, like AMP):

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \dots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation
- This constellation does not exist in solution space of $(H^1, \dots, H^T)$

(prob $1 - e^{-cN}$)

What does this actually imply for $p_{\mathrm{solve}} = \mathbb{P}(\mathcal{A}$ solves $H^1)$?

If $\mathcal{A}$ **concentrates well** (e.g. local or Lipschitz algorithm, like AMP):

- overlap $(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^j)$ concentrates $\forall\ (i, j)$

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \ldots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation **(prob $1 - e^{-cN}$)**
- This constellation does not exist in solution space of $(H^1, \ldots, H^T)$ **(prob $1 - e^{-cN}$)**

What does this actually imply for $p_{\text{solve}} = \mathbb{P}(\mathcal{A}$ solves $H^1)$?

If $\mathcal{A}$ **concentrates well** (e.g. local or Lipschitz algorithm, like AMP):

- overlap $(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^j)$ concentrates $\forall\ (i, j)$

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \dots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation **(prob** $1 - e^{-cN}$**)**
- This constellation does not exist in solution space of $(H^1, \dots, H^T)$ **(prob** $1 - e^{-cN}$**)**

What does this actually imply for $p_{\mathsf{solve}} = \mathbb{P}(\mathcal{A}$ **solves** $H^1)$**?**

If $\mathcal{A}$ **concentrates well** (e.g. local or Lipschitz algorithm, like AMP):

- overlap $(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^j)$ concentrates $\forall$ (i, j)
- algorithm's achieved value $H^i(\boldsymbol{\sigma}^i)$ concentrates $\forall$ i

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \ldots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation **(prob $1 - e^{-cN}$)**
- This constellation does not exist in solution space of $(H^1, \ldots, H^T)$ **(prob $1 - e^{-cN}$)**

What does this actually imply for $p_{\text{solve}} = \mathbb{P}(\mathcal{A}$ solves $H^1)$?

If $\mathcal{A}$ **concentrates well** (e.g. local or Lipschitz algorithm, like AMP):

- overlap $(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^j)$ concentrates $\forall\ (i,j)$
- algorithm's achieved value $H^i(\boldsymbol{\sigma}^i)$ concentrates $\forall\ i$

$\Big\}\ p_{\text{solve}} \leqslant e^{-cN}$

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \dots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation
- This constellation does not exist in solution space of $(H^1, \dots, H^T)$ **(prob $1 - e^{-cN}$)**

What does this actually imply for $p_{\mathrm{solve}} = \mathbb{P}(\mathcal{A}$ solves $H^1)$?

If $\mathcal{A}$ **concentrates well** (e.g. local or Lipschitz algorithm, like AMP):

- overlap $(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^j)$ concentrates $\forall\ (i,j)$
- algorithm's achieved value $H^i(\boldsymbol{\sigma}^i)$ concentrates $\forall\ i$

$\Big\}\ p_{\mathrm{solve}} \leqslant e^{-cN}$

What if we just know $\mathcal{A}$ is stable? (e.g. low degree polynomial)

A closer look at the OGP methodology

Following does not occur simultaneously:

- Algorithm $\mathcal{A}$ solves all $H^1, \dots, H^T$ in the correlated ensemble
- Outputs $\boldsymbol{\sigma}^i = \mathcal{A}(H^i)$ form the desired constellation
- This constellation does not exist in solution space of $(H^1, \dots, H^T)$

(prob $1 - e^{-cN}$)

What does this actually imply for $p_{\text{solve}} = \mathbb{P}(\mathcal{A}$ solves $H^1)$?

If $\mathcal{A}$ **concentrates well** (e.g. local or Lipschitz algorithm, like AMP):

- overlap $(\boldsymbol{\sigma}^i, \boldsymbol{\sigma}^j)$ concentrates $\forall\ (i, j)$
- algorithm's achieved value $H^i(\boldsymbol{\sigma}^i)$ concentrates $\forall\ i$

$\Big\}\ p_{\text{solve}} \leqslant e^{-cN}$

What if we just know $\mathcal{A}$ is stable? (e.g. low degree polynomial)
Union bound: $p_{\text{solve}} \leqslant 1 - 1/T$ ☹

A closer look at the OGP methodology

This issue quietly plagued the OGP literature for years. Numerous works prove $p_{\text{solve}} \leqslant 1 - e^{-\boldsymbol{D}}$ for degree $\boldsymbol{D}$ polynomials:

- Mean-field spin glass optimization (Gamarnik Jagannath Wein 20)
- Max independent set on $G(N, d/N)$ (GJW 20, Wein 20)
- Number partitioning problem (Gamarnik Kızıldağ 21)
- Random k-SAT (Bresler H 21)
- Symmetric / negative Ising perceptron (Gamarnik Perkins Kızıldağ Xu 22, Li Schramm Zhou 24)

A closer look at the OGP methodology

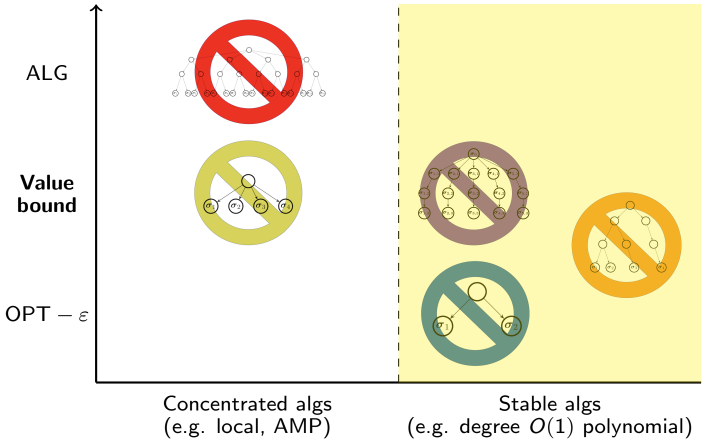
This issue quietly plagued the OGP literature for years. Numerous works prove $p_{\mathrm{solve}} \leqslant 1 - e^{-\boldsymbol{D}}$ for degree $\boldsymbol{D}$ polynomials:

- Mean-field spin glass optimization (Gamarnik Jagannath Wein 20)
- Max independent set on $G(N, d/N)$ (GJW 20, Wein 20)
- Number partitioning problem (Gamarnik Kızıldağ 21)
- Random k-SAT (Bresler H 21)
- Symmetric / negative Ising perceptron (Gamarnik Perkins Kızıldağ Xu 22, Li Schramm Zhou 24)

Open problem in Dec 2024 AIM workshop: *Low degree polynomial methods in average-case complexity*

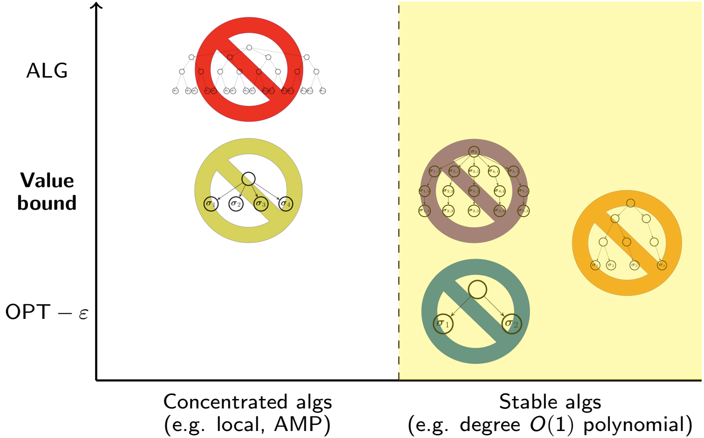
Strong low degree hardness

We give general method to overcome this issue, for **all stability-based** OGPs



Strong low degree hardness

We give general method to overcome this issue, for **all stability-based** OGPs



Actually, show $p_{\text{solve}} = o(1)$ for degrees **much larger** than $O(1)$.

Strong low degree hardness

Theorem (H Sellke 25, informal)

If a stability-based OGP obstruction holds with probability $1 - p_{\mathrm{ogp}}$, *then*

$$\mathbb{P}(\textit{a degree } \boldsymbol{D} = \widetilde{o}(\log \tfrac{1}{p_{\mathrm{ogp}}}) \textit{ algorithm succeeds}) = o(1)$$

Strong low degree hardness

Theorem (H Sellke 25, informal)

If a stability-based OGP obstruction holds with probability $1 - p_{\text{ogp}}$*, then*

$$\mathbb{P}(\textit{a degree } \boldsymbol{D} = \widetilde{o}(\log \tfrac{1}{p_{\text{ogp}}}) \textit{ algorithm succeeds}) = o(1)$$

- Max ind set on $G(N, d/N)$; random k-SAT; mean-field spin glasses: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N)$
- Symmetric / negative Ising perceptron: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N/\log N)$
- Max-clique in $G(N, 1/2)$: $p_{\text{ogp}} = e^{-c\log^2 N}$, $\boldsymbol{D} = o(\log^2 N)$

Strong low degree hardness

Theorem (H Sellke 25, informal)

If a stability-based OGP obstruction holds with probability $1 - p_{\text{ogp}}$*, then*

$$\mathbb{P}(\textit{a degree } \boldsymbol{D} = \widetilde{o}(\log \tfrac{1}{p_{\text{ogp}}}) \textit{ algorithm succeeds}) = o(1)$$

- Max ind set on $G(N, d/N)$; random k-SAT; mean-field spin glasses: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N)$
- Symmetric / negative Ising perceptron: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N/\log N)$
- Max-clique in $G(N, 1/2)$: $p_{\text{ogp}} = e^{-c\log^2 N}$, $\boldsymbol{D} = o(\log^2 N)$

Low degree heuristic (Hopkins 18): degree $\boldsymbol{D}$ polys $\approx e^{\boldsymbol{D}}$ time algs

Strong low degree hardness

Theorem (H Sellke 25, informal)

If a stability-based OGP obstruction holds with probability $1 - p_{\text{ogp}}$*, then*

$$\mathbb{P}(\textit{a degree } \boldsymbol{D} = \widetilde{o}(\log \tfrac{1}{p_{\text{ogp}}}) \textit{ algorithm succeeds}) = o(1)$$

- Max ind set on $G(N, d/N)$; random k-SAT; mean-field spin glasses: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N)$
- Symmetric / negative Ising perceptron: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N/\log N)$
- Max-clique in $G(N, 1/2)$: $p_{\text{ogp}} = e^{-c\log^2 N}$, $\boldsymbol{D} = o(\log^2 N)$

Low degree heuristic (Hopkins 18): degree $\boldsymbol{D}$ polys $\approx e^{\boldsymbol{D}}$ time algs
Suggests e.g. $e^{o(N)}$ time algorithms don't beat ALG in max ind set

Strong low degree hardness

Theorem (H Sellke 25, informal)

If a stability-based OGP obstruction holds with probability $1 - p_{\text{ogp}}$, *then*

$$\mathbb{P}(\textit{a degree } \boldsymbol{D} = \widetilde{o}(\log \tfrac{1}{p_{\text{ogp}}}) \textit{ algorithm succeeds}) = o(1)$$

- Max ind set on $G(N, d/N)$; random k-SAT; mean-field spin glasses: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N)$
- Symmetric / negative Ising perceptron: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N/\log N)$
- Max-clique in $G(N, 1/2)$: $p_{\text{ogp}} = e^{-c \log^2 N}$, $\boldsymbol{D} = o(\log^2 N)$

Low degree heuristic (Hopkins 18): degree $\boldsymbol{D}$ polys $\approx e^{\boldsymbol{D}}$ time algs
Suggests e.g. $e^{o(N)}$ time algorithms don't beat ALG in max ind set

This is tight: $\boldsymbol{D} = O(N)$ / time $e^{O(N)}$ can brute force!

Strong low degree hardness

Theorem (H Sellke 25, informal)

If a stability-based OGP obstruction holds with probability $1 - p_{\text{ogp}}$, *then*

$$\mathbb{P}(\textit{a degree } \boldsymbol{D} = \tilde{o}(\log \tfrac{1}{p_{\text{ogp}}}) \textit{ algorithm succeeds}) = o(1)$$

- Max ind set on $G(N, d/N)$; random k-SAT; mean-field spin glasses: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N)$
- Symmetric / negative Ising perceptron: $p_{\text{ogp}} = e^{-cN}$, $\boldsymbol{D} = o(N/\log N)$
- Max-clique in $G(N, 1/2)$: $p_{\text{ogp}} = e^{-c\log^2 N}$, $\boldsymbol{D} = o(\log^2 N)$

Low degree heuristic (Hopkins 18): degree $\boldsymbol{D}$ polys $\approx e^{\boldsymbol{D}}$ time algs
Suggests e.g. $e^{o(N)}$ time algorithms don't beat ALG in max ind set

This is tight: $\boldsymbol{D} = O(N)$ / time $e^{O(N)}$ can brute force!
(and in max-clique: $\boldsymbol{D} = O(\log^2 N)$ / time $e^{O(\log^2 N)}$ can brute force)

Strong low degree hardness: proof ideas

Let's revisit **ladder** OGP: consider Markovian sequence of Hamiltonians

$$H^0 \to H^1 \to \cdots \to H^T$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated.

Strong low degree hardness: proof ideas

Let's revisit **ladder** OGP: consider Markovian sequence of Hamiltonians

$$H^0 \to H^1 \to \cdots \to H^T$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Following doesn't occur simultaneously:

- $\mathcal{A}(H^i)$ solves H^i for all i
- $\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\|$ small for all i
- no forbidden structure in soln space of $(H^0, \ldots, H^T)$ (prob $1 - p_{\text{ogp}}$)
  (e.g. $1 - e^{-cN}$)

Strong low degree hardness: proof ideas

Let's revisit **ladder** OGP: consider Markovian sequence of Hamiltonians

$$H^0 \to H^1 \to \cdots \to H^T$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Following doesn't occur simultaneously:

- $\mathcal{A}(H^i)$ solves H^i for all i
- $\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\|$ small for all i
- no forbidden structure in soln space of $(H^0, \ldots, H^T)$ (prob $1 - p_{\text{ogp}}$)
  (e.g. $1 - e^{-cN}$)

💡 lower bound $\mathbb{P}$(solve all) and $\mathbb{P}$(all steps stable) by
positive correlation inequalities instead of **union bound**

Strong low degree hardness: proof ideas

Let's revisit **ladder** OGP: consider Markovian sequence of Hamiltonians

$$H^0 \to H^1 \to \cdots \to H^T$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Following doesn't occur simultaneously:

- $\mathcal{A}(H^i)$ solves H^i for all i
- $\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\|$ small for all i
- no forbidden structure in soln space of $(H^0, \ldots, H^T)$ (prob $1 - p_{\text{ogp}}$)
  (e.g. $1 - e^{-cN}$)

lower bound $\mathbb{P}$(solve all) and $\mathbb{P}$(all steps stable) by **positive correlation inequalities** instead of **union bound**

Proof of concept: for $\text{Stab}(i, i+1) = \{\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small}\}$

Strong low degree hardness: proof ideas

Let's revisit **ladder** OGP: consider Markovian sequence of Hamiltonians

$$H^0 \to H^1 \to \cdots \to H^T$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Following doesn't occur simultaneously:

- $\mathcal{A}(H^i)$ solves H^i for all i
- $\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\|$ small for all i
- no forbidden structure in soln space of $(H^0, \ldots, H^T)$ (prob $1 - p_{\text{ogp}}$)
  (e.g. $1 - e^{-cN}$)

💡 lower bound $\mathbb{P}$(solve all) and $\mathbb{P}$(all steps stable) by **positive correlation inequalities** instead of **union bound**

Proof of concept: for $\text{Stab}(i, i+1) = \{\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small}\}$

$$\mathbb{P}(\text{Stab}(0,1) \cap \text{Stab}(1,2)) = \mathbb{E}[\mathbb{P}(\text{Stab}(0,1) \cap \text{Stab}(1,2) | H^1)]$$

Strong low degree hardness: proof ideas

Let's revisit **ladder** OGP: consider Markovian sequence of Hamiltonians

$$H^0 \to H^1 \to \cdots \to H^T$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Following doesn't occur simultaneously:

- $\mathcal{A}(H^i)$ solves H^i for all i
- $\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\|$ small for all i
- no forbidden structure in soln space of $(H^0, \dots, H^T)$ (prob $1 - p_{\text{ogp}}$)
  (e.g. $1 - e^{-cN}$)

lower bound $\mathbb{P}$(solve all) and $\mathbb{P}$(all steps stable) by **positive correlation inequalities** instead of **union bound**

Proof of concept: for $\text{Stab}(i, i+1) = \{\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small}\}$

$$\mathbb{P}(\text{Stab}(0,1) \cap \text{Stab}(1,2)) = \mathbb{E}[\mathbb{P}(\text{Stab}(0,1) \cap \text{Stab}(1,2) | H^1)]$$

(reversibility) $= \mathbb{E}[\mathbb{P}(\text{Stab}(0,1)|H^1)^2]$

Strong low degree hardness: proof ideas

Let's revisit **ladder** OGP: consider Markovian sequence of Hamiltonians

$$H^0 \to H^1 \to \cdots \to H^T$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Following doesn't occur simultaneously:

- $\mathcal{A}(H^i)$ solves H^i for all i
- $\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\|$ small for all i
- no forbidden structure in soln space of $(H^0, \ldots, H^T)$ (prob $1 - p_{\text{ogp}}$)
  (e.g. $1 - e^{-cN}$)

💡 lower bound $\mathbb{P}(\text{solve all})$ and $\mathbb{P}(\text{all steps stable})$ by **positive correlation inequalities** instead of **union bound**

Proof of concept: for $\text{Stab}(i, i+1) = \{\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small}\}$

$$\begin{aligned} \mathbb{P}(\text{Stab}(0,1) \cap \text{Stab}(1,2)) &= \mathbb{E}[\mathbb{P}(\text{Stab}(0,1) \cap \text{Stab}(1,2) | H^1)] \\ \textbf{(reversibility)} \quad &= \mathbb{E}[\mathbb{P}(\text{Stab}(0,1) | H^1)^2] \\ \textbf{(Jensen)} \quad &\geqslant \mathbb{E}[\mathbb{P}(\text{Stab}(0,1) | H^1)]^2 \end{aligned}$$

Strong low degree hardness: proof ideas

Let's revisit **ladder** OGP: consider Markovian sequence of Hamiltonians

$$H^0 \to H^1 \to \cdots \to H^T$$

(H^i, H^{i+1}) is $(1-\varepsilon)$-correlated. Following doesn't occur simultaneously:

- $\mathcal{A}(H^i)$ solves H^i for all i
- $\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\|$ small for all i
- no forbidden structure in soln space of $(H^0, \ldots, H^T)$ (prob $1 - p_{\mathrm{ogp}}$) (e.g. $1 - e^{-cN}$)

lower bound $\mathbb{P}$(solve all) and $\mathbb{P}$(all steps stable) by **positive correlation inequalities** instead of **union bound**

Proof of concept: for $\mathrm{Stab}(i, i+1) = \{\|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small}\}$

$$\begin{aligned}
\mathbb{P}(\mathrm{Stab}(0,1) \cap \mathrm{Stab}(1,2)) &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,1) \cap \mathrm{Stab}(1,2) | H^1)] \\
\textbf{(reversibility)} \quad &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,1) | H^1)^2] \\
\textbf{(Jensen)} \quad &\geqslant \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,1) | H^1)]^2 = \mathbb{P}(\mathrm{Stab}(0,1))^2
\end{aligned}$$

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\mathbb{P}(\mathsf{Stab}(0,\ldots,T)) = \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\ldots,\tfrac{T}{2}) \cap \mathsf{Stab}(\tfrac{T}{2},\ldots,T) | H^{\frac{T}{2}})]$$

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\mathbb{P}(\mathsf{Stab}(0,\ldots,T)) = \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\ldots,\tfrac{T}{2}) \cap \mathsf{Stab}(\tfrac{T}{2},\ldots,T) | H^{\frac{T}{2}})]$$

(reversibility)
$$= \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\ldots,\tfrac{T}{2}) | H^{\frac{T}{2}})^2]$$

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\begin{aligned}
\mathbb{P}(\mathrm{Stab}(0,\ldots,T)) &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\ldots,\tfrac{T}{2}) \cap \mathrm{Stab}(\tfrac{T}{2},\ldots,T) | H^{\frac{T}{2}})] \\
\textbf{(reversibility)} \quad &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\ldots,\tfrac{T}{2}) | H^{\frac{T}{2}})^2] \\
\textbf{(Jensen)} \quad &\geqslant \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\ldots,\tfrac{T}{2}) | H^{\frac{T}{2}})]^2
\end{aligned}$$

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\mathbb{P}(\mathsf{Stab}(0,\dots,T)) = \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\dots,\tfrac{T}{2}) \cap \mathsf{Stab}(\tfrac{T}{2},\dots,T) | H^{\frac{T}{2}})]$$

$$\textbf{(reversibility)} \qquad = \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})^2]$$

$$\textbf{(Jensen)} \qquad \geqslant \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})]^2 = \mathbb{P}(\mathsf{Stab}(0,\dots,\tfrac{T}{2}))^2$$

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\begin{aligned}
\mathbb{P}(\mathsf{Stab}(0,\ldots,T)) &= \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\ldots,\tfrac{T}{2}) \cap \mathsf{Stab}(\tfrac{T}{2},\ldots,T)|H^{\frac{T}{2}})] \\
\textbf{(reversibility)} \quad &= \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\ldots,\tfrac{T}{2})|H^{\frac{T}{2}})^2] \\
\textbf{(Jensen)} \quad &\geqslant \mathbb{E}[\mathbb{P}(\mathsf{Stab}(0,\ldots,\tfrac{T}{2})|H^{\frac{T}{2}})]^2 = \mathbb{P}(\mathsf{Stab}(0,\ldots,\tfrac{T}{2}))^2 \\
\textbf{(iterate)} \quad &\geqslant \mathbb{P}(\mathsf{Stab}(0,1))^T \equiv p_{\mathsf{stable}}^T
\end{aligned}$$

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\mathbb{P}(\mathrm{Stab}(0,\dots,T)) = \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) \cap \mathrm{Stab}(\tfrac{T}{2},\dots,T) | H^{\frac{T}{2}})]$$

(reversibility) $$= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})^2]$$

(Jensen) $$\geqslant \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})]^2 = \mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}))^2$$

(iterate) $$\geqslant \mathbb{P}(\mathrm{Stab}(0,1))^T \equiv p_{\mathrm{stable}}^T$$

- Same proof: $\mathbb{P}(\mathcal{A} \text{ solve all } H^0,\dots,H^T) \geqslant \mathbb{P}(\mathcal{A} \text{ solve } H^0)^T \equiv p_{\mathrm{solve}}^T$

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\begin{aligned}
\mathbb{P}(\mathrm{Stab}(0,\dots,T)) &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) \cap \mathrm{Stab}(\tfrac{T}{2},\dots,T) | H^{\frac{T}{2}})] \\
\textbf{(reversibility)} \quad &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})^2] \\
\textbf{(Jensen)} \quad &\geqslant \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})]^2 = \mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}))^2 \\
\textbf{(iterate)} \quad &\geqslant \mathbb{P}(\mathrm{Stab}(0,1))^T \equiv p_{\mathrm{stable}}^T
\end{aligned}$$

- Same proof: $\mathbb{P}(\mathcal{A} \text{ solve all } H^0,\dots,H^T) \geqslant \mathbb{P}(\mathcal{A} \text{ solve } H^0)^T \equiv p_{\mathrm{solve}}^T$
- $\mathbb{P}(\text{all steps stable}) \geqslant p_{\mathrm{stable}}^T$
- $\mathbb{P}(\nexists \text{ forbidden structure}) = 1 - p_{\mathrm{ogp}}$ (e.g. $1 - e^{-cN}$)

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\begin{aligned}
\mathbb{P}(\mathrm{Stab}(0,\dots,T)) &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) \cap \mathrm{Stab}(\tfrac{T}{2},\dots,T) | H^{\frac{T}{2}})] \\
\textbf{(reversibility)} \quad &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})^2] \\
\textbf{(Jensen)} \quad &\geqslant \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})]^2 = \mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}))^2 \\
\textbf{(iterate)} \quad &\geqslant \mathbb{P}(\mathrm{Stab}(0,1))^T \equiv p_{\mathrm{stable}}^T
\end{aligned}$$

- Same proof: $\mathbb{P}(\mathcal{A} \text{ solve all } H^0, \dots, H^T) \geqslant \mathbb{P}(\mathcal{A} \text{ solve } H^0)^T \equiv p_{\mathrm{solve}}^T$
- $\mathbb{P}(\text{all steps stable}) \geqslant p_{\mathrm{stable}}^T$
- $\mathbb{P}(\nexists \text{ forbidden structure}) = 1 - p_{\mathrm{ogp}}$ (e.g. $1 - e^{-cN}$)

don't occur simultaneously $\Rightarrow$ $p_{\mathrm{solve}}^T + p_{\mathrm{stable}}^T \leqslant 1 + p_{\mathrm{ogp}}$

Strong low degree hardness: initial attempt

We can iterate this dyadically!

$$\begin{aligned}
\mathbb{P}(\mathrm{Stab}(0,\dots,T)) &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) \cap \mathrm{Stab}(\tfrac{T}{2},\dots,T) | H^{\frac{T}{2}})] \\
\textbf{(reversibility)} \quad &= \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})^2] \\
\textbf{(Jensen)} \quad &\geqslant \mathbb{E}[\mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}) | H^{\frac{T}{2}})]^2 = \mathbb{P}(\mathrm{Stab}(0,\dots,\tfrac{T}{2}))^2 \\
\textbf{(iterate)} \quad &\geqslant \mathbb{P}(\mathrm{Stab}(0,1))^T \equiv p_{\mathrm{stable}}^T
\end{aligned}$$

- Same proof: $\mathbb{P}(\mathcal{A} \text{ solve all } H^0, \dots, H^T) \geqslant \mathbb{P}(\mathcal{A} \text{ solve } H^0)^T \equiv p_{\mathrm{solve}}^T$
- $\mathbb{P}(\text{all steps stable}) \geqslant p_{\mathrm{stable}}^T$
- $\mathbb{P}(\nexists \text{ forbidden structure}) = 1 - p_{\mathrm{ogp}}$ (e.g. $1 - e^{-cN}$)

don't occur simultaneously $\Rightarrow$ $p_{\mathrm{solve}}^T + p_{\mathrm{stable}}^T \leqslant 1 + p_{\mathrm{ogp}}$

Doesn't yet imply $p_{\mathrm{solve}} = o(1)$ ☹

Strong low degree hardness via dyadic Jensen

do dyadic Jensen on merged event Solve&Stab$(0, \ldots, T)$:

$$\Big\{\mathcal{A} \text{ solves } H^0, \ldots H^T \text{ and } \|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small for } 0 \leqslant i \leqslant T-1\Big\}$$

Strong low degree hardness via dyadic Jensen

do dyadic Jensen on merged event Solve&Stab$(0, \ldots, T)$:

$$\left\{ \mathcal{A} \text{ solves } H^0, \ldots H^T \text{ and } \|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small for } 0 \leqslant i \leqslant T-1 \right\}$$

$$\begin{aligned} \mathbb{P}(\text{Solve\&Stab}(0, \ldots, T)) &\geqslant \mathbb{P}(\text{Solve\&Stab}(0, \ldots, \tfrac{T}{2}))^2 \\ &\geqslant \mathbb{P}(\text{Solve\&Stab}(0, 1))^T \end{aligned}$$

Strong low degree hardness via dyadic Jensen

do dyadic Jensen on merged event Solve&Stab$(0, \dots, T)$:

$$\left\{\mathcal{A} \text{ solves } H^0, \dots H^T \text{ and } \|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small for } 0 \leqslant i \leqslant T-1\right\}$$

$$\begin{aligned}\mathbb{P}(\text{Solve\&Stab}(0, \dots, T)) &\geqslant \mathbb{P}(\text{Solve\&Stab}(0, \dots, \tfrac{T}{2}))^2 \\ &\geqslant \mathbb{P}(\text{Solve\&Stab}(0, 1))^T\end{aligned}$$

Since $\mathbb{P}(\nexists \text{ forbidden structure}) = 1 - e^{-cN}$ & don't occur simultaneously:

$$\mathbb{P}(\text{Solve\&Stab}(0, 1)) \leqslant e^{-cN/T} \qquad \text{is small}$$

Strong low degree hardness via dyadic Jensen

💡 do dyadic Jensen on merged event Solve&Stab$(0, \ldots, T)$:

$$\left\{\mathcal{A} \text{ solves } H^0, \ldots H^T \text{ and } \|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small for } 0 \leqslant i \leqslant T-1\right\}$$

$$\begin{aligned}\mathbb{P}(\text{Solve\&Stab}(0, \ldots, T)) &\geqslant \mathbb{P}(\text{Solve\&Stab}(0, \ldots, \tfrac{T}{2}))^2 \\ &\geqslant \mathbb{P}(\text{Solve\&Stab}(0, 1))^T\end{aligned}$$

Since $\mathbb{P}(\nexists \text{ forbidden structure}) = 1 - e^{-cN}$ & don't occur simultaneously:

$$\mathbb{P}(\text{Solve\&Stab}(0, 1)) \leqslant e^{-cN/T} \qquad \text{is small}$$

Set suitable parameters $\Rightarrow p_{\text{solve}} = o(1)$ for degree $\boldsymbol{D} = o(N)$ polynomial

Strong low degree hardness via dyadic Jensen

do dyadic Jensen on merged event Solve&Stab$(0, \dots, T)$:

$$\left\{\mathcal{A} \text{ solves } H^0, \dots H^T \text{ and } \|\mathcal{A}(H^i) - \mathcal{A}(H^{i+1})\| \text{ small for } 0 \leqslant i \leqslant T-1\right\}$$

$$\begin{aligned} \mathbb{P}(\text{Solve\&Stab}(0, \dots, T)) &\geqslant \mathbb{P}(\text{Solve\&Stab}(0, \dots, \tfrac{T}{2}))^2 \\ &\geqslant \mathbb{P}(\text{Solve\&Stab}(0, 1))^T \end{aligned}$$

Since $\mathbb{P}(\nexists$ forbidden structure$) = 1 - e^{-cN}$ & don't occur simultaneously:

$$\mathbb{P}(\text{Solve\&Stab}(0, 1)) \leqslant e^{-cN/T} \qquad \text{is small}$$

Set suitable parameters $\Rightarrow p_{\text{solve}} = o(1)$ for degree $\boldsymbol{D} = o(N)$ polynomial

(more generally, $\boldsymbol{D} = o(\log \frac{1}{p_{\text{ogp}}})$ if $\mathbb{P}(\nexists$ forbidden structure$) = 1 - p_{\text{ogp}}$)

Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\mathrm{ogp}}$, *then*

$$\mathbb{P}\Big(\textit{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\mathrm{ogp}}} / \log\log \tfrac{1}{p_{\mathrm{ogp}}}) \textit{ algorithm succeeds}\Big) = o(1)$$

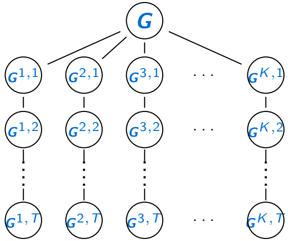
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\mathrm{ogp}}$, *then*

$$\mathbb{P}\Big(\textit{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\mathrm{ogp}}} / \log\log \tfrac{1}{p_{\mathrm{ogp}}}) \textit{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



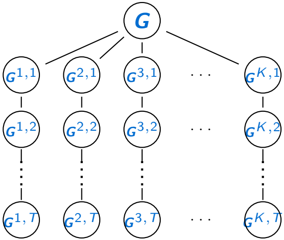
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\mathrm{ogp}}$, *then*

$$\mathbb{P}\Big(\textit{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\mathrm{ogp}}} / \log\log \tfrac{1}{p_{\mathrm{ogp}}}) \textit{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



$\text{Solve\&Stab}^{(k)} = \{$on k-th arm, $\mathcal{A}$ solves all & all steps stable$\}$

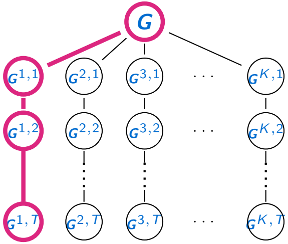
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\text{ogp}}$*, then*

$$\mathbb{P}\Big(\textit{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\text{ogp}}} / \log\log \tfrac{1}{p_{\text{ogp}}}) \textit{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



$\text{Solve\&Stab}^{(k)} = \{\text{on } k\text{-th arm},$
$\mathcal{A} \text{ solves all \& all steps stable}\}$

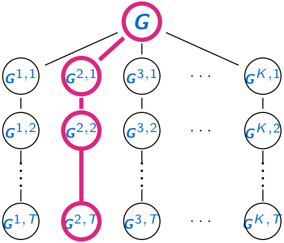
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\text{ogp}}$, *then*

$$\mathbb{P}\Big(\textit{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\text{ogp}}} / \log\log \tfrac{1}{p_{\text{ogp}}}) \textit{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



$\text{Solve\&Stab}^{(k)} = \{$on k-th arm, $\mathcal{A}$ solves all & all steps stable$\}$

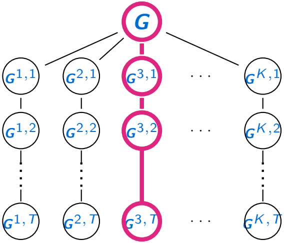
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\rm ogp}$, *then*

$$\mathbb{P}\Big(\textit{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\rm ogp}} / \log\log \tfrac{1}{p_{\rm ogp}}) \textit{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



$\text{Solve\&Stab}^{(k)} = \{$on k-th arm,
$\mathcal{A}$ solves all & all steps stable$\}$

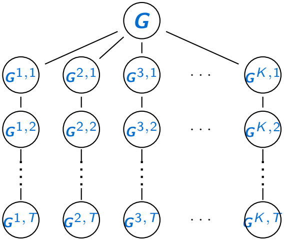
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\mathrm{ogp}}$, *then*

$$\mathbb{P}\Big(\textit{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\mathrm{ogp}}} / \log\log \tfrac{1}{p_{\mathrm{ogp}}}) \textit{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



$\mathrm{Solve\&Stab}^{(k)} = \{$on k-th arm, $\mathcal{A}$ solves all & all steps stable$\}$

As before: $\mathbb{P}(\mathrm{Solve\&Stab}^{(k)}) \geqslant \mathbb{P}(\mathrm{Solve\&Stab(one\ step)})^T$

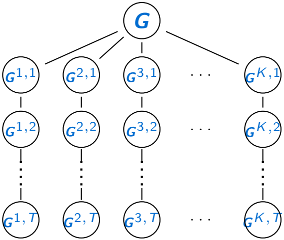
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\text{ogp}}$*, then*

$$\mathbb{P}\Big(\textit{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\text{ogp}}} / \log\log \tfrac{1}{p_{\text{ogp}}}) \textit{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



$\text{Solve\&Stab}^{(k)} = \{$on k-th arm, $\mathcal{A}$ solves all & all steps stable$\}$

As before: $\mathbb{P}(\text{Solve\&Stab}^{(k)}) \geqslant \mathbb{P}(\text{Solve\&Stab(one step)})^T$

Suppose $p_{\text{solve}} = \Omega(1)$. For K large, can show:

$$\mathbb{P}\left(\text{Solve\&Stab}^{(k)} \text{ holds for } \geqslant (Tm)^T \text{ arms}\right) = \Omega(1)$$

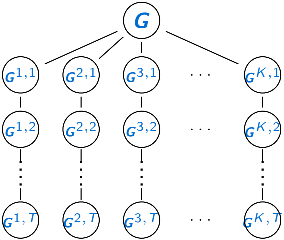
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\mathrm{ogp}}$, *then*

$$\mathbb{P}\Big(\text{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\mathrm{ogp}}} / \log\log \tfrac{1}{p_{\mathrm{ogp}}}) \text{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



$\text{Solve\&Stab}^{(k)} = \{$on k-th arm, $\mathcal{A}$ solves all & all steps stable$\}$

As before: $\mathbb{P}(\text{Solve\&Stab}^{(k)}) \geqslant \mathbb{P}(\text{Solve\&Stab(one step)})^T$

Suppose $p_{\mathrm{solve}} = \Omega(1)$. For K large, can show:

$$\mathbb{P}\,(\text{Solve\&Stab}^{(k)} \text{ holds for } \geqslant (Tm)^T \text{ arms}) = \Omega(1)$$

But on this event, Ramsey trick constructs OGP structure! Contradiction.

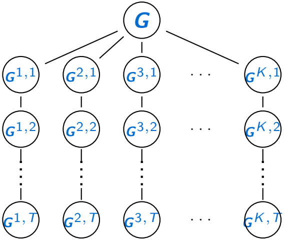
Strong low degree hardness from star OGP

Theorem (H Sellke 25)

If a star OGP holds with probability $1 - p_{\text{ogp}}$, *then*

$$\mathbb{P}\Big(\text{a degree } \boldsymbol{D} = o(\log \tfrac{1}{p_{\text{ogp}}} / \log\log \tfrac{1}{p_{\text{ogp}}}) \text{ algorithm succeeds}\Big) = o(1)$$

Construct $K \gg 1$ independent resample paths



$\text{Solve\&Stab}^{(k)} = \{$on k-th arm, $\mathcal{A}$ solves all & all steps stable$\}$

As before: $\mathbb{P}(\text{Solve\&Stab}^{(k)}) \geqslant \mathbb{P}(\text{Solve\&Stab(one step)})^T$

Suppose $p_{\text{solve}} = \Omega(1)$. For K large, can show:

$$\mathbb{P}\left(\text{Solve\&Stab}^{(k)} \text{ holds for } \geqslant \boldsymbol{(Tm)^T} \text{ arms}\right) = \Omega(1)$$

But on this event, Ramsey trick constructs OGP structure! Contradiction.

SLDH for number partitioning problem

NPP: given $g_1, \ldots, g_N \overset{IID}{\sim} \mathcal{N}(0,1)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ minimizing

$$\mathrm{discr}(\boldsymbol{\sigma}) = \Big| \sum_{i=1}^{N} g_i \sigma_i \Big|$$

SLDH for number partitioning problem

NPP: given $g_1, \dots, g_N \overset{IID}{\sim} \mathcal{N}(0,1)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ minimizing

$$\mathrm{discr}(\boldsymbol{\sigma}) = \Big| \sum_{i=1}^{N} g_i \sigma_i \Big|$$

- Best $\boldsymbol{\sigma}$ that **exists**: $\Theta(\sqrt{N} 2^{-N})$ (Karmarkar Karp Lueker Odlyzko 86)
- Best known **algorithm** finds: $2^{-\Theta(\log^2 N)}$ (Karmarkar Karp 83)

SLDH for number partitioning problem

NPP: given $g_1, \ldots, g_N \overset{IID}{\sim} \mathcal{N}(0,1)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ minimizing

$$\mathrm{discr}(\boldsymbol{\sigma}) = \Big| \sum_{i=1}^{N} g_i \sigma_i \Big|$$

- Best $\boldsymbol{\sigma}$ that **exists**: $\Theta(\sqrt{N} 2^{-N})$ (Karmarkar Karp Lueker Odlyzko 86)
- Best known **algorithm** finds: $2^{-\Theta(\log^2 N)}$ (Karmarkar Karp 83)

- Stable algorithms cannot reach $2^{-\Theta(N)}$ (Gamarnik Kızıldağ 21)
- Algorithms cannot beat $2^{-\Theta(\log^3 N)}$, assuming **worst case** hardness of approx shortest vector in lattices (Vafa Vaikuntanathan 25)

SLDH for number partitioning problem

NPP: given $g_1, \dots, g_N \overset{IID}{\sim} \mathcal{N}(0,1)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ minimizing

$$\mathrm{discr}(\boldsymbol{\sigma}) = \Big| \sum_{i=1}^{N} g_i \sigma_i \Big|$$

- Best $\boldsymbol{\sigma}$ that **exists**: $\Theta(\sqrt{N} 2^{-N})$ (Karmarkar Karp Lueker Odlyzko 86)
- Best known **algorithm** finds: $2^{-\Theta(\log^2 N)}$ (Karmarkar Karp 83)

- Stable algorithms cannot reach $2^{-\Theta(N)}$ (Gamarnik Kızıldağ 21)
- Algorithms cannot beat $2^{-\Theta(\log^3 N)}$, assuming **worst case** hardness of approx shortest vector in lattices (Vafa Vaikuntanathan 25)

Theorem (Mallarapu Sellke 25)

For any $1 \ll \boldsymbol{D} \ll N$, $\mathbb{P}$(*a degree* $\boldsymbol{D}$ *alg beats* $2^{-\tilde{\Omega}(\boldsymbol{D})}) = o(1)$.

SLDH for number partitioning problem

NPP: given $g_1, \ldots, g_N \overset{IID}{\sim} \mathcal{N}(0,1)$, find $\boldsymbol{\sigma} \in \{\pm 1\}^N$ minimizing

$$\mathrm{discr}(\boldsymbol{\sigma}) = \left| \sum_{i=1}^{N} g_i \sigma_i \right|$$

- Best $\boldsymbol{\sigma}$ that **exists**: $\Theta(\sqrt{N} 2^{-N})$ (Karmarkar Karp Lueker Odlyzko 86)
- Best known **algorithm** finds: $2^{-\Theta(\log^2 N)}$ (Karmarkar Karp 83)

- Stable algorithms cannot reach $2^{-\Theta(N)}$ (Gamarnik Kızıldağ 21)
- Algorithms cannot beat $2^{-\Theta(\log^3 N)}$, assuming **worst case** hardness of approx shortest vector in lattices (Vafa Vaikuntanathan 25)

Theorem (Mallarapu Sellke 25)

For any $1 \ll \boldsymbol{D} \ll N$, $\mathbb{P}($*a degree* $\boldsymbol{D}$ *alg beats* $2^{-\tilde{\Omega}(\boldsymbol{D})}) = o(1)$.

This is sharp **for all** $1 \ll \boldsymbol{D} \ll N$: deg $\boldsymbol{D}$ achieves $2^{-\tilde{\Omega}(\boldsymbol{D})}$ by brute force.

Relation to shortest path OGP

Li Schramm 24: shortest path on $G(N, \frac{C \log N}{N})$ satisfies OGP but is easy
$\Rightarrow$ **When does OGP actually imply hardness?**

Relation to shortest path OGP

Li Schramm 24: shortest path on $G(N, \frac{C \log N}{N})$ satisfies OGP but is easy
$\Rightarrow$ **When does OGP actually imply hardness?**

H Sellke 25: OGP with prob $1 - p_{\text{ogp}} \Rightarrow$ hard for deg $\boldsymbol{D} \ll \log \frac{1}{p_{\text{ogp}}}$
(and $\boldsymbol{D} \asymp \log \frac{1}{p_{\text{ogp}}}$ can brute force in many problems)

Relation to shortest path OGP

Li Schramm 24: shortest path on $G(N, \frac{C \log N}{N})$ satisfies OGP but is easy
$\Rightarrow$ **When does OGP actually imply hardness?**

H Sellke 25: OGP with prob $1 - p_{\text{ogp}} \Rightarrow$ hard for deg $\boldsymbol{D} \ll \log \frac{1}{p_{\text{ogp}}}$
(and $\boldsymbol{D} \asymp \log \frac{1}{p_{\text{ogp}}}$ can brute force in many problems)

Our perspective: probability of OGP could matter

- shortest path OGP holds with $p_{\text{ogp}} \asymp \frac{\log \log N}{\log N} \Rightarrow \boldsymbol{D} \asymp \log \log N$

Relation to shortest path OGP

Li Schramm 24: shortest path on $G(N, \frac{C \log N}{N})$ satisfies OGP but is easy
$\Rightarrow$ **When does OGP actually imply hardness?**

H Sellke 25: OGP with prob $1 - p_{\text{ogp}} \Rightarrow$ hard for deg $\boldsymbol{D} \ll \log \frac{1}{p_{\text{ogp}}}$
(and $\boldsymbol{D} \asymp \log \frac{1}{p_{\text{ogp}}}$ can brute force in many problems)

Our perspective: probability of OGP could matter

- shortest path OGP holds with $p_{\text{ogp}} \asymp \frac{\log \log N}{\log N} \Rightarrow \boldsymbol{D} \asymp \log \log N$
- heuristically corresponds to runtime $e^{\boldsymbol{D}} \ll \text{poly}(N)$

Relation to shortest path OGP

Li Schramm 24: shortest path on $G(N, \frac{C \log N}{N})$ satisfies OGP but is easy
$\Rightarrow$ **When does OGP actually imply hardness?**

H Sellke 25: OGP with prob $1 - p_{\text{ogp}} \Rightarrow$ hard for deg $\boldsymbol{D} \ll \log \frac{1}{p_{\text{ogp}}}$
(and $\boldsymbol{D} \asymp \log \frac{1}{p_{\text{ogp}}}$ can brute force in many problems)

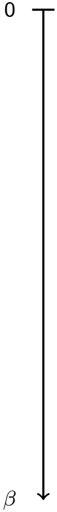
Our perspective: probability of OGP could matter

- shortest path OGP holds with $p_{\text{ogp}} \asymp \frac{\log \log N}{\log N} \Rightarrow \boldsymbol{D} \asymp \log \log N$
- heuristically corresponds to runtime $e^{\boldsymbol{D}} \ll \text{poly}(N)$

Possible reconciliation: $p_{\text{ogp}} = N^{-\omega(1)}$ necessary for "genuine" hardness

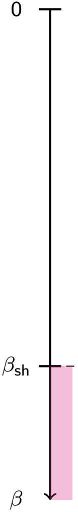
Phase diagram for sampling

For sampling from spin glass Gibbs measure $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$:



Phase diagram for sampling

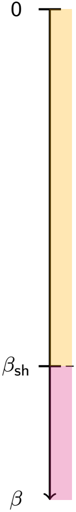
For sampling from spin glass Gibbs measure $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$:



Gibbs measure shatters; sampling predicted hard.
Rigorous hardness for **stable** samplers.
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23,
Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Phase diagram for sampling

For sampling from spin glass Gibbs measure $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$:



Prediction: Langevin succeeds from random init

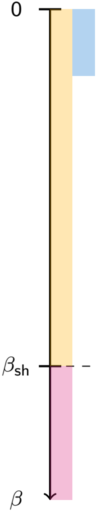
Gibbs measure shatters; sampling predicted hard.

Rigorous hardness for **stable** samplers.

(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

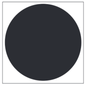
Phase diagram for sampling

For sampling from spin glass Gibbs measure $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$:



Langevin mixes rapidly from worst-case init

(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



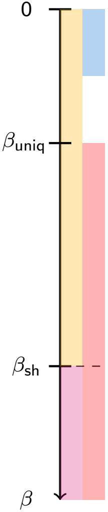
Prediction: Langevin succeeds from random init

Gibbs measure shatters; sampling predicted hard.
Rigorous hardness for **stable** samplers.

(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

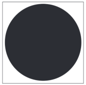
Phase diagram for sampling

For sampling from spin glass Gibbs measure $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$:



Langevin mixes rapidly from worst-case init

(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



Langevin mixes slowly from worst-case init

(Ben Arous Jagannath 24)

Prediction: Langevin succeeds from random init

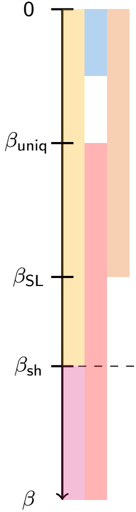
Gibbs measure shatters; sampling predicted hard.

Rigorous hardness for **stable** samplers.

(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

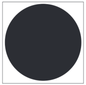
Phase diagram for sampling

For sampling from spin glass Gibbs measure $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$:



Langevin mixes rapidly from worst-case init

(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



Algorithmic stochastic localization (↔ diffusion) succeeds

(El Alaoui Montanari Sellke 22+23, H Montanari Pham 24)

Langevin mixes slowly from worst-case init

(Ben Arous Jagannath 24)

Prediction: Langevin succeeds from random init

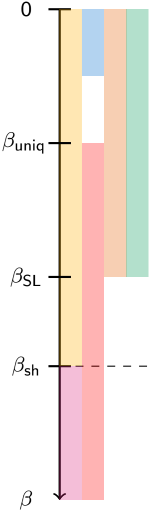
Gibbs measure shatters; sampling predicted hard.

Rigorous hardness for **stable** samplers.

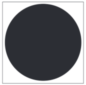
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Phase diagram for sampling

For sampling from spin glass Gibbs measure $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$:



Langevin mixes rapidly from worst-case init
(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



Algorithmic stochastic localization ($\leftrightarrow$ diffusion) succeeds
(El Alaoui Montanari Sellke 22+23, H Montanari Pham 24)

Simulated annealing succeeds
(H Mohanty Rajaraman Wu 24)

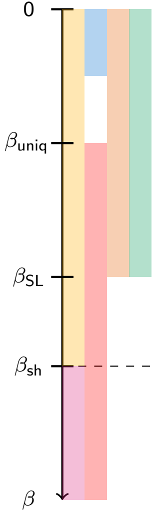
Langevin mixes slowly from worst-case init
(Ben Arous Jagannath 24)

Prediction: Langevin succeeds from random init

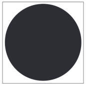
Gibbs measure shatters; sampling predicted hard.
Rigorous hardness for **stable** samplers.
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Phase diagram for sampling

For sampling from spin glass Gibbs measure $\mu_\beta(\boldsymbol{\sigma}) \propto e^{\beta H(\boldsymbol{\sigma})}$:



Langevin mixes rapidly from worst-case init
(Bauerschmidt Bodineau 17, Gheissari Jagannath 19, Eldan Koehler Zeitouni 22, Anari Jain Pham Koehler Vuong 24, AKV 24)



Algorithmic stochastic localization ($\leftrightarrow$ diffusion) succeeds
(El Alaoui Montanari Sellke 22+23, H Montanari Pham 24)

Simulated annealing succeeds
(H Mohanty Rajaraman Wu 24)

Langevin mixes slowly from worst-case init
(Ben Arous Jagannath 24)

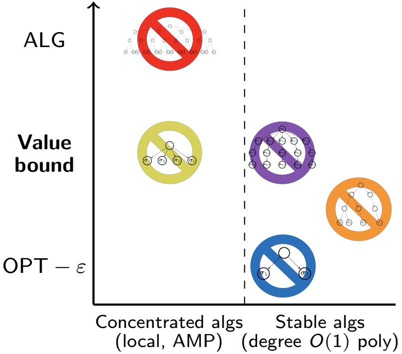
Prediction: Langevin succeeds from random init

Gibbs measure shatters; sampling predicted hard.
Rigorous hardness for **stable** samplers.
(Crisanti Horner Sommers 93, El Alaoui Montanari Sellke 23, Gamarnik Jagannath Kızıldağ 23, El Alaoui 24)

Open problem: sample for $\beta \in (\beta_{\mathsf{SL}}, \beta_{\mathsf{sh}})$

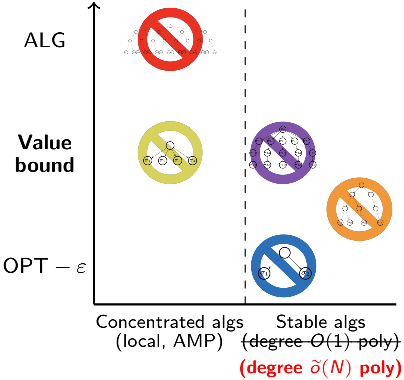
Conclusion

OGP is a powerful geometric framework for computational limits in random search / optimization problems.



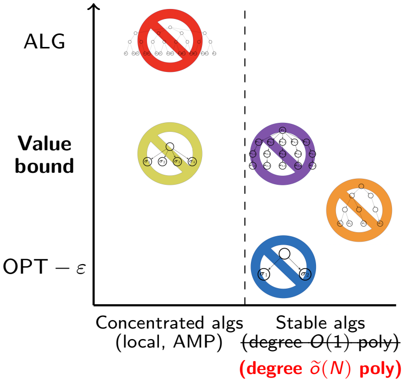
Conclusion

OGP is a powerful geometric framework for computational limits in random search / optimization problems.



Conclusion

OGP is a powerful geometric framework for computational limits in random search / optimization problems.

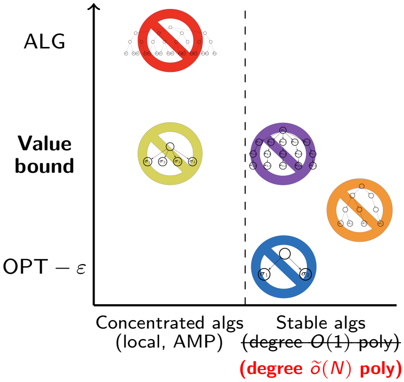


Outstanding challenges:

- strong low degree hardness for branching OGP
- long-time analysis of Glauber / Langevin dynamics
- hardness of finding isolated solutions
- quantum systems (see Anschuetz Gamarnik Kiani 24)

Conclusion

OGP is a powerful geometric framework for computational limits in random search / optimization problems.

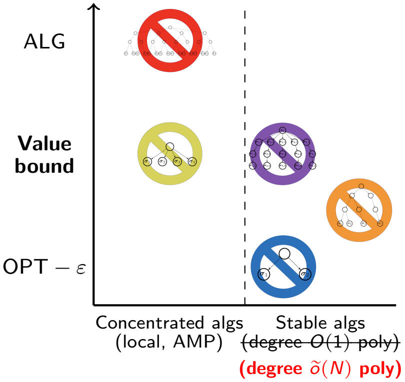


Outstanding challenges:

- strong low degree hardness for branching OGP
- long-time analysis of Glauber / Langevin dynamics
- hardness of finding isolated solutions
- quantum systems (see Anschuetz Gamarnik Kiani 24)
- hardness at ALG (even for conc algs) in random CSPs
- sample up to β_{sh}

Conclusion

OGP is a powerful geometric framework for computational limits in random search / optimization problems.



Outstanding challenges:

- strong low degree hardness for branching OGP
- long-time analysis of Glauber / Langevin dynamics
- hardness of finding isolated solutions
- quantum systems (see Anschuetz Gamarnik Kiani 24)
- hardness at ALG (even for conc algs) in random CSPs
- sample up to β_{sh}

Thank you!