Unifying Statistical Physics and Low-Degree Polynomials?

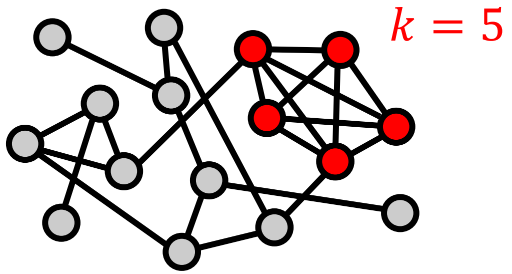
Alex Wein

University of California, Davis

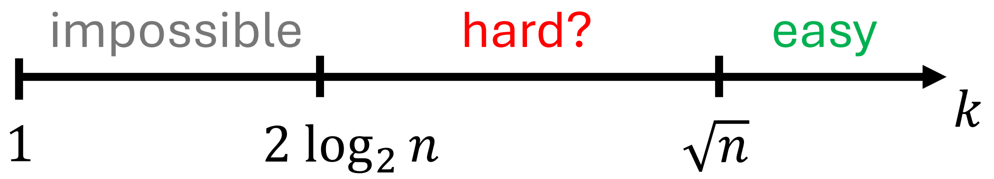
“High-Dimensional Statistics”

- Spiked Wigner model: “rank-1 matrix plus noise”
  - $\mathbb{P}$ (“planted”): $Y = \lambda u u^\top + W$
  - $\mathbb{Q}$ (“null”): $Y = W$
  - “Noise” W is symmetric, i.i.d. $N(0,1)$
  - Planted signal $u \in \mathbb{R}^n$ drawn from some prior

- Planted clique problem
  - $\mathbb{P}$: $G(n, 1/2) + \{k\text{-clique}\}$
  - $\mathbb{Q}$: $G(n, 1/2)$



- Statistical-computational gaps…



Heuristics for Hardness

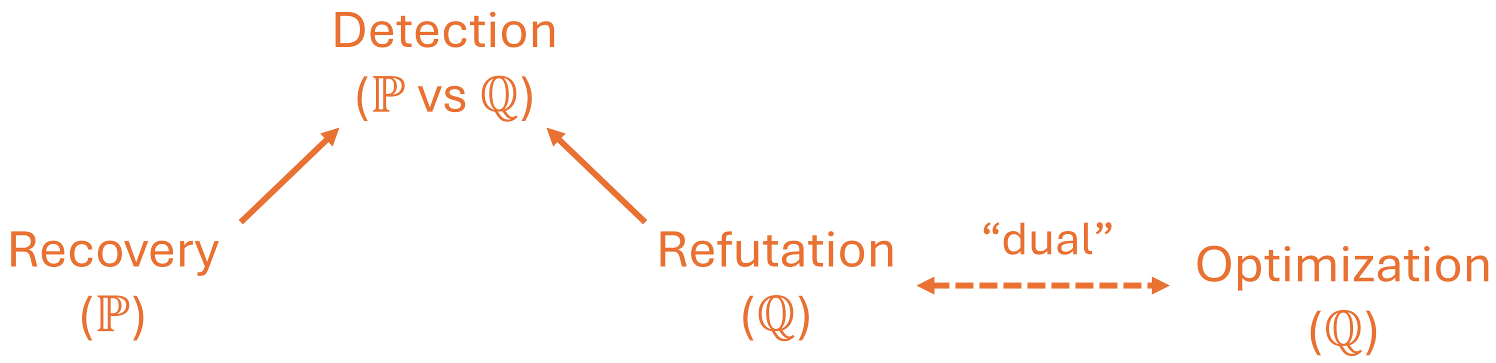
- AMP (approximate message passing + replica symmetric potential)
- OGP (overlap gap property)
- SOS (sum-of-squares hierarchy)
- SQ (statistical query model)
- LD (low-degree polynomials)

- "Unify" these: Can we prove they all make the same predictions?

- Two issues:
  1. Sometimes they are NOT equivalent...
  2. Often they are not even answering the same question...

Tasks

- Using planted clique as a running example…
- **Detection (a.k.a. Testing)**: Decide if a given graph came from $\mathbb{P}$ or $\mathbb{Q}$
- **Recovery (a.k.a. Estimation)**: Given $G \sim \mathbb{P}$, find the planted clique
- **(Non-planted) optimization**: Given $G \sim \mathbb{Q}$, find a k-clique
- **Refutation (or Certification)**: Given $G \sim \mathbb{Q}$, prove there's no k-clique



- "Gaps": trivial detection via total edge count, but recovery is harder

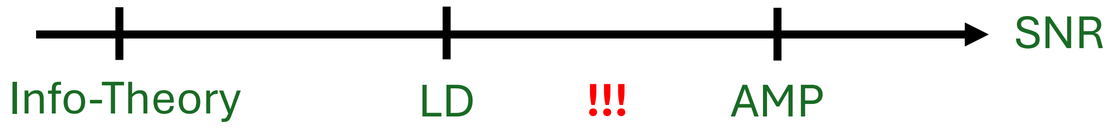
Frameworks vs Tasks

- Which frameworks can give **hardness results** for which tasks?

	AMP	OGP	SOS	SQ	LD
Detection			✓	✓	✓
Recovery	✓	✓		✓	✓
Optimization	✓	✓			✓
Refutation			✓		✓

Tensor PCA

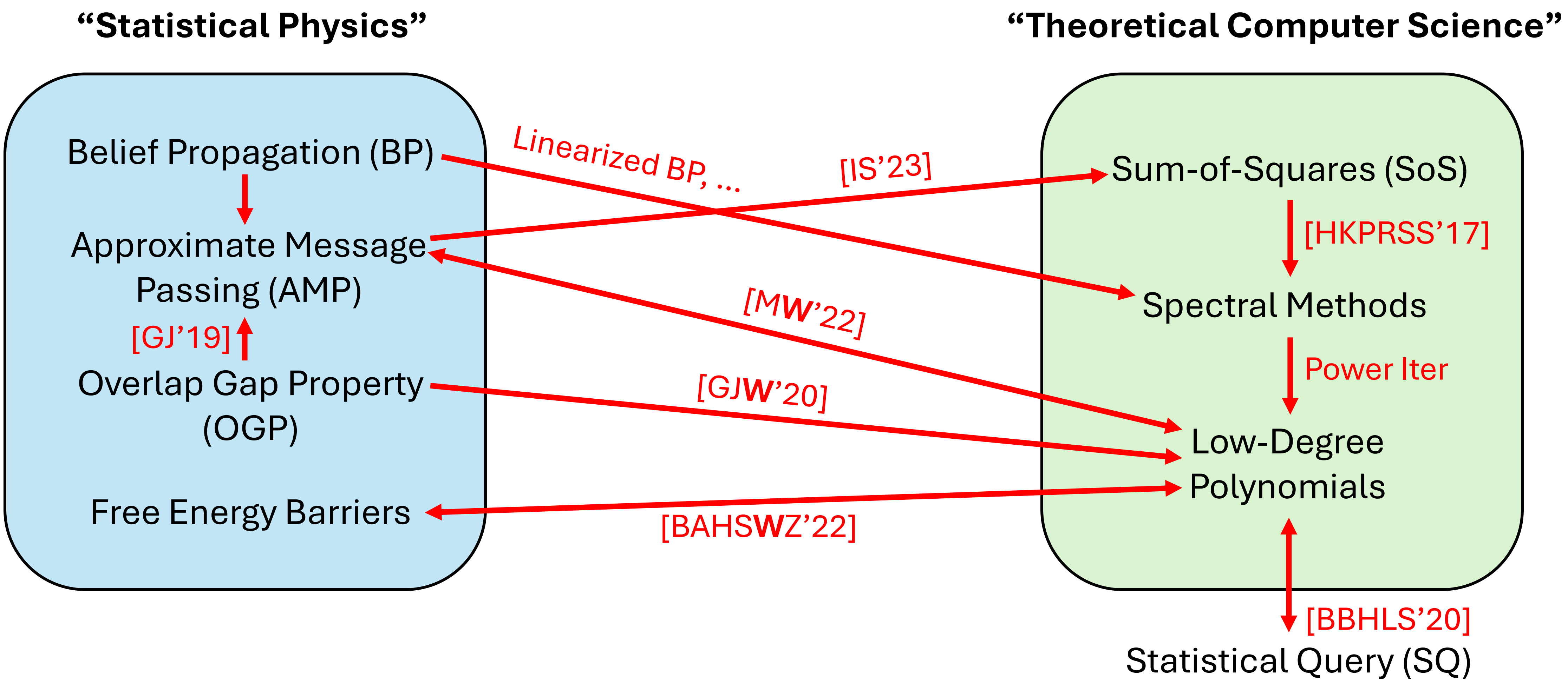
- AMP (and other "local search" algorithms) get the "wrong" threshold!
- LD (and SOS, spectral methods) get the "correct" threshold



- "Redemption"
  - Kikuchi hierarchy (in place of Bethe free energy) [W,Alaoui,Moore'19]
  - Averaged gradient descent [Biroli,Cammarota,Ricci-Tersenghi'19]
  - ... but these are somewhat problem specific (?)

Known Connections

- Despite **many** caveats, some known connections among frameworks



This Talk

- Two stories about connecting physics with low-degree polynomials

- Part 1: “Annealed Franz-Parisi Potential”
  - Connecting low-degree **detection** with a “physics-style” object

- Part 2: AMP vs Low-Degree Estimation
  - Can low-degree polynomials recover the sharp **estimation** predictions made by physics?

Part 1. “Annealed Franz-Parisi Potential”

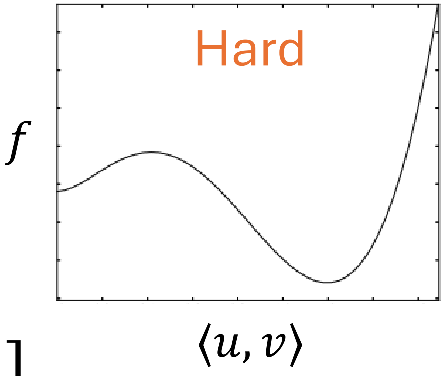
Joint with: Afonso Bandeira, Ahmed El Alaoui, Sam Hopkins,
Tselil Schramm, Ilias Zadik

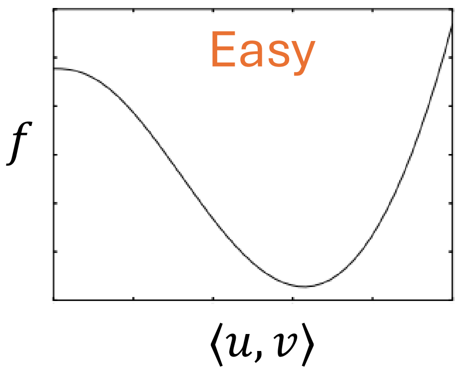
“The Franz-Parisi Criterion and Computational Trade-offs in High Dimensional Statistics”

Testing $Y \sim \mathbb{P}$ vs $Y \sim \mathbb{Q}$

- Optimal test statistic is likelihood ratio $L(Y) = (\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q})(Y)$
- If $\mathbb{P}$ involves planted signal u then $L = \mathbb{E}_u L_u$ where $L_\mathrm{u} = \mathrm{d}\mathbb{P}_u/\mathrm{d}\mathbb{Q}$
  - Physics: This is a *partition function*
- $\chi^2(\mathbb{P}\|\mathbb{Q}) + 1 = \mathbb{E}_{Y\sim\mathbb{P}}[L(Y)] = \mathbb{E}_\mathbb{Q}[L^2] = \mathbb{E}_{u,v}\mathbb{E}_\mathbb{Q}[L_u L_v]$
  - Physics: *Annealed free energy*
  - Statistics: If $O(1)$ then impossible to distinguish w.h.p.; if $\omega(1)$ then ???
- $\chi^2_{\le D}(\mathbb{P}\|\mathbb{Q}) + 1 = \mathbb{E}_{Y\sim\mathbb{P}}[L^{\le D}(Y)] = \mathbb{E}_\mathbb{Q}[(L^{\le D})^2] = \mathbb{E}_{u,v}\mathbb{E}_\mathbb{Q}[L_u^{\le D} L_v^{\le D}]$
  - Think $D = (\log n)^{1+\epsilon}$
  - If $O(1)$ then “low-degree hard” to distinguish; if $\omega(1)$ then ???

Two Different "Truncations"





$$\chi^2(\mathbb{P}\|\mathbb{Q}) + 1 = \mathbb{E}_{u,v}\mathbb{E}_{\mathbb{Q}}[L_u L_v]$$

Low degree ↙ ↘ Low overlap

$$\mathrm{LD}(D) = \mathbb{E}_{u,v}\mathbb{E}_{\mathbb{Q}}[L_u^{\leq D} L_v^{\leq D}] \quad \approx \quad \mathrm{FP}(\delta) = \mathbb{E}_{u,v}\mathbb{1}_{|\langle u,v\rangle| \leq \delta}\mathbb{E}_{\mathbb{Q}}[L_u L_v]$$

"Annealed Franz-Parisi Potential"

- **Thm**: Two truncations are "equivalent" for various models
  - General additive Gaussian models, "planted sparse models"
- Relation D vs δ: $\Pr(|\langle u, v\rangle| \geq \delta) \approx e^{-D}$
- If $\mathrm{FP}(\delta) = O(1)$ then "FP-hard", implies "LD-hard"; if $\omega(1)$ then ???

Summary: FP vs LD

- Connects “algebraic” hardness with “landscape” hardness
- But some caveats:
  - FP is not directly related to AMP or RS potential
  - FP does not say anything precise about recovery error or MSE
  - In fact, FP is more tied to detection rather than recovery
- FP makes the “right” prediction for tensor PCA, while AMP fails
- FP is also a useful tool for proving low-degree lower bounds

Part 2. AMP vs Low-Degree Estimation

Joint with: Andrea Montanari

"Equivalence of AMP and Low-Degree Polynomials in Rank-One Matrix Estimation"

Another Approach...

- Let's meet AMP on its home turf: **estimation**
- Focus on a problem (spiked Wigner) where we expect AMP is optimal
- RS potential makes precise predictions about MSE
- For a start, can we recover these using low-degree polynomials?

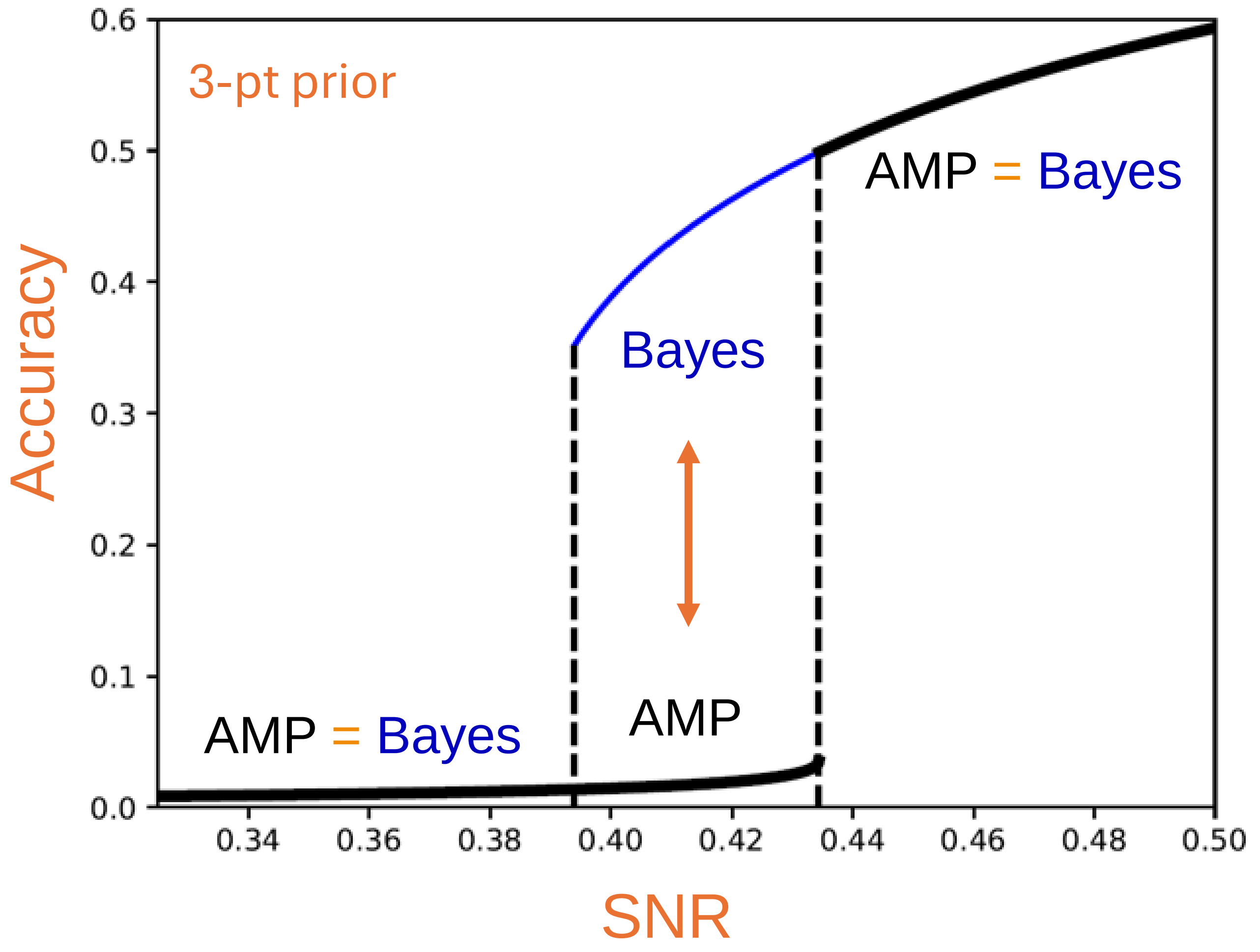
- Spiked Wigner: $Y = \lambda u u^\top + W$ with u i.i.d. from some (fixed) prior
- Estimator $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ multivariate polynomial of deg $\leq D$
- Degree-D MMSE:

$$\mathrm{MMSE}_{\leq D} := \inf_{\deg(f) \leq D} \mathbb{E}[(f(Y) - u_1)^2]$$

AMP for (Rank-1) Spiked Wigner

- Signal $u_i \sim \pi$ where $\mathbb{E}[\pi] \neq 0$
- AMP with optimal denoiser, t iter
- $n \to \infty$ followed by $t \to \infty$

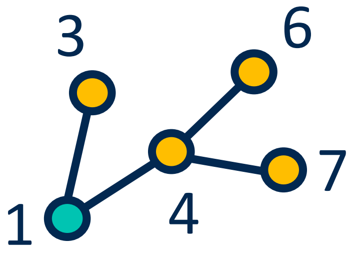
- **Conj**: AMP has best MSE among poly-time algorithms
- **Thm**: AMP has best MSE among degree-$O(1)$ polynomials
- **Conj**: AMP has best MSE among degree-$n^{1-o(1)}$ polynomials



Why AMP ≈ LD?

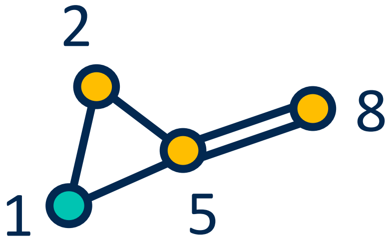
- AMP is equivalent to “tree-structured” polynomials

tree

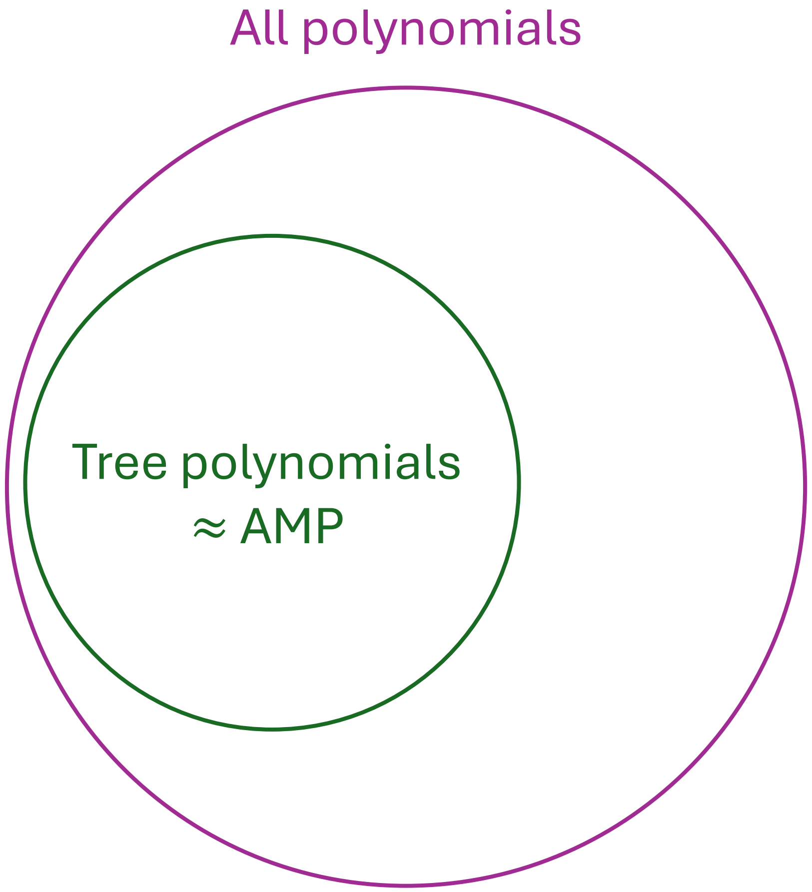


$$f(Y) = Y_{13}Y_{14}Y_{46}Y_{47}$$

non-tree



$$g(Y) = Y_{12}Y_{15}Y_{25}Y_{58}^2$$



- In spiked Wigner, tree polynomials are optimal among all polynomials

Higher Degree

- Ideally we should rule out polynomials of higher degree, say $n^{\Omega(1)}$

Forthcoming work with Byron Chin, Elchanan Mossel, Youngtak Sohn...

- LD matches some sharp phase transitions predicted by physics
  - But we don't (yet) match the exact MSE...
- Spiked Wigner: LD estimation fails below BBP transition $\lambda = 1$
  - Even for any sub-extensive rank $m \ll n$
- Stochastic Block Model: LD estimation fails below KS threshold
  - Even for growing number of communities

Concluding Thoughts

- We know AMP makes extremely sharp predictions
  - But not applicable to all settings (tensor PCA, extensive rank, …)
- LD gives a stronger form of hardness (all polyn’s vs tree-polyn’s)
  - But current results are less sharp than AMP
- Ongoing challenge: Sharpen LD lower bounds to match AMP
  - And in the process, understand when AMP is optimal (and when it’s not)

Thanks!