Optimality of AMP Among Low-Degree Polynomials

Alex Wein

University of California, Davis

Department of Mathematics

Joint work with Andrea Montanari (Stanford)

arXiv: Equivalence of Approximate Message Passing and Low-Degree Polynomials in Rank-One Matrix Estimation

UCDAVIS

High-Dimensional Statistical Inference

- Find hidden structure in random graph

  E.g. planted clique in G(n,1/2), stochastic block model, graph matching

- Find low-dimensional structure in random data

  E.g. spiked matrix models, matrix factorization, tensor decomposition

- Regression / linear models

  E.g. compressed sensing / sparse regression, phase retrieval

Common features: large input, many unknowns, planted signal

Spiked Wigner Model

signal-to-noise ratio $s > 0$

$$Y = \frac{s}{\sqrt{n}} \theta\theta^\top + Z$$

observed data, n-by-n matrix

rank-1 “signal”

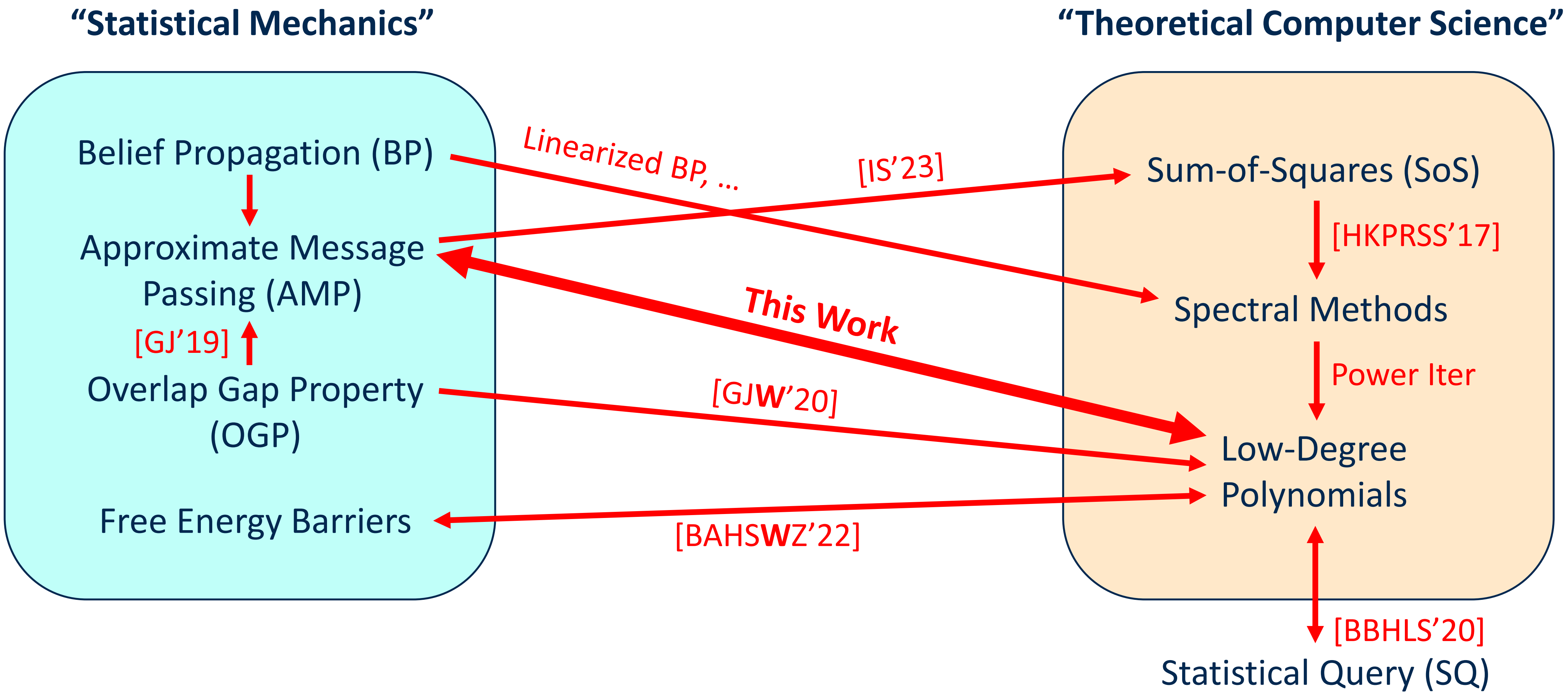
iid Gaussian “noise”

θ – unknown vector with entries iid from known fixed prior π

Goal: given Y, estimate θ

Simple “signal plus noise” model, testbed

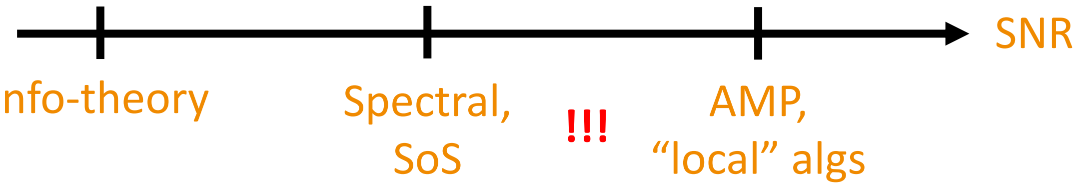
What Are The Best Algorithms?



A Unified Theory?

Many connections, but also caveats and counterexamples…

- Detection vs recovery vs optimization vs refutation vs sampling
- Physics predictions are “wrong” for tensor PCA (!)
  [Montanari,Richard’14; Ben Arous,Gheissari,Jagannath’18]



“Redemption”

- Kikuchi hierarchy (in place of Bethe free energy) [**W**,Alaoui,Moore’19]
- Averaged gradient descent [Biroli,Cammarota,Ricci-Tersenghi’19]

AMP for Spiked Wigner Model

$$Y = \frac{s}{\sqrt{n}}\theta\theta^\top + Z$$

Iterate

$$x^0 = 0$$

vector, estimate for θ

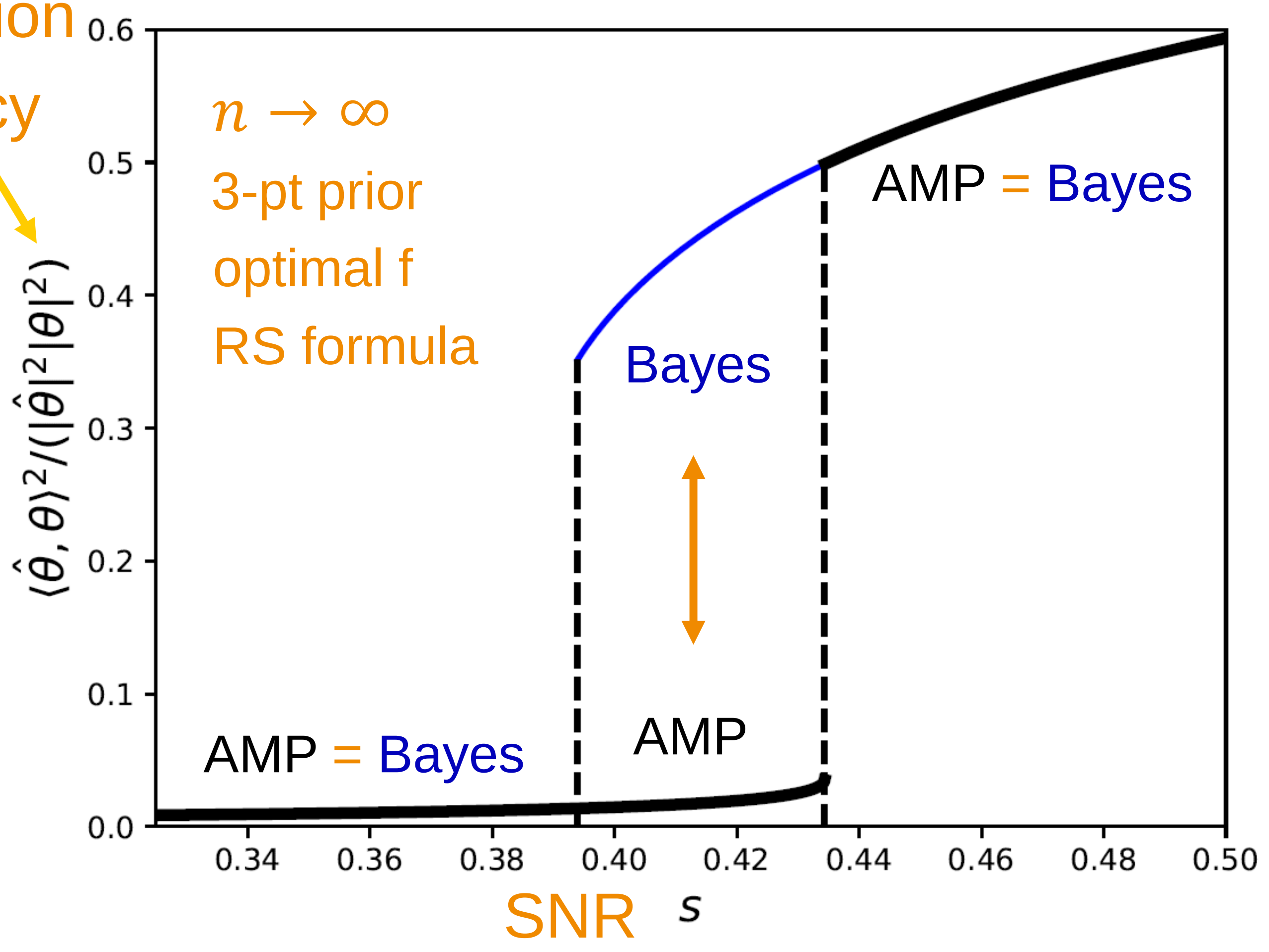
$$x^t = \frac{1}{\sqrt{n}} Y f(x^{t-1}) - \mathrm{b}_t f(x^{t-2})$$

entrywise transform

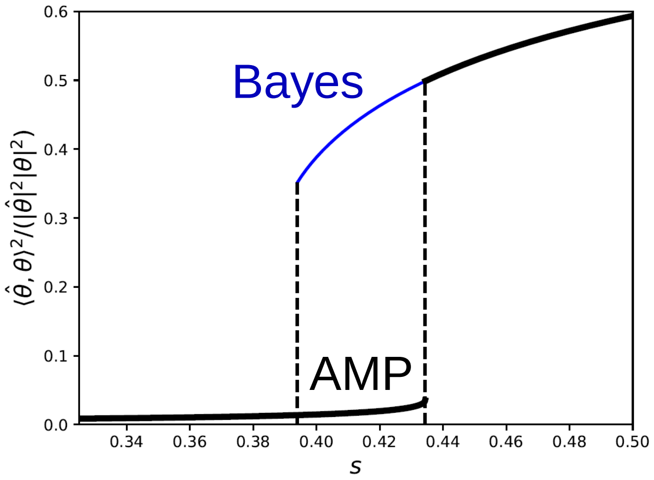
Onsager term

Bayes is not comp. efficient, gap

estimation accuracy



Main Result



Conjecture *(… Lesieur,Krzakala,Zdeborová’15 …)*

AMP has optimal MSE among all poly-time algorithms

Theorem *(Montanari, W ‘22)*

AMP has optimal MSE among all constant-degree polynomials

AMP (with const num iter) takes the form $(\hat{\theta}_1(Y), \dots, \hat{\theta}_n(Y))$ where $\hat{\theta}_i$ is a const-deg multivariate polynomial in the entries of Y

We show AMP is the best estimator of this form; sharp constant

Comments

Biased prior: $\mathbb{E}[\pi] \neq 0$

Open: mean-zero prior π, $O(\log n)$ iterations/degree

Open: rule out higher degree polynomials

Conjecture: need degree $n^{1-o(1)}$ to beat AMP

AMP is sub-optimal for some problems (tensor PCA, …)

Proof suggests how to test if AMP is optimal for a given problem

Low-Degree Estimation Lower Bounds

Given Y, estimate θ_1

Want to understand $\mathrm{MMSE}_{\leq D} := \inf_{p \deg D} \mathbb{E}[(p(Y) - \theta_1)^2]$

- Planted submatrix, planted dense subgraph [Schramm,**W**'20]
- Hypergraphic planted dense subgraph [Luo,Zhang'20]
- Tensor decomposition [**W**'22]

This work: exact value of $\lim_{D\to\infty} \lim_{n\to\infty} \mathrm{MMSE}_{\leq D}$

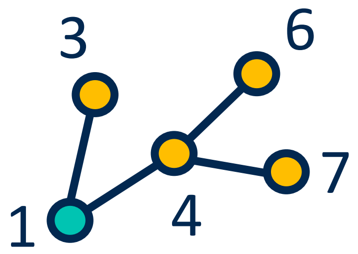
$$Y = \frac{s}{\sqrt{n}} \theta \theta^\top + Z$$

Proof Sketch: AMP vs Low-Deg

I. AMP is as powerful as any “tree-shaped” polynomial

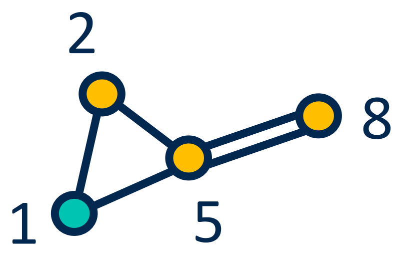
II. Tree-shaped polynomials are as powerful as all polynomials (of the same degree)

tree



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

non-tree



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

$$Y = \frac{s}{\sqrt{n}}\theta\theta^\top + Z$$

I. AMP vs Tree Polynomials

Claim: $\lim_{t\to\infty}\lim_{n\to\infty} \mathrm{MSE}_t^{\mathrm{AMP}} = \lim_{D\to\infty}\lim_{\mathrm{n}\to\infty} \mathrm{MMSE}_{\le D}^{\mathrm{Tree}}$

($\geq$) AMP is a tree polynomial

($\leq$) Consider the best tree polynomial, WLOG symmetric

Given any symmetric const-deg tree polynomial, can construct a "message-passing" (MP) scheme to compute it

Prior work: AMP has best MSE among all MP schemes

[Celentano,Montanari,Wu'20; Montanari,Wu'22]

$$Y = \frac{s}{\sqrt{n}}\theta\theta^\top + Z$$

II. Tree Poly vs All Poly

Remains to prove: $\lim_{n\to\infty} \mathrm{MMSE}_{\leq D}^{\mathrm{Tree}} = \lim_{n\to\infty} \mathrm{MMSE}_{\leq D}$ (rest of talk)

Conclude:

$$\lim_{t\to\infty}\lim_{n\to\infty} \mathrm{MSE}_t^{\mathrm{AMP}} = \lim_{D\to\infty}\lim_{n\to\infty} \mathrm{MMSE}_{\leq D}^{\mathrm{Tree}} = \lim_{D\to\infty}\lim_{n\to\infty} \mathrm{MMSE}_{\leq D}$$

AMP | Tree Poly | All Poly

$$Y = \frac{s}{\sqrt{n}} \theta\theta^\top + Z$$

II. Tree Poly vs All Poly

Remains to prove: $\lim_{n\to\infty} \mathrm{MMSE}_{\le D}^{\mathrm{Tree}} = \lim_{n\to\infty} \mathrm{MMSE}_{\le D}$

$$\mathrm{MMSE}_{\le D} := \inf_{p \deg D} \mathbb{E}[(p(Y) - \theta_1)^2] = \mathbb{E}[\theta_1^2] - c^\top M^{-1} c$$

where:

$\{H_A\}$ – basis for (symmetric) const-deg polynomials

$c_A := \mathbb{E}[H_A(Y) \cdot \theta_1]$ $\qquad$ $M_{AB} := \mathbb{E}[H_A(Y) \cdot H_B(Y)]$

Goal: $\lim_{n\to\infty} \mathrm{MMSE}_{\leq D}^{\mathrm{Tree}} = \lim_{n\to\infty} \mathrm{MMSE}_{\leq D}$

$$\mathrm{MMSE}_{\leq D} = \mathbb{E}[\theta_1^2] - c^\top M^{-1} c$$

$$c_A := \mathbb{E}[H_A(Y) \cdot \theta_1] \qquad M_{AB} := \mathbb{E}[H_A(Y) \cdot H_B(Y)]$$

tree non-tree

$$M = \begin{array}{c} \text{tree} \\ \text{non-tree} \end{array} \begin{bmatrix} P & R \\ R^\top & Q \end{bmatrix} = \begin{bmatrix} \Theta(1) & o(1) \\ o(1) & \Theta(1) \end{bmatrix} \qquad c = \begin{bmatrix} d \\ e \end{bmatrix} \begin{array}{c} \Theta(1) \\ o(1) \end{array}$$

$$\mathbb{E}[\theta_1^2] - \mathrm{MMSE}_{\leq D} = c^\top M^{-1} c \approx d^\top P^{-1} d^\top = \mathbb{E}[\theta_1^2] - \mathrm{MMSE}_{\leq D}^{\mathrm{Tree}}$$

Summary

Equivalence of constant-iter AMP and constant-degree polynomials in the spiked Wigner model with any fixed prior

AMP = tree polynomials = all polynomials

Key property of Wigner model for "tree = all": block diagonal

Use this to test if AMP is optimal for a given problem?

Evidence for AMP conjecture; connection stat mech $\leftrightarrow$ TCS

Thanks!