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

or: Optimality of AMP

Alex Wein

University of California, Davis

Department of Mathematics

Joint work with Andrea Montanari (Stanford)

UCDAVIS

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”

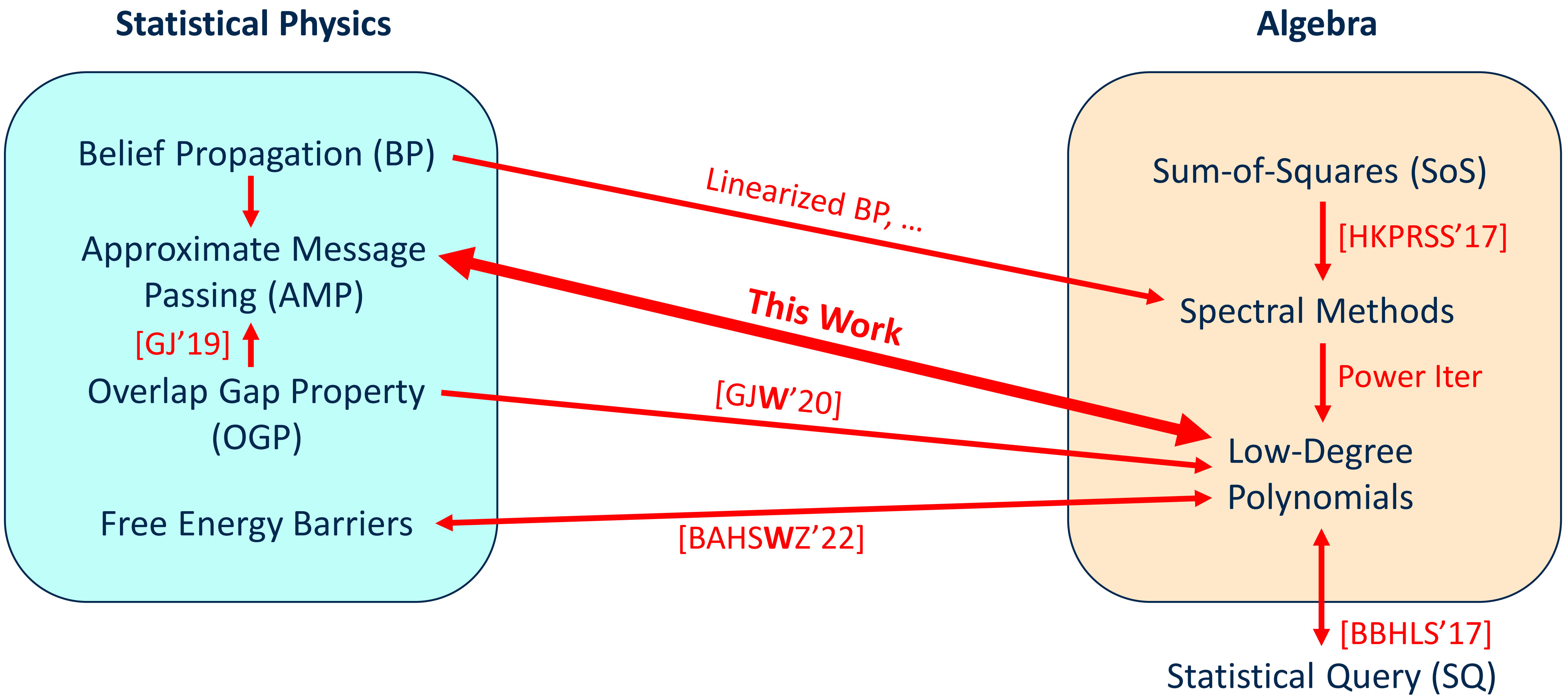
iid Gaussian “noise”

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?



AMP for Spiked Wigner Model

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

Iterate

$$x^0 = 0$$

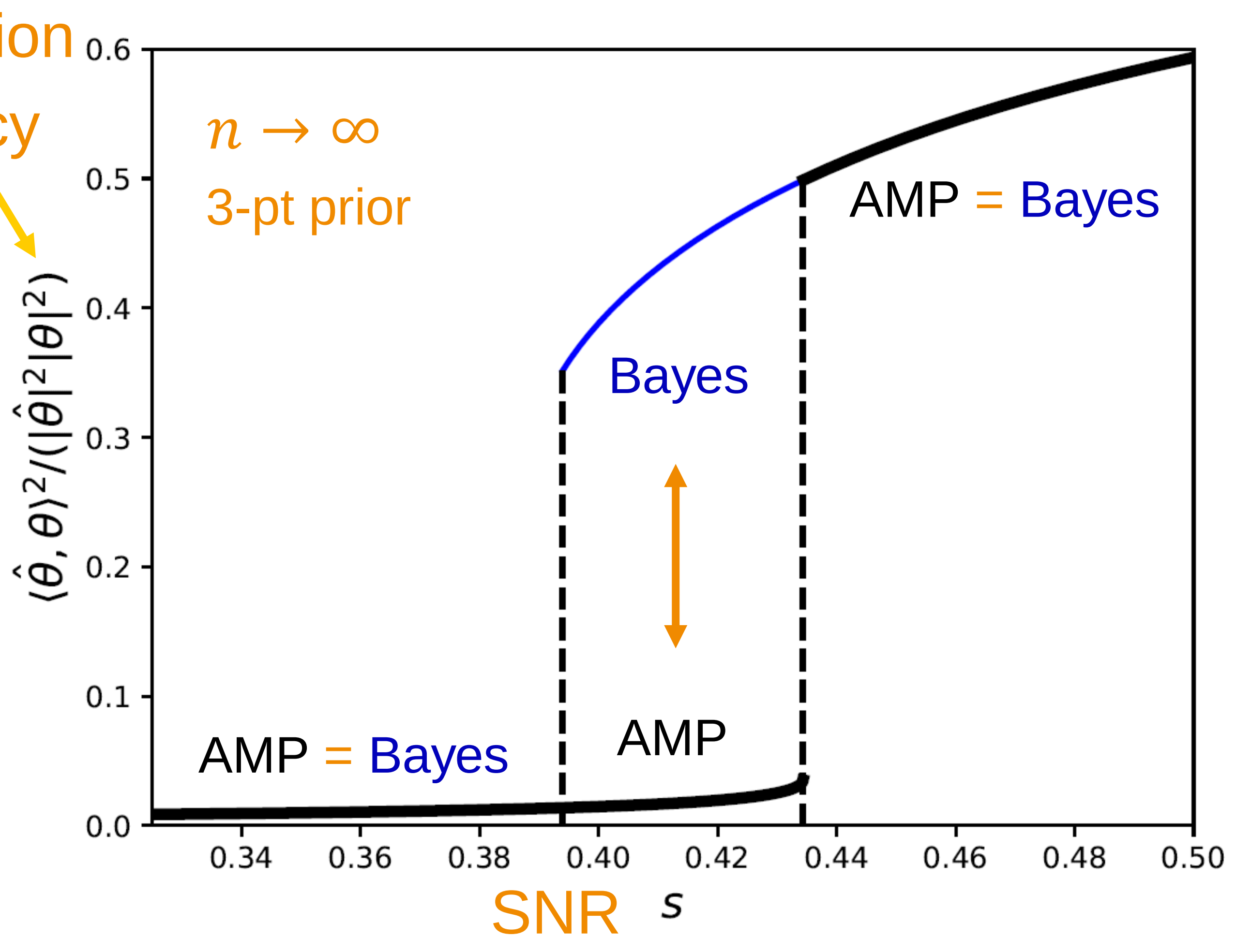
vector, estimate for θ

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

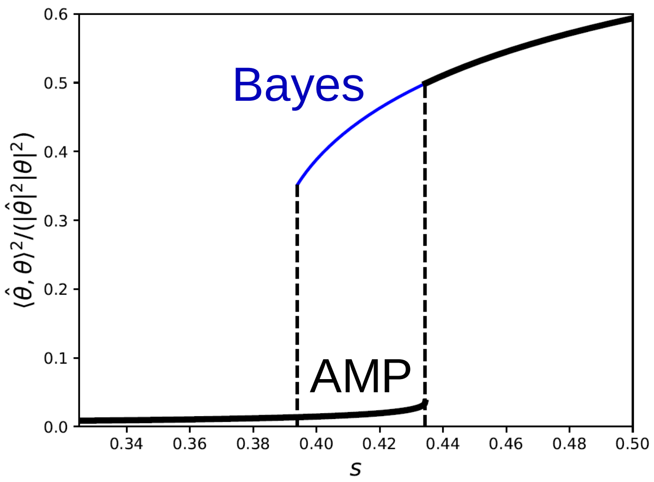
entrywise transform

Onsager term

Bayes is not comp. efficient, gap



Main Result



Conjecture

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 $(\hat{\theta}_1(Y), \ldots, \hat{\theta}_n(Y))$ where $\hat{\theta}_i$ is a const-deg multivariate polynomial in the entries of Y

We show AMP is the <u>best</u> 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 tensor PCA [Montanari,Richard'14]

Kikuchi hierarchy "redeems" physics [**W**,Alaoui,Moore'19]

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 $\lim_{\mathrm{D} \to \infty} \lim_{\mathrm{n} \to \infty} \mathrm{MMSE}_{\leq D}$

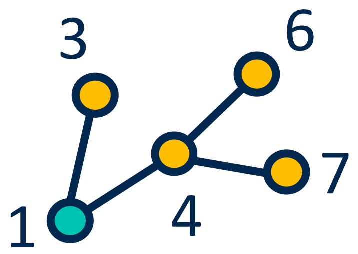
$$Y = \frac{s}{\sqrt{n}}\theta\theta^\top + Z$$

Proof Sketch: AMP vs Low-Deg

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

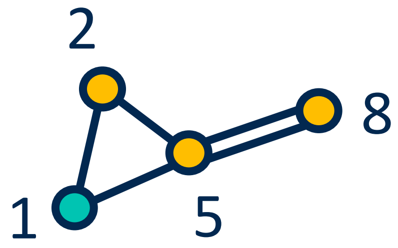
II. Tree-shaped polynomials are as powerful as all polynomials (of the same degree)

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

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

Conclude:

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

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

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

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

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

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

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?

Thanks!