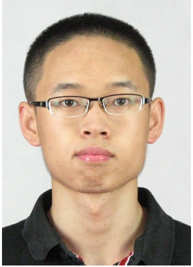
A Non-asymptotic Framework for Approximate Message Passing Algorithm

Yuting Wei

Statistics & Data Science, Wharton
University of Pennsylvania

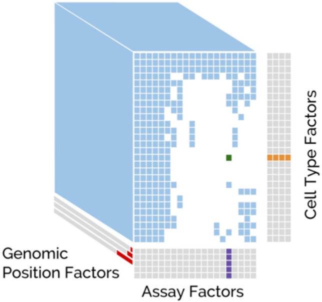
UC Davis, 2022

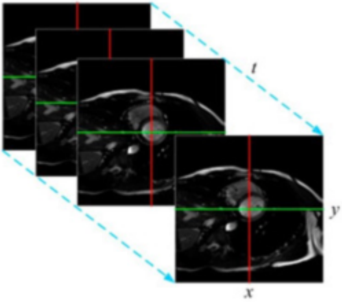


Gen Li, UPenn Statistics

"A Non-Asymptotic Framework for Approximate Message Passing in Spiked Models," Gen Li, Yuting Wei, *arxiv.2208.03313*

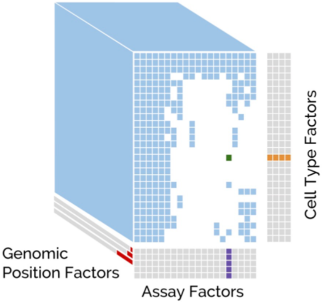
High-dimensional statistical tasks

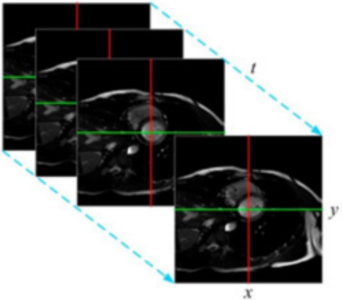




Statistical tasks: linear regression, generalized linear models, low-rank matrix estimation, phase retrieval, tensor decomposition...

High-dimensional statistical tasks



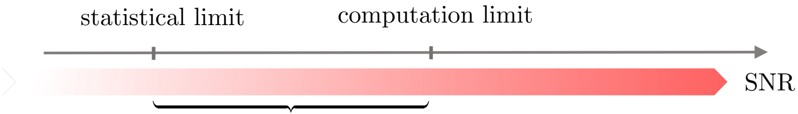


Statistical tasks: linear regression, generalized linear models, low-rank matrix estimation, phase retrieval, tensor decomposition...

When problem sizes are large, computation complexity is an issue!

Statistical accuracy vs. computation complexity

Problems with combinatorial nature (e.g. community detection, planted cliques, sparse principal component analysis, structured matrix models, sparse tensor models...)



"I can't find an efficient algorithm, but neither can all these people."

— *see survey Bandeira, Perry, Wein (2018)*

Approximate message passing (AMP) algorithm

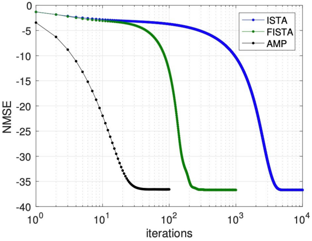
- AMP is a low-complexity, iterative algorithm
  [Donoho, Maleki, Montanari (2009, 2010a, 2011b), Bayati & Montanari (2011)]

Approximate message passing (AMP) algorithm

- AMP is a low-complexity, iterative algorithm
  [Donoho, Maleki, Montanari (2009, 2010a, 2011b), Bayati & Montanari (2011)]
- Theoretically optimal vs. computationally feasible estimators
  [Reeves, Pfister (2019), Barbier et al. (2017), Lelarge & Miolane (2019), Montanari & Ramji (2019), Celentano & Montanari (2019)]

Approximate message passing (AMP) algorithm

- AMP is a low-complexity, iterative algorithm
  [Donoho, Maleki, Montanari (2009, 2010a, 2011b), Bayati & Montanari (2011)]
- Theoretically optimal vs. computationally feasible estimators
  [Reeves, Pfister (2019), Barbier et al. (2017), Lelarge & Miolane (2019), Montanari & Ramji (2019), Celentano & Montanari (2019)]



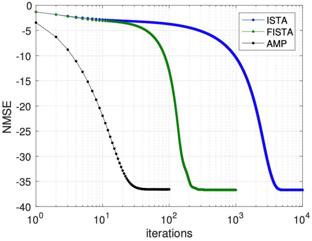
AMP in computing LASSO

Advantages of AMP:

- fast convergence
- asymptotically exact characterization
- easily combine with prior info on signal structure

Approximate message passing (AMP) algorithm

- AMP is a low-complexity, iterative algorithm
  [Donoho, Maleki, Montanari (2009, 2010a, 2011b), Bayati & Montanari (2011)]
- Theoretically optimal vs. computationally feasible estimators
  [Reeves, Pfister (2019), Barbier et al. (2017), Lelarge & Miolane (2019), Montanari & Ramji (2019), Celentano & Montanari (2019)]



AMP in computing LASSO

Advantages of AMP:

- fast convergence
- asymptotically exact characterization
- easily combine with prior info on signal structure

— *tutorial, Feng, Venkataramanan, Rush, Samworth (2022)*

Prior theory of AMP

Exact asymptotics: for constant # iterations t (e.g. $t = 20$), empirical distribution of the coordinates of AMP iterate x_t is approximately Gaussian ($n \to \infty$), with variance given by low-dimensional recursion:

state evolution: $\tau_{t+1} = F(\tau_t)$

τ_t captures the variance at iteration t

[Bayati & Montanari (2011), Javanmard & Montanari (2013), Schniter & Rangan (2014)]

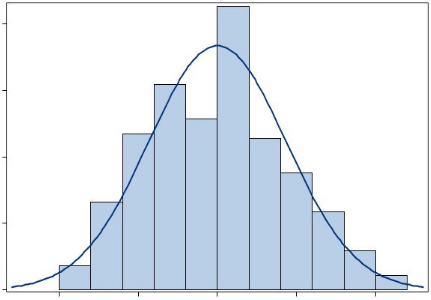
Prior theory of AMP

Exact asymptotics: for constant # iterations t (e.g. $t = 20$), empirical distribution of the coordinates of AMP iterate x_t is approximately Gaussian ($n \to \infty$), with variance given by low-dimensional recursion:

state evolution: $\tau_{t+1} = F(\tau_t)$

τ_t captures the variance at iteration t

[Bayati & Montanari (2011), Javanmard & Montanari (2013), Schniter & Rangan (2014)]



histogram of coordinates of x_t

Prior theory of AMP

Exact asymptotics: for constant # iterations t (e.g. $t = 20$), empirical distribution of the coordinates of AMP iterate x_t is approximately Gaussian ($n \to \infty$), with variance given by low-dimensional recursion:

state evolution: $\tau_{t+1} = F(\tau_t)$

τ_t captures the variance at iteration t

[Bayati & Montanari (2011), Javanmard & Montanari (2013), Schniter & Rangan (2014)]

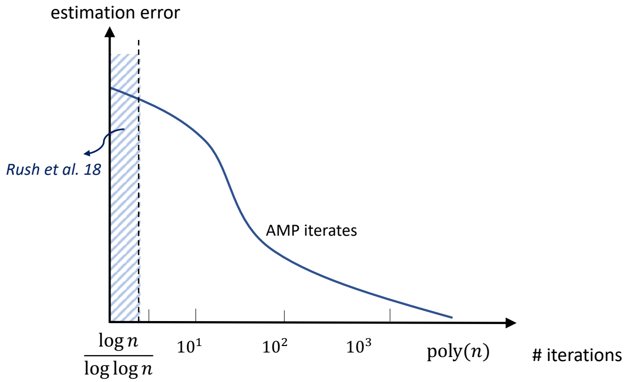
Another benefit: AMP as a tool to analyze statistical procedures

[Donoho, Maleki, Montanari (2009), Donoho & Montanari (2016), Sur, Chen, Candès. (2017), Bu et al. (2020), Fan & Wu (2021), Li & Wei (2021)...]

Non-asymptotic analyses are quite limited so far...

- compare to other optimization methods
- compare to other analysis techniques

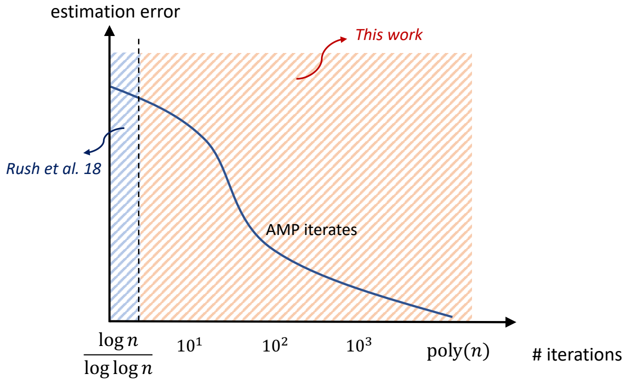
Non-asymptotic analysis?



Non-asymptotic result: Rush & Venkataramanan (2018)

$\#$iterations $= o(\log n / \log\log n)$ (*based on state-evolution analysis*)

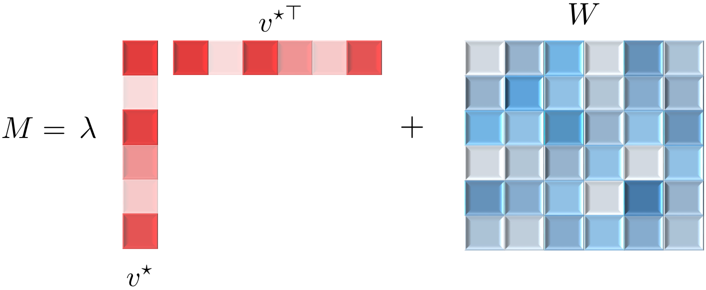
Non-asymptotic analysis?



Question: Is it possible to develop non-asymptotic analysis of AMP beyond $o(\log n/\log\log n)$ iterations?

AMP for signal recovery in spiked models

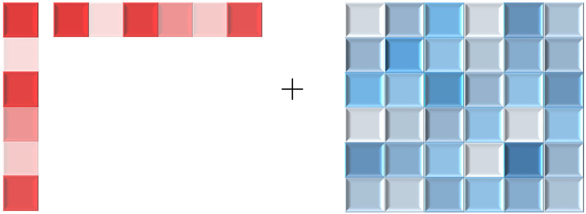
Spiked Wigner model



$$M = \lambda v^{\star} v^{\star\top} + W$$

Johnstone (2001),

Spiked Wigner model

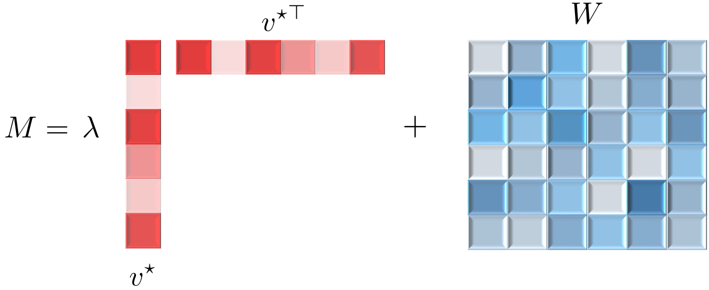


$$M = \lambda\, v^\star v^{\star\top} + W$$

- $W_{ij} = W_{ji} \sim \mathcal{N}(0, \frac{1}{n})$ and $W_{ii} \sim \mathcal{N}(0, \frac{2}{n})$

Johnstone (2001),

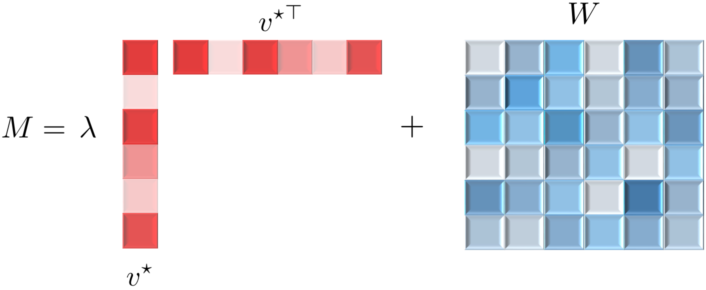
Spiked Wigner model



- $W_{ij} = W_{ji} \sim \mathcal{N}(0, \frac{1}{n})$ and $W_{ii} \sim \mathcal{N}(0, \frac{2}{n})$
- λ = SNR (signal-to-noise ratio) with $\|v^{\star}\|_2 = 1$
- Goal: estimate $v^{\star}$ from M

Johnstone (2001),

Spiked Wigner model

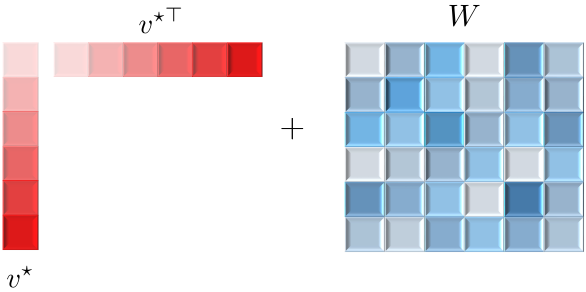


- $W_{ij} = W_{ji} \sim \mathcal{N}(0, \frac{1}{n})$ and $W_{ii} \sim \mathcal{N}(0, \frac{2}{n})$
- λ = SNR (signal-to-noise ratio) with $\|v^\star\|_2 = 1$
- Goal: estimate $v^\star$ from M
- Phase transition at $\lambda > 1$: the top eigenvalue separates from bulk, eigenvector correlates non-trivially with $v^\star$

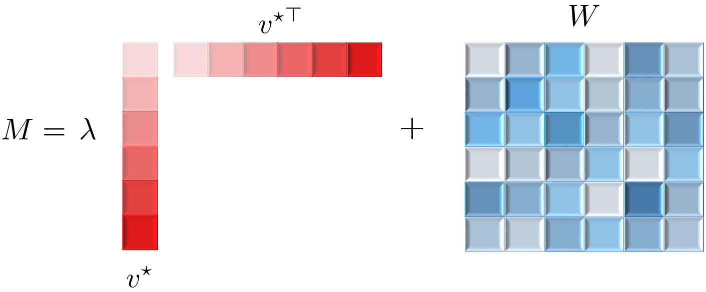
Johnstone (2001), Johnstone & Lu (2004), Péché (2006), Baik & Silverstein (2006), Capitaine, Donati-Martin & Féral (2009), Féral & Péché (2007)...

Spiked Wigner model with structures

$M = \lambda$

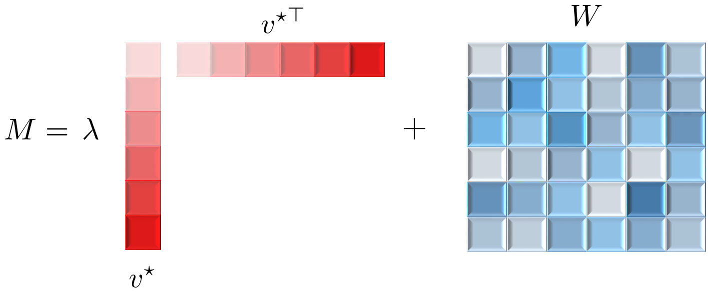


Spiked Wigner model with structures



Applications: spin-glass problems, community detection, image alignment, angular synchronization

Spiked Wigner model with structures

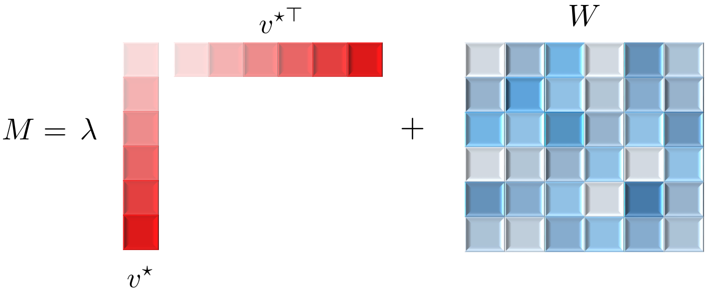


Applications: spin-glass problems, community detection, image alignment, angular synchronization

- $\mathbb{Z}_2$ synchronization: $\sqrt{n}v_i^\star \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\{+1, -1\}$

Singer (2011), Panchenko (2013), Deshpande, Abbe & Montanari (2016), Perry, Wein, Bandeira, Moitra (2018), Javanmard, Montanari & Ricci-Tersenghi (2016)...

Spiked Wigner model with structures

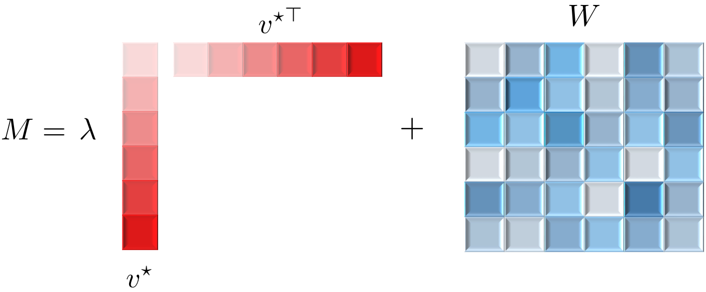


Applications: spin-glass problems, community detection, image alignment, angular synchronization

- $\mathbb{Z}_2$ synchronization: $\sqrt{n}v_i^\star \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\{+1, -1\}$
- sparse Wigner model: $\|v^\star\|_0 = k$

Singer (2011), Panchenko (2013), Deshpande, Abbe & Montanari (2016), Perry, Wein, Bandeira, Moitra (2018), Javanmard, Montanari & Ricci-Tersenghi (2016)...

Spiked Wigner model with structures

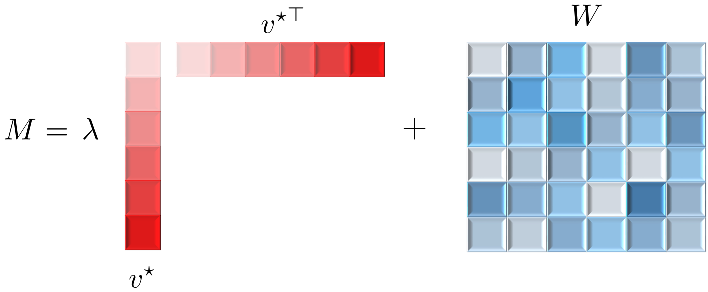


Applications: spin-glass problems, community detection, image alignment, angular synchronization

- $\mathbb{Z}_2$ synchronization: $\sqrt{n}v_i^{\star} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\{+1, -1\}$
- sparse Wigner model: $\|v^{\star}\|_0 = k$
- non-negative Wigner model: $v_i^{\star} \geq 0$

Singer (2011), Panchenko (2013), Deshpande, Abbe & Montanari (2016), Perry, Wein, Bandeira, Moitra (2018), Javanmard, Montanari & Ricci-Tersenghi (2016)...

Spiked Wigner model with structures



Applications: spin-glass problems, community detection, image alignment, angular synchronization

- $\mathbb{Z}_2$ synchronization: $\sqrt{n}v_i^\star \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\{+1, -1\}$
- sparse Wigner model: $\|v^\star\|_0 = k$
- non-negative Wigner model: $v_i^\star \geq 0$
- cone-constrained spiked models: $v^\star \in \mathcal{K}$ (e.g. monotone, convex)

Singer (2011), Panchenko (2013), Deshpande, Abbe & Montanari (2016), Perry, Wein, Bandeira, Moitra (2018), Javanmard, Montanari & Ricci-Tersenghi (2016)...

An incomplete list of prior art

$\mathbb{Z}_2$ synchronization:

- Panchenko'13
- Baik, Arous, Péché'05
- Javanmard et al.'16
- Montanari & Sen'16
- Lelarge & Miolane'19
- Deshpande, Abbe, Montanari'17
- Celentano, Fan, Mei'21

general convex cones:

- Deshpande, Montanari, Richard'14
- Lesieur, Krzakala, Zdeborová'17
- Bandeira, Kunisky, Wein'19

sparse PCA (Wigner / Wishart)

- Johnstone & Lu'09
- d'Aspremont et al.'04
- Vu & Lei'12
- Berthet & Rigollet'13
- Ma'13
- Lesieur, Krzakala, Zdeborová'15
- Deshpande & Montanari'14
- Wang, Berthet, Samworth'16
- Ding, Kunisky, Wein, Bandeira'19

positive Wigner models

- Montanari & Richard'16

Idealistic estimators

$$\text{Maximum likelihood estimator} := \arg \min_{\substack{v \in \mathcal{S}^{n-1} \\ v \text{ with structures}}} \|M - \lambda vv^\top\|_F^2$$

$$\text{Bayes optimal estimator} := \mathbb{E}[vv^\top \mid M]$$

AMP for spiked models

AMP for spiked models:

$$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t) \rangle \cdot \eta_{t-1}(x_{t-1}), \text{ for } t \geq 1$$

where $\langle x \rangle := \frac{1}{n}\sum_{i=1}^{n} x_i$.

- η_t: denoising function selected *a priori* (tailored to structure of $v^\star$)
  - $\mathbb{Z}_2$ **synchronization:** $\eta_t(x) = \rho_t \tanh(x)$
  - **sparse estimation:** $\eta_t(x) = \rho_t \cdot \mathrm{sign}(x)(|x| - \tau_t)_+$
  - **general cone:** $\eta_t(x) = \rho_t \cdot \mathsf{Proj}_{\mathcal{K}}(x)$

AMP for spiked models

AMP for spiked models:

$$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t) \rangle \cdot \eta_{t-1}(x_{t-1}), \text{ for } t \geq 1$$

where $\langle x \rangle := \frac{1}{n}\sum_{i=1}^{n} x_i$.

- η_t: denoising function selected *a priori* (tailored to structure of $v^\star$)
  - $\mathbb{Z}_2$ **synchronization:** $\eta_t(x) = \rho_t \tanh(x)$
  - **sparse estimation:** $\eta_t(x) = \rho_t \cdot \mathrm{sign}(x)(|x| - \tau_t)_+$
  - **general cone:** $\eta_t(x) = \rho_t \cdot \mathsf{Proj}_{\mathcal{K}}(x)$
- effectiveness of AMP — *Onsager correction term* $\langle \eta_t'(x_t) \rangle \cdot \eta_{t-1}(x_{t-1})$

Prior results: exact asymptotics

Theorem (Montanari & Venkataramanan (2019))

Suppose the empirical distribution $\{v_i^\star\}_{i=1}^n \to \mu_V$ *on* $\mathbb{R}$, *with* $\mathbb{E}[V^2] = 1$.
For constant # iterations t *(independent of* n*), it satisfies,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(x_{t,i} - v_i^\star)^2 = \mathbb{E}\Big[\big(\alpha_t V + \beta_t G - V\big)^2\Big], \qquad \text{almost surely}$$

where $V \sim \mu_V$ *and* $G \sim \mathcal{N}(0,1)$ *are independent.*

Prior results: exact asymptotics

Theorem (Montanari & Venkataramanan (2019))

Suppose the empirical distribution $\{v_i^\star\}_{i=1}^n \to \mu_V$ *on* $\mathbb{R}$*, with* $\mathbb{E}[V^2] = 1$*. For constant* $\#$ *iterations* t *(independent of* n*), it satisfies,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n (x_{t,i} - v_i^\star)^2 = \mathbb{E}\Big[\big(\alpha_t V + \beta_t G - V\big)^2\Big], \qquad \textit{almost surely}$$

where $V \sim \mu_V$ *and* $G \sim \mathcal{N}(0,1)$ *are independent.*

- State Evolution (SE) via the recursion

$$(\alpha_{t+1}, \beta_{t+1}) = F(\alpha_t, \beta_t) \;=\; \begin{cases} \alpha_{t+1} = \lambda \mathbb{E}\big[V \cdot \eta_t(\alpha_t V + \beta_t G)\big] \\ \beta_{t+1}^2 = \mathbb{E}\big[\eta_t^2(\alpha_t V + \beta_t G)\big] \end{cases}$$

- **Challenges for non-asymptotic guarantees:** deal with statistical dependence between iterations

AMP for spiked models:

$$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t) \rangle \cdot \eta_{t-1}(x_{t-1}), \text{ for } t \geq 1$$

where $\langle x \rangle := \frac{1}{n}\sum_{i=1}^{n} x_i$.

This work: a new decomposition of AMP

Theorem (Li & Wei (2022))

Initialize AMP with x_1 independent of W. For every $1 \le t \le n$, AMP yields the decomposition

$$x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t, \qquad (\star)$$

for $\phi_k \overset{i.i.d.}{\sim} \mathcal{N}(0, \frac{1}{n} I_n)$.

This work: a new decomposition of AMP

Theorem (Li & Wei (2022))

Initialize AMP with x_1 independent of W. For every $1 \le t \le n$, AMP yields the decomposition

$$x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t, \qquad (\star)$$

for $\phi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{n} I_n)$.

here $(\alpha_{t+1}, \beta_t, \xi_t)$ obeys

$$\alpha_{t+1} = \lambda {v^\star}^\top \eta_t(x_t),$$

$$\beta_t^k = \langle \eta_t(x_t), z_k \rangle \qquad \text{for an explicit-defined basis } \{z_k\}$$

$$\|\xi_t\|_2 = \Big\langle \sum_{k=1}^{t-1} \mu^k \phi_k, \delta_t \Big\rangle - \langle \delta_t' \rangle \sum_{k=1}^{t-1} \mu^k \beta_{t-1}^k + \Delta_t + O\Big(\sqrt{\frac{t \log n}{n}} \|\beta_t\|_2\Big) \quad \text{w.h.p.}$$

This work: a new decomposition of AMP

Theorem (Li & Wei (2022))

Initialize AMP with x_1 independent of W. For every $1 \le t \le n$, AMP yields the decomposition

$$x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t, \qquad (\star)$$

for $\phi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{n} I_n)$.

- x_t behaves like $\alpha_t v^\star + \sum_{k=1}^{t-1} \beta_{t-1}^k \phi_k$ if $\|\xi_{t-1}\|_2$ is small

This work: a new decomposition of AMP

Theorem (Li & Wei (2022))

Initialize AMP with x_1 independent of W. For every $1 \le t \le n$, AMP yields the decomposition

$$x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t, \qquad (\star)$$

for $\phi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{n} I_n)$.

- x_t behaves like $\alpha_t v^\star + \sum_{k=1}^{t-1} \beta_{t-1}^k \phi_k$ if $\|\xi_{t-1}\|_2$ is small
- if η_t are nice (*smooth & with finite jumps*), we can track how $\|\xi_t\|_2$ depends on λ, t, n

This work: a new decomposition of AMP

Theorem (Li & Wei (2022))

Initialize AMP with x_1 independent of W. For every $1 \le t \le n$, AMP yields the decomposition

$$x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t, \qquad (\star)$$

for $\phi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{n} I_n)$.

- x_t behaves like $\alpha_t v^\star + \sum_{k=1}^{t-1} \beta_{t-1}^k \phi_k$ if $\|\xi_{t-1}\|_2$ is small
- if η_t are nice (*smooth & with finite jumps*), we can track how $\|\xi_t\|_2$ depends on λ, t, n
- decomposition $(\star)$ can be extended for spectral initialization

Applications in two examples:
$\mathbb{Z}_2$ *synchronization* & *sparse Wigner model*

Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim \mathsf{Unif}(\{\pm 1\})$
- Goal: recover $v^\star$ given M

Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim$ Unif$(\{\pm 1\})$
- Goal: recover $v^\star$ given M

Deshpande, Abbe, Montanari (2016) characterizes the theoretical limit of this problem

$$\lim_{n \to \infty} \mathbb{E}\left[\left\| v^\star v^{\star\top} - \underbrace{\mathbb{E}[v v^\top \mid M]}_{\text{Bayes estimator}} \right\|_F^2 \right] = \begin{cases} 1 & \lambda \le 1; \\ 1 - q^\star(\lambda)^2 & \lambda > 1. \end{cases}$$

Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim$ Unif$(\{\pm 1\})$
- Goal: recover $v^\star$ given M

Deshpande, Abbe, Montanari (2016) characterizes the theoretical limit of this problem

$$\lim_{n\to\infty} \mathbb{E}\Big[\big\| v^\star v^{\star\top} - \underbrace{\mathbb{E}[vv^\top \mid M]}_{\text{Bayes estimator}} \big\|_F^2\Big] = \begin{cases} 1 & \lambda \le 1; \\ 1 - q^\star(\lambda)^2 & \lambda > 1. \end{cases}$$

Starting from an informative initialization, AMP is ϵ-close to the Bayes optimal estimator (in large n limit)

Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim$ Unif$(\{\pm 1\})$
- Goal: recover $v^\star$ given M

Deshpande, Abbe, Montanari (2016) characterizes the theoretical limit of this problem

$$\lim_{n\to\infty} \mathbb{E}\Big[\big\| v^\star v^{\star\top} - \underbrace{\mathbb{E}[vv^\top \mid M]}_{\text{Bayes estimator}} \big\|_F^2\Big] = \begin{cases} 1 & \lambda \le 1; \\ 1 - q^\star(\lambda)^2 & \lambda > 1. \end{cases}$$

Starting from an informative initialization, AMP is ϵ-close to the Bayes optimal estimator (in large n limit)

— ✓ *initialize by spectral methods* *Montanari & Venkataramanan (2019)*

Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim$ Unif$(\{\pm 1\})$
- Goal: recover $v^\star$ given M

— *AMP succeeds in a fixed t, large n limit*

Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim \text{Unif}(\{\pm 1\})$
- Goal: recover $v^\star$ given M

— *AMP succeeds in a fixed t, large n limit*

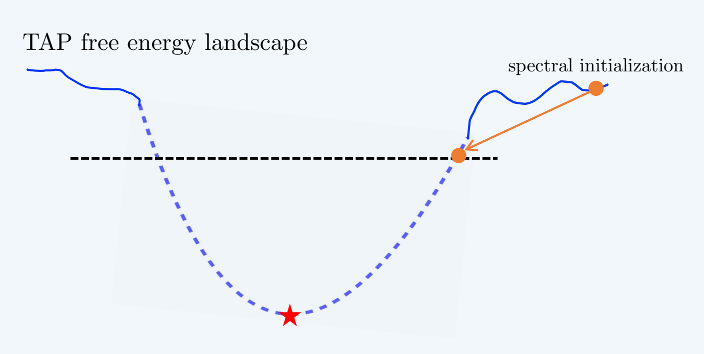
A hybrid procedure proposed in Celentano, Fan, Mei (2021)

Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim \mathsf{Unif}(\{\pm 1\})$
- Goal: recover $v^\star$ given M

— *AMP succeeds in a fixed t, large n limit*

A hybrid procedure proposed in Celentano, Fan, Mei (2021)

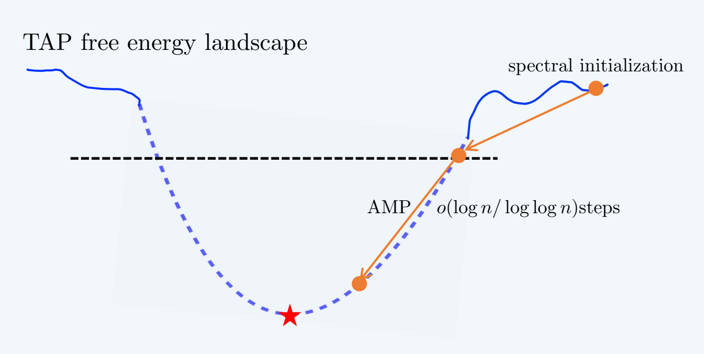


Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim$ Unif$(\{\pm 1\})$
- Goal: recover $v^\star$ given M

— *AMP succeeds in a fixed t, large n limit*

A hybrid procedure proposed in Celentano, Fan, Mei (2021)

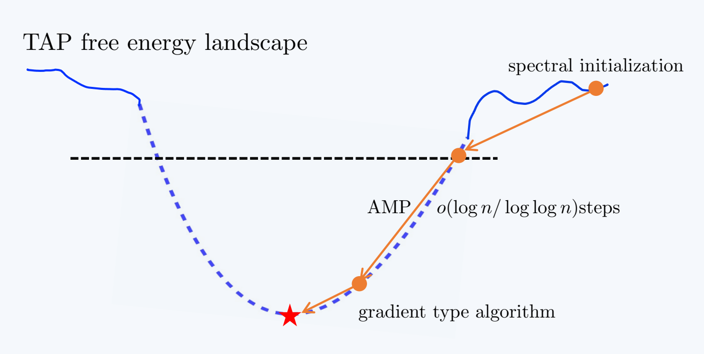


Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim \mathsf{Unif}(\{\pm 1\})$
- Goal: recover $v^\star$ given M

— *AMP succeeds in a fixed t, large n limit*

A hybrid procedure proposed in Celentano, Fan, Mei (2021)

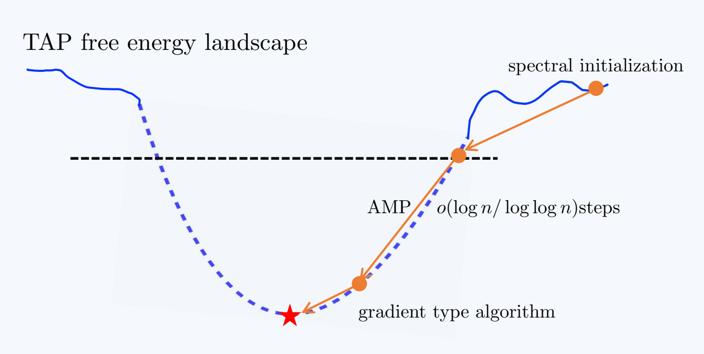


Example 1: $\mathbb{Z}_2$ Synchronization

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\sqrt{n} v_i^\star \sim$ Unif$(\{\pm 1\})$
- Goal: recover $v^\star$ given M

— *AMP succeeds in a fixed t, large n limit*

A hybrid procedure proposed in Celentano, Fan, Mei (2021)



Open question: spectral-initialized AMP converges for $\lambda > 1$?

$\mathbb{Z}_2$ Synchronization: our results

Theorem (Li & Wei (2022))

Spectrally initialized AMP satisfies

$$x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t,$$

with

$$\alpha_{t+1} = \mathbb{E}\left[\lambda v^{\star\top} \eta_t\left(\alpha_t v^\star + \frac{1}{\sqrt{n}} G\right)\right] + O\left(\sqrt{\frac{t \log n}{(\lambda-1)^3 n}}\right),$$

$$\|\beta_t\|_2 = 1, \qquad \|\xi_t\|_2 \lesssim O\left(\sqrt{\frac{t \log n}{(\lambda-1)^3 n}} + \sqrt{\frac{\log^7 n}{(\lambda-1)^9 n}}\right)$$

w.h.p. provided that $t \lesssim \frac{(\lambda-1)^{10}}{\log^7 n} n$.

- denoising functions:

$$\eta_t(x) := \tanh(\pi_t x)/\|\tanh(\pi_t x)\|_2, \quad \text{where } \pi_t^2 = n(\|x_t\|_2^2 - 1)$$

- record (asymptotic) State Evolution:

$$\tau_{t+1} \coloneqq \lambda^2 \int \tanh(\tau_t + \sqrt{\tau_t}x)\varphi(\mathrm{d}x)$$

then

$$\alpha_t^2 - \tau_{t+1} = O\left(\sqrt{\frac{t \log n}{(\lambda-1)^8 n}} + \sqrt{\frac{\log^7 n}{(\lambda-1)^{14} n}}\right)$$

- record (asymptotic) State Evolution:

$$\tau_{t+1} \coloneqq \lambda^2 \int \tanh(\tau_t + \sqrt{\tau_t} x) \varphi(\mathrm{d}x)$$

then

$$\alpha_t^2 - \tau^\star = (\lambda^2 - 1)(1 - (\lambda - 1))^t + O\left(\sqrt{\frac{t \log n}{(\lambda - 1)^8 n}} + \sqrt{\frac{\log^7 n}{(\lambda - 1)^{14} n}}\right)$$

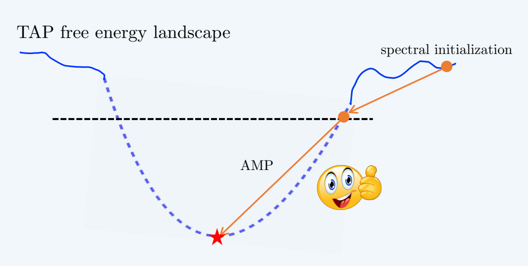
- record (asymptotic) State Evolution:

$$\tau_{t+1} := \lambda^2 \int \tanh(\tau_t + \sqrt{\tau_t} x)\varphi(\mathrm{d}x)$$

then

$$\alpha_t^2 - \tau^\star = (\lambda^2 - 1)(1 - (\lambda - 1))^t + O\left(\sqrt{\frac{t \log n}{(\lambda - 1)^8 n}} + \sqrt{\frac{\log^7 n}{(\lambda - 1)^{14} n}}\right)$$

- answer the open question (Celentano, Fan & Mei (2021)) positively: spectral-initialized AMP is enough!



$\mathbb{Z}_2$ Synchronization: simulations

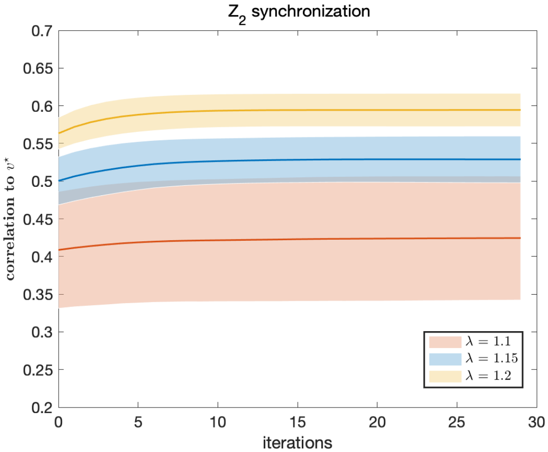


Figure: Convergence of spectrally initialized AMP for different signal strengths with $n = 10000$. Repeat 40 times.

Example 2: sparse PCA

- Setting: $M = \lambda v^\star v^{\star\top} + W$ where $\|v^\star\|_0 = k$
- Goal: recover $v^\star$ given M

$\lambda \approx \sqrt{\frac{k \log n}{n}}$

statistical limit

$\lambda \approx \sqrt{\frac{k^2}{n}}$

reduction to planted cliques: Berthet & Rigollet (2013)

computation limit

SNR

Zou et al. (2006)
Amini and Wainwright (2008)
Ma (2013)
Deshpande and Montanari (2014b)
Hopkins et al. (2017)

"I can't find an efficient algorithm, but neither can all these people."

Sparse PCA: our results

Theorem (Li & Wei (2022))

Suppose $0 < \lambda \lesssim 1$. *Given an informative initialization (with non-vanishing correlation with* $v^\star$*), AMP satisfies*

$$x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t,$$

with

$$\alpha_{t+1} = \mathbb{E}\left[\lambda v^{\star\top} \eta_t\left(\alpha_t v^\star + \frac{1}{\sqrt{n}} G\right)\right] + \sqrt{\frac{k + t\log^3 n}{n}},$$

$$\|\beta_t\|_2 = 1, \qquad \|\xi_t\|_2 \lesssim \sqrt{\frac{k + t\log^3 n}{n}} \qquad \textit{w.h.p.}$$

provided that $\frac{t\log^3 n}{n\lambda^2} \lesssim 1$ *and* $\frac{k\log n}{n\lambda^2} \lesssim 1$.

Sparse PCA: our results

Theorem (Li & Wei (2022))

Suppose $0 < \lambda \lesssim 1$. *Given an informative initialization (with non-vanishing correlation with* $v^\star$*), AMP satisfies*

$$x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t,$$

with

$$\alpha_{t+1} = \mathbb{E}\left[\lambda v^{\star\top} \eta_t\Big(\alpha_t v^\star + \frac{1}{\sqrt{n}} G\Big)\right] + \sqrt{\frac{k + t\log^3 n}{n}},$$

$$\|\beta_t\|_2 = 1, \qquad \|\xi_t\|_2 \lesssim \sqrt{\frac{k + t\log^3 n}{n}} \qquad \textit{w.h.p.}$$

provided that $\frac{t\log^3 n}{n\lambda^2} \lesssim 1$ *and* $\frac{k\log n}{n\lambda^2} \lesssim 1$.

denoising functions:

$$\eta_t(x) = \gamma_t \mathrm{sign}(x)(|x| - \tau_t)_+ \quad \text{where } \gamma_t^{-1} := \|(|x_t| - \tau_t)_+\|_2, \tau_t \asymp \sqrt{\frac{\log n}{n}}$$

Several remarks

- record (asymptotic) State Evolution:

$$\alpha_{t+1}^{\star} := \frac{\lambda v^{\star\top} \int \mathsf{ST}_{\tau_t}\left(\alpha_t^{\star} v^{\star} + \frac{x}{\sqrt{n}}\right) \varphi_n(\mathrm{d}x)}{\sqrt{\int \left\| \mathsf{ST}_{\tau_t}\left(\alpha_t^{\star} v^{\star} + \frac{x}{\sqrt{n}}\right) \right\|_2^2 \varphi_n(\mathrm{d}x)}}$$

  then

$$\left|\alpha_{t+1} - \alpha_{t+1}^{\star}\right| \lesssim \sqrt{\frac{k \log n + t \log^3 n}{n}}$$

Several remarks

- record (asymptotic) State Evolution:

$$\alpha_{t+1}^{\star} := \frac{\lambda v^{\star\top} \int \mathsf{ST}_{\tau_t}\left(\alpha_t^{\star} v^{\star} + \frac{x}{\sqrt{n}}\right) \varphi_n(\mathrm{d}x)}{\sqrt{\int \left\|\mathsf{ST}_{\tau_t}\left(\alpha_t^{\star} v^{\star} + \frac{x}{\sqrt{n}}\right)\right\|_2^2 \varphi_n(\mathrm{d}x)}}$$

  then

$$\left|\alpha_{t+1} - \alpha_{t+1}^{\star}\right| \lesssim \sqrt{\frac{k \log n + t \log^3 n}{n}}$$

- two sufficient initialization schemes:
  - AMP with diagonal maximization: $\lambda \|v^{\star}\|_{\infty} \gtrsim \sqrt{\frac{k \log n}{n}}$
  - AMP with sample-split initialization: $\lambda \gtrsim \sqrt{\frac{k^2}{n}}$ and $\|v^{\star}\|_{\infty} \lesssim \frac{\log n}{k}$

Sparse PCA: simulations

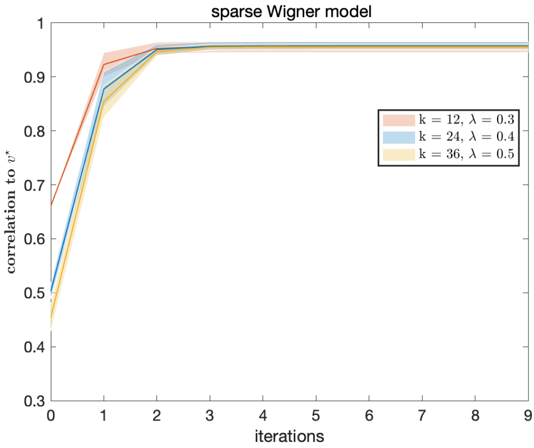


Figure: Convergence of AMP with diagonal maximization for different signal strengths with $n = 10000$. Repeat 40 times.

Questions?

A glimpse of our main proof idea...

— *decomposition:* $x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t$

Main proof idea: a new decomposition

- $\{x_t\}$ is the sequence generated by AMP

Main proof idea: a new decomposition

- $\{x_t\}$ is the sequence generated by AMP
- define an orthonormal basis $\{z_t\}$ where

$$z_1 := \frac{\eta_1(x_1)}{\|\eta_1(x_1)\|_2} \in \mathbb{R}^n, \qquad \text{and} \qquad W_1 := W$$

Main proof idea: a new decomposition

- $\{x_t\}$ is the sequence generated by AMP
- define an orthonormal basis $\{z_t\}$ where

$$z_1 \coloneqq \frac{\eta_1(x_1)}{\|\eta_1(x_1)\|_2} \in \mathbb{R}^n, \qquad \text{and} \qquad W_1 \coloneqq W$$

- write $U_{t-1} \coloneqq [z_k]_{1 \le k \le t-1} \in \mathbb{R}^{n \times (t-1)}$ and denote

$$z_t \coloneqq \frac{\left(I - U_{t-1}U_{t-1}^\top\right)\eta_t(x_t)}{\left\|\left(I - U_{t-1}U_{t-1}^\top\right)\eta_t(x_t)\right\|_2} \qquad \textit{Gram–Schmidt orthogonalization},$$

$$W_t \coloneqq \left(I - z_{t-1}z_{t-1}^\top\right) W_{t-1} \left(I - z_{t-1}z_{t-1}^\top\right)$$

Main proof idea: a new decomposition

- $\{x_t\}$ is the sequence generated by AMP
- define an orthonormal basis $\{z_t\}$ where

$$z_1 := \frac{\eta_1(x_1)}{\|\eta_1(x_1)\|_2} \in \mathbb{R}^n, \qquad \text{and} \qquad W_1 := W$$

- write $U_{t-1} := [z_k]_{1 \le k \le t-1} \in \mathbb{R}^{n \times (t-1)}$ and denote

$$z_t := \frac{\left(I - U_{t-1}U_{t-1}^\top\right)\eta_t(x_t)}{\left\|\left(I - U_{t-1}U_{t-1}^\top\right)\eta_t(x_t)\right\|_2} \qquad \textit{Gram–Schmidt orthogonalization},$$

$$W_t := \left(I - z_{t-1}z_{t-1}^\top\right)W_{t-1}\left(I - z_{t-1}z_{t-1}^\top\right)$$

- write $\eta_t(x_t) = \sum_{k=1}^{t} \beta_t^k z_k$, for $\beta_t^k := \langle \eta_t(x_t), z_k \rangle$

Main proof idea: a new decomposition

- AMP updates:
  $$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t) \rangle \cdot \eta_{t-1}(x_{t-1}), \text{ where } M = \lambda v^\star v^{\star\top} + W$$
- Goal: $x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t$

Main proof idea: a new decomposition

- AMP updates:

  $$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t) \rangle \cdot \eta_{t-1}(x_{t-1}), \text{ where } M = \lambda v^\star v^{\star\top} + W$$

- Goal: $x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t$

$M\eta_t(x_t)$

Main proof idea: a new decomposition

- AMP updates:

$$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t) \rangle \cdot \eta_{t-1}(x_{t-1}), \text{ where } M = \lambda v^\star v^{\star\top} + W$$

- Goal: $x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^t \beta_t^k \phi_k + \xi_t$

$M\eta_t(x_t)$

$$= v^\star \underbrace{\lambda v^{\star\top} \eta_t(x_t)}_{\alpha_{t+1}} + \Big\{ W_t + \sum_{k=1}^{t-1} \underbrace{\Big[W_k z_k z_k^\top + z_k z_k^\top W_k - z_k z_k^\top W_k z_k z_k^\top \Big]}_{W_k - W_{k+1}} \Big\} \cdot \underbrace{\sum_{k=1}^{t} \beta_t^k z_k}_{\eta_t(x_t)}$$

Main proof idea: a new decomposition

- AMP updates:

$$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t)\rangle \cdot \eta_{t-1}(x_{t-1}), \text{ where } M = \lambda v^\star v^{\star\top} + W$$

- Goal: $x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t$

$$\begin{aligned} & M\eta_t(x_t) \\ &= v^\star \underbrace{\lambda v^{\star\top}\eta_t(x_t)}_{\alpha_{t+1}} + \Big\{ W_t + \sum_{k=1}^{t-1} \underbrace{\Big[W_k z_k z_k^\top + z_k z_k^\top W_k - z_k z_k^\top W_k z_k z_k^\top \Big]}_{W_k - W_{k+1}} \Big\} \cdot \underbrace{\sum_{k=1}^{t} \beta_t^k z_k}_{\eta_t(x_t)} \\ &= \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k W_k z_k + \dots \end{aligned}$$

Main proof idea: a new decomposition

- AMP updates:

  $$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t) \rangle \cdot \eta_{t-1}(x_{t-1}), \text{ where } M = \lambda v^\star v^{\star\top} + W$$

- Goal: $x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t$

$$\begin{aligned}
&M\eta_t(x_t) \\
&= v^\star \underbrace{\lambda v^{\star\top}\eta_t(x_t)}_{\alpha_{t+1}} + \Big\{ W_t + \sum_{k=1}^{t-1} \underbrace{\Big[W_k z_k z_k^\top + z_k z_k^\top W_k - z_k z_k^\top W_k z_k z_k^\top \Big]}_{W_k - W_{k+1}} \Big\} \cdot \underbrace{\sum_{k=1}^{t} \beta_t^k z_k}_{\eta_t(x_t)} \\
&= \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k W_k z_k + \dots \\
&= \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \underbrace{(W_k z_k + \zeta_k)}_{\phi_k \sim \mathcal{N}(0, \frac{1}{n}\mathbf{I}_n)} + \dots
\end{aligned}$$

Main proof idea: a new decomposition

- AMP updates:

$$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t)\rangle \cdot \eta_{t-1}(x_{t-1}), \text{ where } M = \lambda v^\star v^{\star\top} + W$$

- Goal: $x_{t+1} = \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \phi_k + \xi_t$

$$M\eta_t(x_t)$$

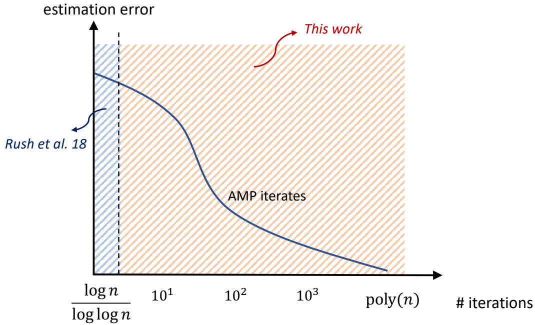
$$= v^\star \underbrace{\lambda v^{\star\top}\eta_t(x_t)}_{\alpha_{t+1}} + \Big\{ W_t + \sum_{k=1}^{t-1} \underbrace{\Big[W_k z_k z_k^\top + z_k z_k^\top W_k - z_k z_k^\top W_k z_k z_k^\top \Big]}_{W_k - W_{k+1}} \Big\} \cdot \underbrace{\sum_{k=1}^{t} \beta_t^k z_k}_{\eta_t(x_t)}$$

$$= \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k W_k z_k + \dots$$

$$= \alpha_{t+1} v^\star + \sum_{k=1}^{t} \beta_t^k \underbrace{(W_k z_k + \zeta_k)}_{\phi_k \sim \mathcal{N}(0, \frac{1}{n}\mathbf{I}_n)} + \dots$$

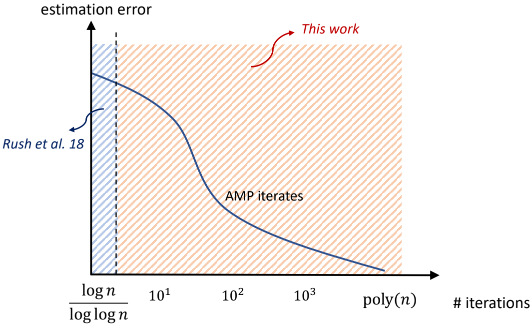
$$\xi_t = \sum_{k=1}^{t-1} z_k \Big[\langle W_k z_k, \eta_t(x_t)\rangle - \langle \eta_t'(x_t)\rangle \beta_{t-1}^k - \beta_t^k z_k^\top W_k z_k \Big] - \sum_{k=1}^{t} \beta_t^k \zeta_k$$

Concluding remarks



- A new non-asymptotic framework of AMP for spiked models that allows for # iterations $O(\frac{n}{\text{poly}(\log n)})$

Concluding remarks



- A new non-asymptotic framework of AMP for spiked models that allows for # iterations $O(\frac{n}{\mathrm{poly}(\log n)})$
- Apply our theory to two specific examples: $\mathbb{Z}_2$ synchronization & sparse PCA

Concluding remarks: future extensions

- Other settings: regression (*upcoming*), GLMs, phase retrieval, etc?

Concluding remarks: future extensions

- Other settings: regression (*upcoming*), GLMs, phase retrieval, etc?
- Can we understand AMP for general cone structures (*non-separable denoising functions*)?

Concluding remarks: future extensions

- Other settings: regression (*upcoming*), GLMs, phase retrieval, etc?
- Can we understand AMP for general cone structures (*non-separable denoising functions*)?
- Universality beyond Gaussian design (e.g. uniform Bernoulli, correlated columns, DFT)

Concluding remarks: future extensions

- Other settings: regression (*upcoming*), GLMs, phase retrieval, etc?
- Can we understand AMP for general cone structures (*non-separable denoising functions*)?
- Universality beyond Gaussian design (e.g. uniform Bernoulli, correlated columns, DFT)
- Connections between AMP and other optimization procedures

Thanks for your attention! Questions?

Paper:

"A Non-Asymptotic Framework for Approximate Message Passing in Spiked Models," G. Li, Y. Wei, *arxiv.2208.03313*

"Non-Asymptotic Analysis for Approximate Message Passing with Applications to Sparse and Robust Regression," G. Li, Y. Wei, *upcoming*

Non-asymptotic guarantees

- best known results: Rush & Venkataramanan (2016)
  $\#\text{iterations} = o(\log n / \log\log n)$ (*based on state-evolution analysis*)

$$\mathbb{P}(\text{residual at time } t \geq \epsilon)$$
$$= \mathbb{P}\Big(\sum_{i=0}^{t-1} r_i^t \geq \epsilon\Big) \overset{\text{statistical dependence}}{\boxed{\leq}} \sum_{i=0}^{t-1} \mathbb{P}\Big(r_i^t \leq \frac{\epsilon}{t}\Big) \overset{\text{induction step}}{\boxed{\leq}} tC_{t-1} \exp\Big(-\frac{c_{t-1}}{t^2} n\epsilon^2\Big)$$

  requires $\frac{n}{(t!)^2} \to \infty \to t = o(\log n / \log\log n)$

- main challenges: deal with statistical dependence between iterations

Conditioning technique

$$\text{AMP updates} \qquad x_{t+1} = Wm_t - \gamma_t m_{t-1}$$
$$\text{where} \qquad m^t = \eta_t(x_t), \quad \gamma_t = \langle \eta_t'(x_t) \rangle$$

- $m_{-1} = 0, x_0 = 0$ and $x_1 = W\eta_t(0)$
- σ-algebra $\mathcal{F}_t$ generated by $\{x_0, x_1, \ldots, x_t\}$, conditioning on $\mathcal{F}$ is equivalent to conditioning on event

$$\mathcal{E}_t := \left\{ x_1 + \gamma_0 m_{-1} = Wm_0,\ x_2 + \gamma_1 m_1 = Wm_1, \ldots,\ x_t + \gamma_{t-1} m_{t-1} = Wm_{t-1} \right\}$$

- W conditioning on linear observations

$$W|_{\mathcal{F}_t} \overset{\mathsf{d}}{=} \mathbb{E}[W|_{\mathcal{F}_t}] + P_t^{\perp} W^{\mathsf{new}} P_t^{\perp}$$
$$W|_{\mathcal{F}_t} m^t \overset{\mathsf{d}}{=} \underbrace{W^{\mathsf{new}} P_t^{\perp} m^t}_{\text{Gaussian term}} + \underbrace{W^{\mathsf{new}} (I - P_t^{\perp}) m^t + \mathbb{E}[W|\mathcal{F}_t] m^t}_{\text{non-Gaussian term}}$$

Bolthausen (2006), Donoho (2006), Bayati & Montanari (2011), Rush & Venkataramanan (2016), Berthier, Montanari & Nguyen (2020)

Auxiliary details

Define $\zeta_k := \left(\frac{\sqrt{2}}{2} - 1\right) z_k z_k^\top W_k z_k + \sum_{i=1}^{k-1} g_i^k z_i$

$$W_k z_k + \zeta_k = \phi_k \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \frac{1}{n} I_n)$$

- conditioning on $\{z_i\}_{i<k}$, W_k is a Wigner matrix in subspace $U_{k-1}^\perp$
- W_k is independent of $\{W_i z_i\}_{i<k}$ and x_k and z_k only depend on $\{W_i z_i\}_{i<k}$
- $W_k z_k$ has zero variance along the directions of $\{z_i\}_{i<k}$ and $\frac{2}{n}$ variance along the direction of z_k