Low-rank Matrix Estimation via Approximate Message Passing

Ramji Venkataramanan
Department of Engineering

(Joint work with Andrea Montanari)

October 31, 2018

CCIMI Seminar

Symmetric Low-rank Model

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \quad \in \mathbb{R}^{n \times n}$$

- $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ are deterministic scalars
- $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ are orthonormal vectors ("spikes")
- $\boldsymbol{W}$ is a symmetric noise matrix

GOAL: To estimate the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ from $\boldsymbol{A}$

Rectangular Low-rank model

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \quad \in \mathbb{R}^{m \times n}$$

- $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ are deterministic scalars
- $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k \in \mathbb{R}^m$ are left singular vectors
  $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ are right singular vectors
- $\boldsymbol{W}$ is a noise matrix

GOAL: Estimate the singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$

Applications

$$\underbrace{\mathbf{A}}_{n \times d} \approx \underbrace{\mathbf{U}}_{n \times k} \; \overbrace{\mathbf{V}^{\mathsf{T}}}^{k \times d}$$

Topic Modelling

- Each row of $\boldsymbol{A}$ is a document
- Each row of $\boldsymbol{V}^{\mathsf{T}}$ is a topic
- Each document convex combination of k topics

[Blei, Ng, Jordan '03]

Applications

$$\underset{n \times d}{\mathbf{A}} \approx \underset{n \times k}{\mathbf{U}} \; \overset{k \times d}{\mathbf{V}^{\mathsf{T}}}$$

Collaborative filtering

- $\boldsymbol{A}$ contains ratings of users for items (e.g, films or books)
- Rows represent users, columns represent items
- Each rating is a combination of weights corresponding to a small number of factors

Hidden clique

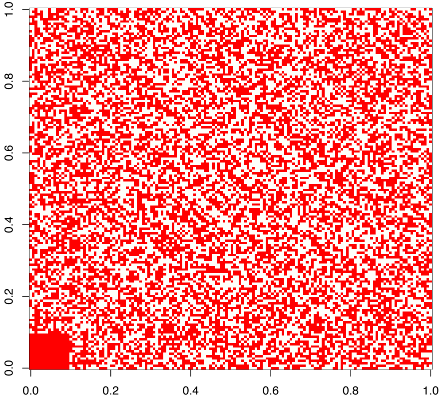


Image from *Statistical Estimation: From Denoising to Sparse Regression and Hidden Cliques* by A. Montanari

[Alon, Krivelivich, Sudakov '98], . . .

Hidden clique

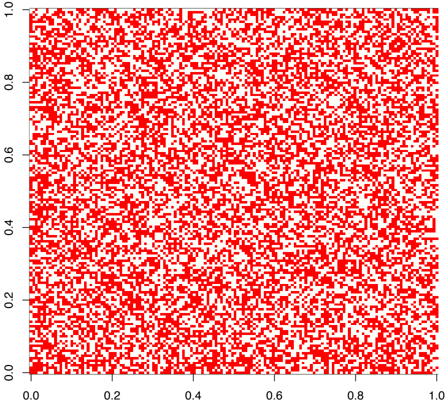


Image from *Statistical Estimation: From Denoising to Sparse Regression and Hidden Cliques* by A. Montanari

[Alon, Krivelivich, Sudakov '98], . . .

Hidden clique

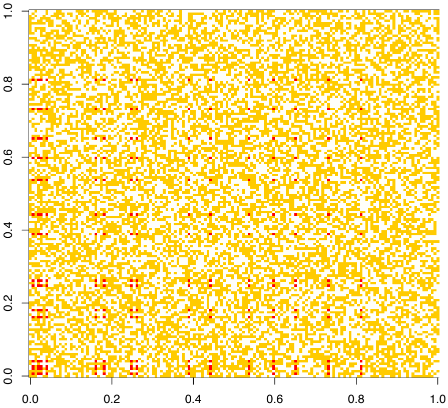
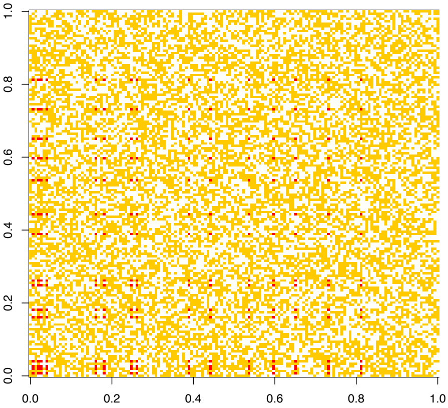


Image from *Statistical Estimation: From Denoising to Sparse Regression and Hidden Cliques* by A. Montanari

[Alon, Krivelivich, Sudakov '98], . . .

Hidden clique



For hidden clique S, adjacency matrix has the form

$$\boldsymbol{A} = \mathbf{1}_S \mathbf{1}_S^{\mathsf{T}} + \boldsymbol{W}$$

[Alon, Krivelivich, Sudakov '98], ...

Symmetric Spiked Model

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \quad \in \mathbb{R}^{n \times n}$$

- $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ are deterministic scalars
- $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ are orthonormal vectors ("spikes")
- $\boldsymbol{W} \sim \mathrm{GOE}(n) \quad \Rightarrow \quad \boldsymbol{W}$ symmetric with $(W_{ii})_{i \leq n} \sim_{i.i.d.} \mathrm{N}(0, \frac{2}{n})$ and $(W_{ij})_{i<j \leq n} \sim_{i.i.d.} \mathrm{N}(0, \frac{1}{n})$

Spectrum of spiked matrix

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^\mathsf{T} + \boldsymbol{W}$$

Random matrix theory and the 'BBAP' phase transition :

- Bulk of eigenvalues of $\boldsymbol{A}$ in $[-2, 2]$ distributed according to Wigner's semicircle
- Outlier eigenvalues corresponding to $|\lambda_i|$'s greater than 1:

$$z_i \to \lambda_i + \frac{1}{\lambda_i} > 2$$

- Eigenvectors $\boldsymbol{\varphi}_i$ corresponding to outliers z_i satisfy

$$|\langle \boldsymbol{\varphi}_i, \boldsymbol{v}_i \rangle| \to \sqrt{1 - \frac{1}{\lambda_i^2}}$$

[Baik, Ben Arous, Péché '05], [Baik, Silverstein '06], [Capitaine, Donati-Martin, Féral '09], [Benaych-Georges and Nadakuditi '11], . . .

Structural information

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W}$$

When $\boldsymbol{v}_i$'s are unstructured, e.g., drawn uniformly at random from the unit sphere,

- Best estimator of $\boldsymbol{v}_i$ is the ith eigenvector $\boldsymbol{\varphi}_i$
- If $|\lambda_i| \geq 1$, then $|\langle \boldsymbol{v}_i, \boldsymbol{\varphi}_i \rangle| \to \sqrt{1 - \frac{1}{\lambda_i^2}}$

Structural information

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W}$$

When $\boldsymbol{v}_i$'s are unstructured, e.g., drawn uniformly at random from the unit sphere,

- Best estimator of $\boldsymbol{v}_i$ is the ith eigenvector $\boldsymbol{\varphi}_i$
- If $|\lambda_i| \geq 1$, then $|\langle \boldsymbol{v}_i, \boldsymbol{\varphi}_i \rangle| \rightarrow \sqrt{1 - \frac{1}{\lambda_i^2}}$

But we often have *structural* information about $\boldsymbol{v}_i$'s

- For example, $\boldsymbol{v}_i$'s may be sparse, bounded, non-negative etc.
- Relevant in sparse PCA, non-negative PCA, hidden clique, community detection under stochastic block model, ...
- Can improve on spectral methods

Prior on eigenvectors

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \equiv \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}$$

$$\boldsymbol{V} = [\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ldots \boldsymbol{v}_k] \quad \mathbb{R}^{n\times k}$$

If each row of $\boldsymbol{V}$ is $\sim_{i.i.d} P_{\underline{V}}$, then Bayes-optimal estimator (for squared error loss) is

$$\widehat{\boldsymbol{V}}_{\mathrm{Bayes}} = \mathbb{E}\,[\boldsymbol{V} \mid \boldsymbol{A}]$$

- Generally not computable
- Closed-form expressions for asymptotic Bayes risk

[Deshpande, Montanari '14], [Barbier *et al.* '16], [Lesieur *et al.* '17], [Miolane, Lelarge '16] . . .

Computable estimators

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \equiv \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}$$

- Convex relaxations generally do not achieve Bayes-optimal performance [Javanmard, Montanari, Ricci-Tersinghi '16]
- MCMC can approximate Bayes estimator, but can have large mixing time and hard to analyze

Computable estimators

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \equiv \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}$$

- Convex relaxations generally do not achieve Bayes-optimal performance [Javanmard, Montanari, Ricci-Tersinghi '16]
- MCMC can approximate Bayes estimator, but can have large mixing time and hard to analyze

In this talk

Approximate Message Passing (AMP) algorithm to estimate $\boldsymbol{V}$

Rank one spiked model

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}, \qquad \boldsymbol{v} \sim_{i.i.d.} P_V, \quad \mathbb{E}V^2 = 1$$

Power iteration for principal eigenvector:

$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t$, with $\boldsymbol{x}^0$ chosen at random

Rank one spiked model

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}, \qquad \boldsymbol{v} \sim_{i.i.d.} P_V, \quad \mathbb{E}V^2 = 1$$

Power iteration for principal eigenvector:
$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t$, with $\boldsymbol{x}^0$ chosen at random

AMP:

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) - \mathsf{b}_t f_{t-1}(\boldsymbol{x}^{t-1}), \qquad \mathsf{b}_t = \frac{1}{n}\sum_{i=1}^{n} f_t'(x_i^t)$$

- Non-linear function $f_t : \mathbb{R} \to \mathbb{R}$ chosen based on structural info on $\boldsymbol{v}$
- Memory term ensures a nice distributional property for the iterates in high dimensions
- Can be derived via approximation of belief propagation equations

State evolution

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) - \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1}), \qquad \text{with } \mathrm{b}_t = \frac{1}{n}\sum_{i=1}^{n} f_t'(x_i^t)$$

If we initialize with $\boldsymbol{x}^0$ *independent* of $\boldsymbol{A}$, then as $n \to \infty$:

$$\boldsymbol{x}^t \;\longrightarrow\; \mu_t \boldsymbol{v} + \sigma_t \mathbf{g}$$

- $\mathbf{g} \sim_{i.i.d.} \mathrm{N}(0,1)$, independent of $\boldsymbol{v} \sim_{i.i.d.} P_V$

[Bayati,Montanari '11], [Rangan, Fletcher '12], [Deshpande, Montanari '14]

State evolution

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) - \mathsf{b}_t f_{t-1}(\boldsymbol{x}^{t-1}), \qquad \text{with } \mathsf{b}_t = \frac{1}{n}\sum_{i=1}^{n} f_t'(x_i^t)$$

If we initialize with $\boldsymbol{x}^0$ *independent* of $\boldsymbol{A}$, then as $n \to \infty$:

$$\boldsymbol{x}^t \longrightarrow \mu_t \boldsymbol{v} + \sigma_t \mathbf{g}$$

- $\mathbf{g} \sim_{i.i.d.} \mathsf{N}(0,1)$, independent of $\boldsymbol{v} \sim_{i.i.d.} P_V$
- Scalars μ_t, σ_t^2 recursively determined as

$$\mu_{t+1} = \lambda\, \mathbb{E}[V\, f_t(\mu_t V + \sigma_t G)], \quad \sigma_{t+1}^2 = \mathbb{E}[f_t(\mu_t V + \sigma_t G)^2]$$

- Initialize with $\mu_0 = \frac{1}{n}|\mathbb{E}\langle \boldsymbol{x}^0, \boldsymbol{v}\rangle|,\ \sigma_0^2 = \mathbb{E}V^2 - \mu_0^2$

[Bayati,Montanari '11], [Rangan, Fletcher '12], [Deshpande, Montanari '14]

Bayes-optimal AMP

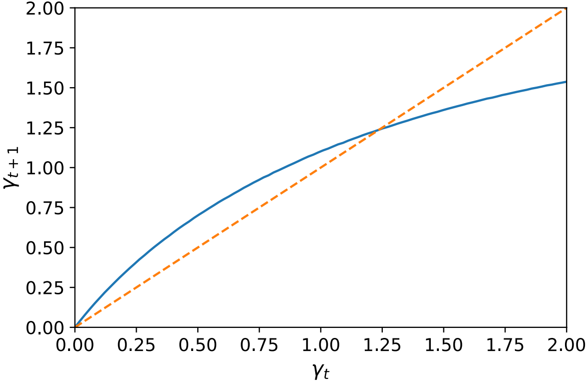
Assuming $\boldsymbol{x}^t = \mu_t \boldsymbol{v} + \sigma_t \mathbf{g}$, choose $f_t(y) = \lambda\, \mathbb{E}[V \mid \mu_t V + \sigma_t G = y]$

Bayes-optimal AMP

Assuming $\boldsymbol{x}^t = \mu_t \boldsymbol{v} + \sigma_t \mathbf{g}$, choose $f_t(y) = \lambda\,\mathbb{E}[V \mid \mu_t V + \sigma_t G = y]$

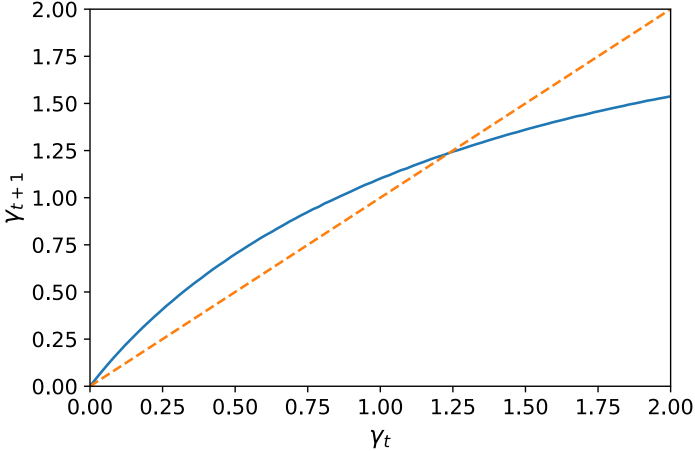
State evolution becomes $\gamma_{t+1} = \lambda^2\{1 - \mathrm{mmse}(\gamma_t)\}$ with $\mu_t = \sigma_t^2 = \gamma_t$

$P_V \sim \text{uniform}\{1, -1\}, \quad \lambda = \sqrt{2}$



Initial value $\gamma_0 \propto \frac{1}{n}|\mathbb{E}\langle \boldsymbol{x}^0, \boldsymbol{v}\rangle|$, what is $\lim_{t\to\infty} \gamma_t$?

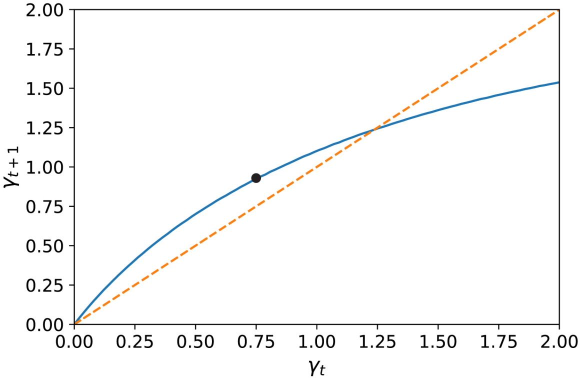
Fixed points of state evolution



- If $\mathbb{E}\langle \boldsymbol{x}^0, \boldsymbol{v} \rangle = 0$, then $\gamma_t = 0$ is an (unstable) fixed point.
- This is the case when $\boldsymbol{v}$ has zero mean, as $\boldsymbol{x}^0$ is independent of $\boldsymbol{v}$

Spectral Initialization

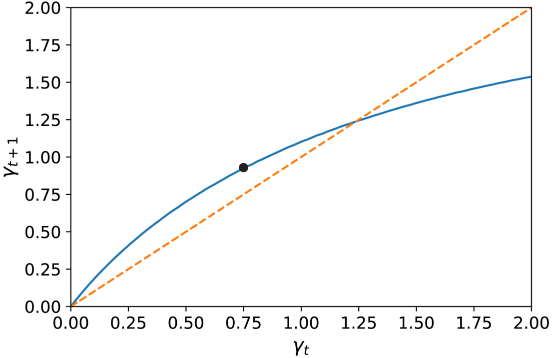
$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}, \qquad \lambda > 1$$



- Compute φ_1, the principal eigenvector of $\boldsymbol{A}$
- Run AMP with initialization $\boldsymbol{x}^0 = \sqrt{n}\varphi_1$
- $\gamma_0 > 0$ as $\frac{1}{n}|\mathbb{E}\langle \boldsymbol{x}^0, \boldsymbol{v}\rangle| \to \sqrt{1-\lambda^{-2}}$

AMP with spectral initialization

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$



$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) \,-\, \mathsf{b}_t f_{t-1}(\boldsymbol{x}^{t-1}), \qquad \boldsymbol{x}^0 = \sqrt{n}\boldsymbol{\varphi}_1$$

Existing AMP analysis does not apply for initialization $\boldsymbol{x}^0$ *correlated* with $\boldsymbol{v}$

Standard AMP analysis

With $\boldsymbol{W} \sim \mathrm{GOE}(n)$, consider

$$\boldsymbol{h}^{t+1} = \boldsymbol{W}\, f_t(\boldsymbol{h}^t) \,-\, \mathrm{b}_t f_{t-1}(\boldsymbol{h}^{t-1})$$

Initialised with $\boldsymbol{h}^0$ *independent* of $\boldsymbol{W}$. Let $\vartheta_t := \{\boldsymbol{h}^0, \dots, \boldsymbol{h}^t\}$

[Bolthausen '10], [Bayati-Montanari '11], [Rush-Venkataramanan '16]

Standard AMP analysis

With $\boldsymbol{W} \sim \mathrm{GOE}(n)$, consider

$$\boldsymbol{h}^{t+1} = \boldsymbol{W} f_t(\boldsymbol{h}^t) - \mathrm{b}_t f_{t-1}(\boldsymbol{h}^{t-1})$$

Initialised with $\boldsymbol{h}^0$ *independent* of $\boldsymbol{W}$. Let $\vartheta_t := \{\boldsymbol{h}^0, \ldots, \boldsymbol{h}^t\}$

- Conditional distribution

$$\boldsymbol{W}|_{\vartheta_t} \stackrel{d}{=} \mathbb{E}[\boldsymbol{W} \mid \vartheta_t] + \boldsymbol{P}_{\vartheta_t}^{\perp} \tilde{\boldsymbol{W}} \boldsymbol{P}_{\vartheta_t}^{\perp}$$

[Bolthausen '10], [Bayati-Montanari '11], [Rush-Venkataramanan '16]

Standard AMP analysis

With $\boldsymbol{W} \sim \mathrm{GOE}(n)$, consider

$$\boldsymbol{h}^{t+1} = \boldsymbol{W}\, f_t(\boldsymbol{h}^t) - \mathrm{b}_t f_{t-1}(\boldsymbol{h}^{t-1})$$

Initialised with $\boldsymbol{h}^0$ *independent* of $\boldsymbol{W}$. Let $\vartheta_t := \{\boldsymbol{h}^0, \dots, \boldsymbol{h}^t\}$

- Conditional distribution

$$\boldsymbol{W}|_{\vartheta_t} \overset{d}{=} \mathbb{E}[\boldsymbol{W} \mid \vartheta_t] + \boldsymbol{P}^{\perp}_{\vartheta_t} \tilde{\boldsymbol{W}} \boldsymbol{P}^{\perp}_{\vartheta_t}$$

- By induction, show that for $t \geq 0$:

$$\boldsymbol{h}^{t+1} = \sum_{i=0}^{t} \alpha_i \boldsymbol{h}^i + \boldsymbol{g}_t + \boldsymbol{\Delta}_t$$

[Bolthausen '10], [Bayati-Montanari '11], [Rush-Venkataramanan '16]

Standard AMP analysis

With $\boldsymbol{W} \sim \mathrm{GOE}(n)$, consider

$$\boldsymbol{h}^{t+1} = \boldsymbol{W}\, f_t(\boldsymbol{h}^t) \,-\, \mathrm{b}_t f_{t-1}(\boldsymbol{h}^{t-1})$$

Initialised with $\boldsymbol{h}^0$ *independent* of $\boldsymbol{W}$. Let $\vartheta_t := \{\boldsymbol{h}^0, \dots, \boldsymbol{h}^t\}$

- Conditional distribution

$$\boldsymbol{W}|_{\vartheta_t} \overset{d}{=} \mathbb{E}[\boldsymbol{W} \mid \vartheta_t] + \boldsymbol{P}_{\vartheta_t}^{\perp} \tilde{\boldsymbol{W}} \boldsymbol{P}_{\vartheta_t}^{\perp}$$

- By induction, show that for $t \geq 0$:

$$\boldsymbol{h}^{t+1} = \sum_{i=0}^{t} \alpha_i \boldsymbol{h}^i + \boldsymbol{g}_t + \boldsymbol{\Delta}_t$$

$$\boldsymbol{h}^{t+1} \overset{d}{\approx} \tau_t \boldsymbol{g} \qquad \tau_t^2 = \mathbb{E}[f_t(\tau_{t-1} G)^2], \quad \tau_0^2 = \|f(\boldsymbol{h}^0)\|^2/n$$

[Bolthausen '10], [Bayati-Montanari '11], [Rush-Venkataramanan '16]

AMP with spectral initialization

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let (φ_1, z_1) be principal eigenvector and eigenvalue of $\boldsymbol{A}$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) \;-\; \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

initialised with $\boldsymbol{x}^0 = \sqrt{n}\,\varphi_1$

AMP with spectral initialization

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let (φ_1, z_1) be principal eigenvector and eigenvalue of $\boldsymbol{A}$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) \, - \, \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

initialised with $\boldsymbol{x}^0 = \sqrt{n}\,\varphi_1$

We write

$$\boldsymbol{A} = z_1\varphi_1\varphi_1^{\mathsf{T}} + \boldsymbol{P}^{\perp}\left(\frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}\right)\boldsymbol{P}^{\perp}$$

- $\boldsymbol{P}^{\perp} = \boldsymbol{I} - \varphi_1\varphi_1^{\mathsf{T}}$

AMP with spectral initialization

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let (φ_1, z_1) be principal eigenvector and eigenvalue of $\boldsymbol{A}$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) \;-\; \mathsf{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

initialised with $\boldsymbol{x}^0 = \sqrt{n}\,\varphi_1$

Instead of $\boldsymbol{A}$, we will analyze AMP on

$$\tilde{\boldsymbol{A}} = z_1\varphi_1\varphi_1^{\mathsf{T}} + \boldsymbol{P}^{\perp}\left(\frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \tilde{\boldsymbol{W}}\right)\boldsymbol{P}^{\perp}$$

- $\boldsymbol{P}^{\perp} = \boldsymbol{I} - \varphi_1\varphi_1^{\mathsf{T}}$
- $\tilde{\boldsymbol{W}} \sim \mathrm{GOE}(n)$ is independent of $\boldsymbol{W}$

1. Conditioned on z_1 and $(\varphi_1^{\mathsf{T}}\boldsymbol{v})^2$ being close to limiting values, total variation distance between $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ is small

AMP with spectral initialization

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let $(\boldsymbol{\varphi}_1, z_1)$ be principal eigenvector and eigenvalue of $\boldsymbol{A}$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) \; - \; \mathsf{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

initialised with $\boldsymbol{x}^0 = \sqrt{n}\,\boldsymbol{\varphi}_1$

Instead of $\boldsymbol{A}$, we will analyze AMP on

$$\tilde{\boldsymbol{A}} = z_1 \boldsymbol{\varphi}_1 \boldsymbol{\varphi}_1^{\mathsf{T}} + \boldsymbol{P}^{\perp} \left(\frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \tilde{\boldsymbol{W}} \right) \boldsymbol{P}^{\perp}$$

- $\boldsymbol{P}^{\perp} = \boldsymbol{I} - \boldsymbol{\varphi}_1 \boldsymbol{\varphi}_1^{\mathsf{T}}$
- $\tilde{\boldsymbol{W}} \sim \mathrm{GOE}(n)$ is independent of $\boldsymbol{W}$

1. Conditioned on z_1 and $(\boldsymbol{\varphi}_1^{\mathsf{T}}\boldsymbol{v})^2$ being close to limiting values, total variation distance between $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ is small
2. Analyze AMP on $\tilde{\boldsymbol{A}}$ by extending standard AMP analysis

Model assumptions

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let $\boldsymbol{v} = \boldsymbol{v}(n) \in \mathbb{R}^n$ be a sequence such that the empirical distribution of entries of $\boldsymbol{v}(n)$ converges weakly to P_V,

Model assumptions

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let $\boldsymbol{v} = \boldsymbol{v}(n) \in \mathbb{R}^n$ be a sequence such that the empirical distribution of entries of $\boldsymbol{v}(n)$ converges weakly to P_V,

Performance of any estimator $\hat{\boldsymbol{v}}$ measured via loss function $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$:

$$\psi(\boldsymbol{v}, \hat{\boldsymbol{v}}) = \frac{1}{n}\sum_{i=1}^{n} \psi(v_i, \hat{v}_i)$$

ψ assumed to be *pseudo-Lipschitz*:

$$|\psi(\boldsymbol{x}) - \psi(\boldsymbol{y})| \le C\|\boldsymbol{x} - \boldsymbol{y}\|_2\,(1 + \|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2), \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$$

L_2 loss, L_1 loss are both pseudo-Lipschitz

Result for rank one case

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Theorem: Let $\lambda > 1$. Consider the AMP

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) \;-\; \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

- Assume $f_t : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous
- Initialize with $\boldsymbol{x}^0 = \sqrt{n}\boldsymbol{\varphi}_1$

Then for any pseudo-Lipschitz loss function ψ and $t \geq 0$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(v_i, x_i^t) = \mathbb{E}\left\{\psi(V, \mu_t V + \sigma_t G)\right\} \quad \text{a.s.}$$

Result for rank one case

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Theorem: Let $\lambda > 1$. Consider the AMP

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) \; - \; \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

- Assume $f_t : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous
- Initialize with $\boldsymbol{x}^0 = \sqrt{n}\boldsymbol{\varphi}_1$

Then for any pseudo-Lipschitz loss function ψ and $t \geq 0$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(v_i, x_i^t) = \mathbb{E}\left\{\psi(V, \mu_t V + \sigma_t G)\right\} \quad \text{a.s.}$$

State evolution parameters: $\mu_0 = \sqrt{1-\lambda^{-2}}, \quad \sigma_0 = 1/\lambda,$

$$\mu_{t+1} = \lambda\, \mathbb{E}[V\, f_t(\mu_t V + \sigma_t G)], \quad \sigma_{t+1}^2 = \mathbb{E}[f_t(\mu_t V + \sigma_t G)^2],$$

Proof Sketch

True vs conditional model

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

$$\tilde{\boldsymbol{A}} = z_1\,\boldsymbol{\varphi}_1\boldsymbol{\varphi}_1^{\mathsf{T}} + \boldsymbol{P}^{\perp}\left(\frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \tilde{\boldsymbol{W}}\right)\boldsymbol{P}^{\perp}$$

Lemma

For
$$(z_1, \boldsymbol{\varphi}_1) \in \left\{|z_1 - (\lambda + \lambda^{-1})| \le \varepsilon, \quad (\boldsymbol{\varphi}_1^{\mathsf{T}}\boldsymbol{v})^2 \ge 1 - \lambda^{-2} - \varepsilon\right\},$$

we have

$$\sup_{(\boldsymbol{z}_{\hat{S}}, \boldsymbol{\Phi}_{\hat{S}}) \in \mathcal{E}_{\varepsilon}} \left\|\mathbb{P}\big(\boldsymbol{A} \in \cdot \,\big|\, z_1, \boldsymbol{\varphi}_1\big) - \mathbb{P}\big(\tilde{\boldsymbol{A}} \in \cdot \,\big|\, z_1, \boldsymbol{\varphi}_1\big)\right\|_{\mathrm{TV}} \le \frac{1}{c(\varepsilon)}\, e^{-nc(\varepsilon)}$$

AMP on conditional model

$$\tilde{\boldsymbol{A}} = z_1 \varphi_1 \varphi_1^{\mathsf{T}} + \boldsymbol{P}^{\perp} \left(\frac{\lambda}{n} \boldsymbol{v} \boldsymbol{v}^{\mathsf{T}} + \tilde{\boldsymbol{W}} \right) \boldsymbol{P}^{\perp}$$

AMP with $\tilde{\boldsymbol{A}}$ instead of $\boldsymbol{A}$:

$$\tilde{\boldsymbol{x}}^{t+1} = \tilde{\boldsymbol{A}} f(\tilde{\boldsymbol{x}}^t; t) - \mathrm{b}_t f(\tilde{\boldsymbol{x}}^{t-1}; t-1), \qquad \tilde{\boldsymbol{x}}^0 = \sqrt{n}\, \varphi_1$$

Analyze using existing AMP analysis + results from random matrix theory

Bayes-optimal AMP

$$\boldsymbol{A} = \frac{\lambda}{n} \boldsymbol{v} \boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A} \, f_t(\boldsymbol{x}^t) \; - \; \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

- Bayes-optimal choice $f_t(y) = \lambda \, \mathbb{E}(V \mid \gamma_t \, V + \sqrt{\gamma_t} \, G = y)$
- State evolution:

$$\gamma_{t+1} = \lambda^2 \{1 - \mathrm{mmse}(\gamma_t)\}, \qquad \gamma_0 = \lambda^2 - 1$$

where $\mathrm{mmse}(\gamma) = \mathbb{E}\{ [V - \mathbb{E}(V \mid \sqrt{\gamma} \, V + G)]^2 \}$

- $\mu_t = \sigma_t^2 = \gamma_t$

Bayes-optimal AMP

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let $\gamma_{\mathrm{AMP}}(\lambda)$ be the *smallest* strictly positive solution of

$$\gamma = \lambda^2[1 - \mathrm{mmse}(\gamma)]. \tag{1}$$

Then the AMP estimate $\hat{\boldsymbol{x}}^t = f_t(\boldsymbol{x}^t)$ achieves

$$\lim_{t\to\infty} \lim_{n\to\infty} \min_{s\in\{+1,-1\}} \frac{1}{n}\|\hat{\boldsymbol{x}}^t - s\boldsymbol{v}\|_2^2 = 1 - \frac{\gamma_{\mathrm{AMP}}(\lambda)}{\lambda^2}$$

Bayes-optimal AMP

$$\boldsymbol{A} = \frac{\lambda}{n} \boldsymbol{v} \boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let $\gamma_{\text{AMP}}(\lambda)$ be the *smallest* strictly positive solution of

$$\gamma = \lambda^2 [1 - \mathsf{mmse}(\gamma)]. \tag{1}$$

Then the AMP estimate $\hat{\boldsymbol{x}}^t = f_t(\boldsymbol{x}^t)$ achieves

$$\text{Overlap}: \quad \lim_{t \to \infty} \lim_{n \to \infty} \frac{|\langle \hat{\boldsymbol{x}}^t, \boldsymbol{v} \rangle|}{\|\hat{\boldsymbol{x}}^t\|_2 \|\boldsymbol{v}\|_2} = \frac{\sqrt{\gamma_{\text{AMP}}(\lambda)}}{\lambda}$$

Bayes-optimal AMP

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

Let $\gamma_{\mathrm{AMP}}(\lambda)$ be the *smallest* strictly positive solution of

$$\gamma = \lambda^2[1 - \mathsf{mmse}(\gamma)]. \qquad (1)$$

Then the AMP estimate $\hat{\boldsymbol{x}}^t = f_t(\boldsymbol{x}^t)$ achieves

$$\text{Overlap :} \quad \lim_{t\to\infty}\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}^t, \boldsymbol{v}\rangle|}{\|\hat{\boldsymbol{x}}^t\|_2\|\boldsymbol{v}\|_2} = \frac{\sqrt{\gamma_{\mathrm{AMP}}(\lambda)}}{\lambda}$$

Bayes-optimal overlap [Miolane-Lelarge '16]

For (almost) all $\lambda > 0$

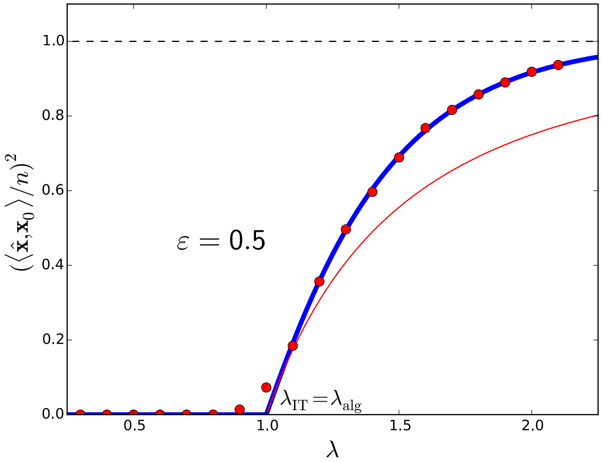
$$\lim_{n\to\infty}\sup_{\hat{\boldsymbol{x}}(\,\cdot\,)} \frac{|\langle \hat{\boldsymbol{x}}^t, \boldsymbol{v}\rangle|}{\|\hat{\boldsymbol{x}}^t\|_2\|\boldsymbol{v}\|_2} = \frac{\sqrt{\gamma_{\mathrm{Bayes}}(\lambda)}}{\lambda}$$

$\gamma_{\mathrm{Bayes}}(\lambda)$: fixed point of (1) that maximizes a free-energy functional

Example: Two-point mixture

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

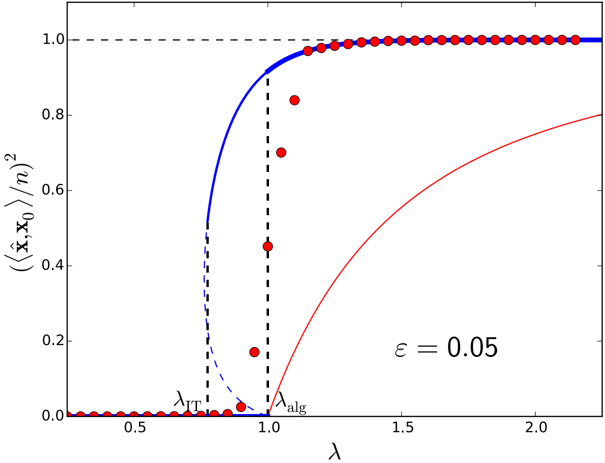
$$P_V = \varepsilon\,\delta_{a_+} + (1-\varepsilon)\delta_{a_-} \qquad a_+ = \sqrt{\frac{1-\varepsilon}{\varepsilon}} \quad a_- = -\sqrt{\frac{\varepsilon}{1-\varepsilon}}\,.$$



Example: Two-point mixture

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

$$P_V = \varepsilon\,\delta_{a_+} + (1-\varepsilon)\delta_{a_-} \qquad a_+ = \sqrt{\frac{1-\varepsilon}{\varepsilon}} \quad a_- = -\sqrt{\frac{\varepsilon}{1-\varepsilon}}\,.$$



Confidence intervals

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

$$\text{AMP:} \quad \boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) - \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

- Convergence result tells us that $\boldsymbol{x}^t \approx \mu_t \boldsymbol{v} + \sigma_t \boldsymbol{g}$

Confidence intervals

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

$$\text{AMP:}\quad \boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) - \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

- Convergence result tells us that $\boldsymbol{x}^t \approx \mu_t \boldsymbol{v} + \sigma_t \boldsymbol{g}$
- State evolution parameters can be estimated:

$$\hat{\sigma}_t^2 \equiv \frac{1}{n}\big\| f_{t-1}(\boldsymbol{x}^{t-1})\big\|_2^2 ,$$
$$\hat{\mu}_t^2 \equiv \frac{1}{n}\big\|\boldsymbol{x}^t\big\|_2^2 - \frac{1}{n}\big\| f_{t-1}(\boldsymbol{x}^{t-1})\big\|_2^2 .$$

Confidence intervals

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} + \boldsymbol{W}$$

$$\text{AMP:} \quad \boldsymbol{x}^{t+1} = \boldsymbol{A}\, f_t(\boldsymbol{x}^t) - \mathrm{b}_t f_{t-1}(\boldsymbol{x}^{t-1})$$

- Convergence result tells us that $\boldsymbol{x}^t \approx \mu_t \boldsymbol{v} + \sigma_t \boldsymbol{g}$
- State evolution parameters can be estimated:

$$\hat{\sigma}_t^2 \equiv \frac{1}{n}\big\|f_{t-1}(\boldsymbol{x}^{t-1})\big\|_2^2,$$
$$\hat{\mu}_t^2 \equiv \frac{1}{n}\big\|\boldsymbol{x}^t\big\|_2^2 - \frac{1}{n}\big\|f_{t-1}(\boldsymbol{x}^{t-1})\big\|_2^2.$$

- Confidence intervals for coverage level $(1-\alpha)$:

$$\hat{I}_i(\alpha; t) = \left[\frac{1}{\hat{\mu}_t}x_i^t - \frac{\hat{\sigma}_t}{\hat{\mu}_t}\Phi^{-1}\left(1-\frac{\alpha}{2}\right),\;\; \frac{1}{\hat{\mu}_t}x_i^t + \frac{\hat{\sigma}_t}{\hat{\mu}_t}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right]$$

- Bayes-optimal choice minimizes length of confidence intervals, but requires knowledge of the empirical distribution of $\boldsymbol{v}$

For $1 \leq i \leq n$,

$$\hat{I}_i(\alpha; t) = \left[\frac{1}{\hat{\mu}_t} x_i^t - \frac{\hat{\sigma}_t}{\hat{\mu}_t} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \quad \frac{1}{\hat{\mu}_t} x_i^t + \frac{\hat{\sigma}_t}{\hat{\mu}_t} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]$$

Corollary:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left(v_i \in \hat{I}_i(\alpha; t)\right) = 1 - \alpha \quad \text{almost surely.}$$

General case

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \equiv \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}\,.$$

- Assume k_* eigenvectors corresponding to outliers $|\lambda_i| > 1$
- Outliers can be estimated from $\boldsymbol{A}$, as $z_i \to \lambda_i + \lambda_i^{-1}$
- Assume empirical distribution of rows of $\boldsymbol{V} \sim P_{\boldsymbol{V}}$

General case

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \equiv \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W} \,.$$

- Assume k_* eigenvectors corresponding to outliers $|\lambda_i| > 1$
- Outliers can be estimated from $\boldsymbol{A}$, as $z_i \to \lambda_i + \lambda_i^{-1}$
- Assume empirical distribution of rows of $\boldsymbol{V} \sim P_{\boldsymbol{V}}$

General case

$$\boldsymbol{A} = \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathsf{T}} + \boldsymbol{W} \equiv \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}\,.$$

- Assume k_* eigenvectors corresponding to outliers $|\lambda_i| > 1$
- Outliers can be estimated from $\boldsymbol{A}$, as $z_i \to \lambda_i + \lambda_i^{-1}$
- Assume empirical distribution of rows of $\boldsymbol{V} \sim P_{\boldsymbol{V}}$

AMP : $\boldsymbol{x}^{t+1} = \boldsymbol{A} f_t(\boldsymbol{x}^t) - f_{t-1}(\boldsymbol{x}^{t-1})\,\mathrm{B}_t^{\mathsf{T}}$

- $\boldsymbol{x}^t \in \mathbb{R}^{n \times k_*}$ are estimates of the outlier eigenvectors
- $f(\cdot;t) : \mathbb{R}^{k_*} \to \mathbb{R}^{k_*}$ applied row by row
- $\mathrm{B}_t = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial f_t}{\partial \boldsymbol{x}}(\boldsymbol{x}_i^t)$, where $\frac{\partial f_t}{\partial \boldsymbol{x}}$ is Jacobian of $f(\cdot;t)$

Spectral initialization: $\boldsymbol{x}^0 = \left[\sqrt{n}\varphi_1 \mid \ldots \mid \sqrt{n}\varphi_{k_*}\right]$

Block model with multiple communities

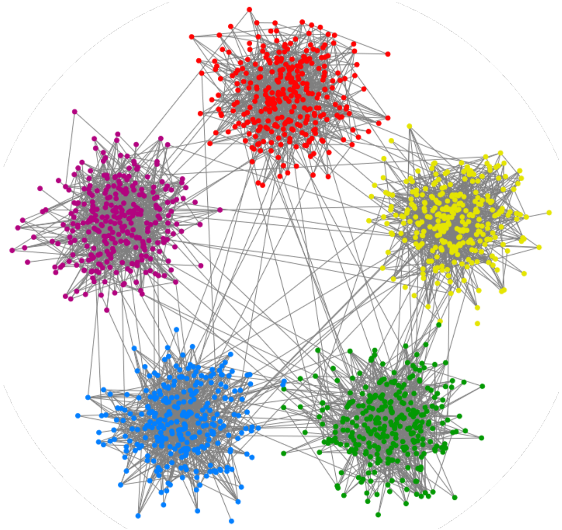
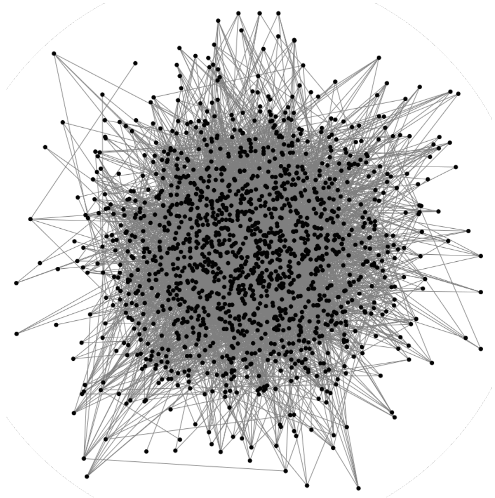


Image from *Community detection and stochastic block models* by E. Abbe

Block model with multiple communities



Wish to recover vertex labels (colours) from adjacency matrix

Image from *Community detection and stochastic block models* by E. Abbe

A closely related model . . .

- Let $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ be vector of vertex labels
- Labels σ_i uniformly distributed in $\{1, 2, 3\}$
- Consider the $n \times n$ matrix $\boldsymbol{A}_0$ with entries

$$A_{0,ij} = \begin{cases} 2/n & \text{if } \sigma_i = \sigma_j \\ -1/n & \text{otherwise.} \end{cases}$$

- $\boldsymbol{A}_0$ is an orthogonal projector onto a two-dimensional subspace $\Rightarrow \boldsymbol{A}_0$ is rank 2

A closely related model . . .

- Let $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ be vector of vertex labels
- Labels σ_i uniformly distributed in $\{1, 2, 3\}$
- Consider the $n \times n$ matrix $\boldsymbol{A}_0$ with entries

$$A_{0,ij} = \begin{cases} 2/n & \text{if } \sigma_i = \sigma_j \\ -1/n & \text{otherwise.} \end{cases}$$

- $\boldsymbol{A}_0$ is an orthogonal projector onto a two-dimensional subspace $\Rightarrow \boldsymbol{A}_0$ is rank 2

Wish to estimate $\boldsymbol{A}_0$ from noisy version:

$$\boldsymbol{A} = \lambda \boldsymbol{A}_0 + \boldsymbol{W}$$

- Degenerate eigenvalues: $\lambda_1 = \lambda_2 = \lambda > 1$
- $\boldsymbol{W} \sim \text{GOE}(n)$
- $\boldsymbol{A}$ similar to rescaled adjacency matrix in block model

AMP

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}$$

Spectral initialization: $\boldsymbol{x}^0 = [\sqrt{n}\varphi_1 \quad \sqrt{n}\varphi_2]$

Main result

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(\boldsymbol{V}_i, \boldsymbol{x}_i^t) = \mathbb{E}\{\psi(\underline{V},\ \boldsymbol{M}_t\underline{V} + \boldsymbol{Q}_t^{1/2}\underline{G})\} \quad \text{a.s.}$$

AMP

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}$$

Spectral initialization: $\boldsymbol{x}^0 = [\sqrt{n}\varphi_1 \quad \sqrt{n}\varphi_2]$

Main result

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(\boldsymbol{V}_i, \boldsymbol{x}_i^t) = \mathbb{E}\{\psi(\underline{V}, \boldsymbol{M}_t\underline{V} + \boldsymbol{Q}_t^{1/2}\underline{G})\} \quad \text{a.s.}$$

State evolution: $\boldsymbol{M}_0 = (\boldsymbol{x}^0)^{\mathsf{T}}\boldsymbol{V}$ and $\boldsymbol{Q}_0 = \lambda^{-1}\boldsymbol{I} \ \in \mathbb{R}^{2\times 2}$

$$\boldsymbol{M}_{t+1} = \lambda\mathbb{E}\Big\{f_t(\boldsymbol{M}_t\underline{V} + \boldsymbol{Q}_t^{1/2}\underline{G})\underline{V}^{\mathsf{T}}\Big\},$$

$$\boldsymbol{Q}_{t+1} = \mathbb{E}\Big\{f_t(\boldsymbol{M}_t\underline{V} + \boldsymbol{Q}_t^{1/2}\boldsymbol{G})f_t(\boldsymbol{M}_t\underline{V} + \boldsymbol{Q}_t^{1/2}\underline{G})^{\mathsf{T}}\Big\}.$$

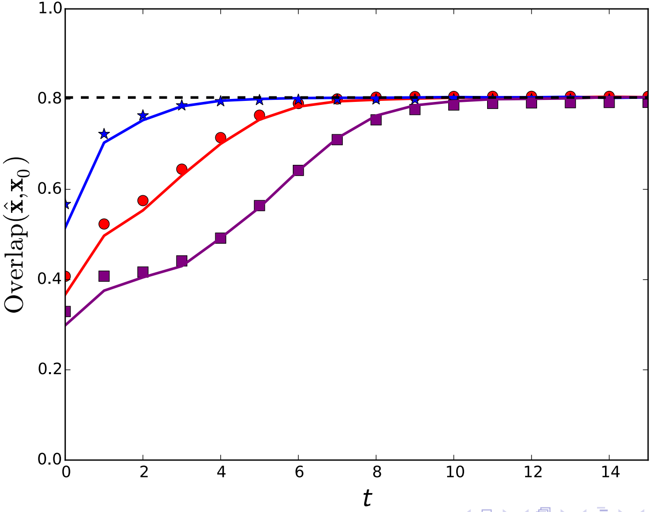
Since $\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}} = \boldsymbol{V}\boldsymbol{R}\boldsymbol{R}^{\mathsf{T}}\boldsymbol{V}^{\mathsf{T}}$ for any 2×2 rotation matrix $\boldsymbol{R}$

$\Rightarrow$ state evolution starts from a *random* initial condition

$$\boldsymbol{M}_0 = (\boldsymbol{x}^0)^{\mathsf{T}}\boldsymbol{V} \stackrel{d}{=} \sqrt{1-\lambda^{-2}}\boldsymbol{R}$$

$$\boldsymbol{A} = \frac{\lambda}{n}\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}$$

Gaussian block model with $\lambda = 1.5, \quad n = 6000$



Summary

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}$$

AMP with spectral initialization

- Distributional property of the iterates gives succinct performance characterization via state evolution
- Can be used to construct confidence intervals
- AMP can achieve Bayes-optimal accuracy

Extensions and Future work

- Can be extended to rectangular low-rank matrix model: $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{T}} + \boldsymbol{W}$
- AMP with spectral initialization for generalized linear models, e.g., phase retrieval

https://arxiv.org/abs/1711.01682