Estimation in Generalized Linear Models: From Spectral Methods to Approximate Message Passing (and back)

Ramji Venkataramanan

University of Cambridge

IT-FuN Workshop, KIT, March 2026

Marco Mondelli

Yihan Zhang

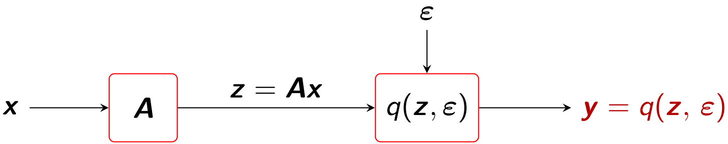
Hong Chang Ji

Generalized Linear Model

$$\text{Data } (\boldsymbol{a}_i,\ y_i) \in \mathbb{R}^d \times \mathbb{R}, \quad i = 1, \ldots n$$

$$\begin{tikzcd} & & \varepsilon_i \arrow[d] & \\ \boldsymbol{x} \arrow[r] & \boxed{\boldsymbol{a}_i} \arrow[r, "z_i = \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle"] & \boxed{q(\cdot, \cdot)} \arrow[r] & y_i = q(z_i,\ \varepsilon_i) \end{tikzcd}$$

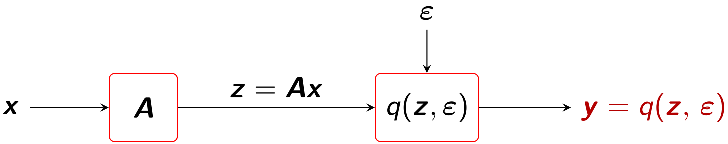
Generalized Linear Model



GOAL: Estimate signal $\boldsymbol{x} \in \mathbb{R}^d$ from observations $\boldsymbol{y} \equiv (y_1, \ldots, y_n)$

Known sensing matrix $\boldsymbol{A} \in \mathbb{R}^{n \times d}$ and output function q

Examples



- Linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon$

Examples

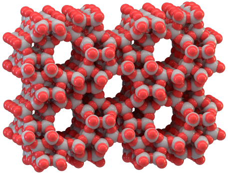
$$\begin{tikzcd} & & & \varepsilon \arrow[d] & \\ \boldsymbol{x} \arrow[r] & \boxed{\boldsymbol{A}} \arrow[rr, "\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}"] & & \boxed{q(\boldsymbol{z}, \varepsilon)} \arrow[r] & \boldsymbol{y} = q(\boldsymbol{z}, \varepsilon) \end{tikzcd}$$

- Linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon$
- 1-bit compressed sensing $\boldsymbol{y} = \mathrm{sign}(\boldsymbol{A}\boldsymbol{x} + \varepsilon)$

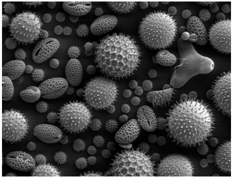
Examples

$$
\begin{tikzcd}
 & & \varepsilon \arrow[d] & \\
\boldsymbol{x} \arrow[r] & \boxed{\boldsymbol{A}} \arrow[r, "\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}"] & \boxed{q(\boldsymbol{z}, \varepsilon)} \arrow[r] & \boldsymbol{y} = q(\boldsymbol{z}, \varepsilon)
\end{tikzcd}
$$

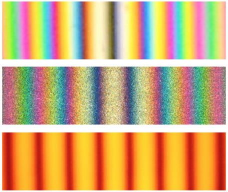
- Linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon$
- 1-bit compressed sensing $\boldsymbol{y} = \mathrm{sign}(\boldsymbol{A}\boldsymbol{x} + \varepsilon)$
- Phase retrieval $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 + \varepsilon$



X-ray crystallography



Microscopy



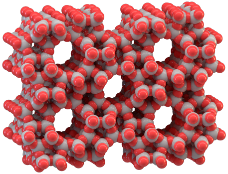
Interferometry

Examples

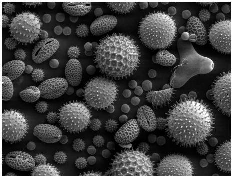
$$\boldsymbol{x} \longrightarrow \boxed{\boldsymbol{A}} \xrightarrow{\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}} \boxed{q(\boldsymbol{z}, \varepsilon)} \longrightarrow \boldsymbol{y} = q(\boldsymbol{z}, \varepsilon)$$

$$\varepsilon \downarrow q(\boldsymbol{z}, \varepsilon)$$

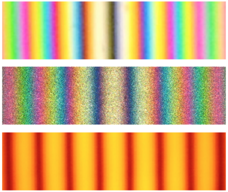
- Linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon$
- 1-bit compressed sensing $\boldsymbol{y} = \text{sign}(\boldsymbol{A}\boldsymbol{x} + \varepsilon)$
- Phase retrieval $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 + \varepsilon$



X-ray crystallography



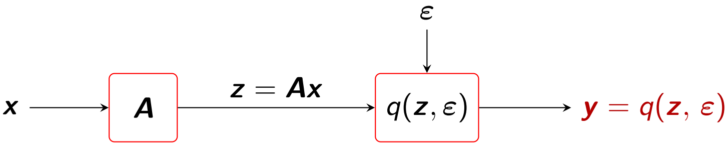
Microscopy



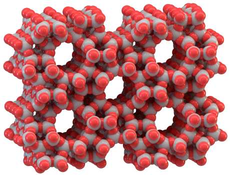
Interferometry

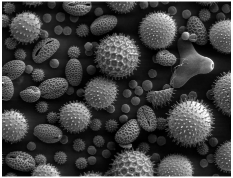
- Communication systems with non-linearities

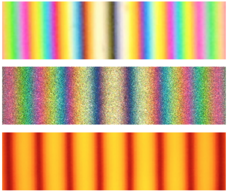
Examples



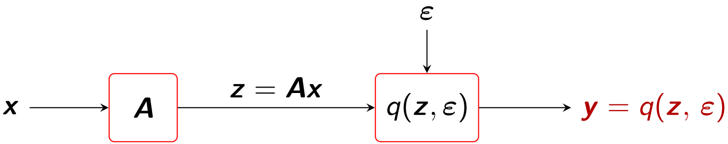
- Linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon$
- 1-bit compressed sensing $\boldsymbol{y} = \mathrm{sign}(\boldsymbol{A}\boldsymbol{x} + \varepsilon)$
- Phase retrieval $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 + \varepsilon$


X-ray crystallography


Microscopy


Interferometry

- Logistic regression, Poisson regression ...



$$\boldsymbol{A} = \begin{bmatrix} \longleftarrow & \boldsymbol{a}_1 & \longrightarrow \\ & \vdots & \\ \longleftarrow & \boldsymbol{a}_n & \longrightarrow \end{bmatrix} \in \mathbb{R}^{n \times d}$$

GOAL : Estimate $\boldsymbol{x}$ from $\boldsymbol{y}$, with $\boldsymbol{A}, q$ known

High-dimensional regime: $\frac{n}{d} \to \delta$ as $n, d \to \infty$

Estimators

- Convex relaxations
- Iterative algorithms for non-convex objectives:
  Alternating minimization, gradient descent, ...
- Approximate Message Passing (AMP)

Iterative methods typically need an informative initialization

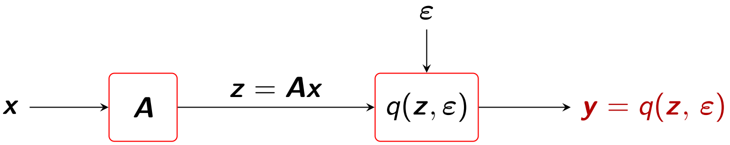
Spectral methods often effective

Phase retrieval: [Netrapalli et al. '13], [Candes et al. '13], [Luo et al. '19], [Mondelli & Montanari '19], ...

1-bit CS: [Plan & Vershynin '13], [Jacques et al. '13], ...

AMP for GLMs: [Rangan '11], [Javanmard & Montanari '13], ...

Weak recovery problem



Find $\boldsymbol{x}$ such that **overlap** $\frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x} \rangle|}{\|\hat{\boldsymbol{x}}\| \cdot \|\boldsymbol{x}\|} > \epsilon$, for some $\epsilon > 0$

Weak recovery problem

$$\boldsymbol{x} \longrightarrow \boxed{\boldsymbol{A}} \xrightarrow{\ \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}\ } \boxed{q(\boldsymbol{z}, \boldsymbol{\varepsilon})} \longrightarrow \boldsymbol{y} = q(\boldsymbol{z}, \boldsymbol{\varepsilon})$$

$$\boldsymbol{\varepsilon} \downarrow$$

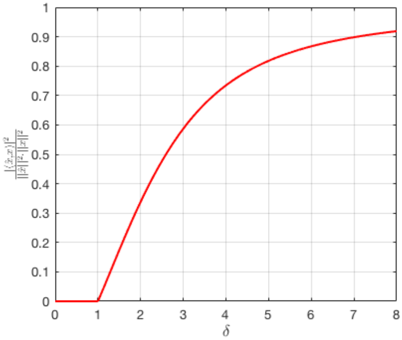
Find $\boldsymbol{x}$ such that **overlap** $\frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x} \rangle|}{\|\hat{\boldsymbol{x}}\| \cdot \|\boldsymbol{x}\|} > \epsilon$, for some $\epsilon > 0$

Example: Phase Retrieval

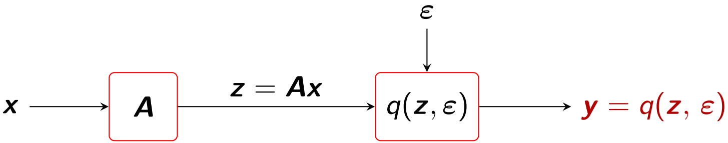
Overlap of spectral estimator

$n, d \to \infty$ with $n/d \to \delta$

Phase transition



Spectral estimator



Rows of $\boldsymbol{A}$ are $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n \in \mathbb{R}^d$. Let

$$\boldsymbol{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{T}}$$

$\mathcal{T} : \mathbb{R} \to \mathbb{R}$ pre-processing function.

Spectral Estimator $\hat{\boldsymbol{x}}^{\mathrm{s}} =$ principal eigenvector of $\boldsymbol{D}_n$

Precise asymptotics for i.i.d. Gaussian design

$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^\top$$

Theorem [Lu-Li '19], [Mondelli-Montanari '19]

Assume $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \sim_{\text{i.i.d.}} \mathrm{N}(0, \boldsymbol{I}_d/d)$. Let $\mathcal{T}$ be Lipschitz and satisfy some mild regularity conditions. Then, as $n, d \to \infty$ with $n/d \to \delta$,

$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x}\rangle|}{\|\hat{\boldsymbol{x}}^{\mathrm{s}}\| \, \|\boldsymbol{x}\|} = \begin{cases} \rho(\mathcal{T}, \delta) > 0 & \text{if } \delta > \delta^*(\mathcal{T}), \\ 0 & \text{otherwise} \end{cases}$$

Precise asymptotics for i.i.d. Gaussian design

$$\boldsymbol{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^\top$$

Theorem [Lu-Li '19], [Mondelli-Montanari '19]

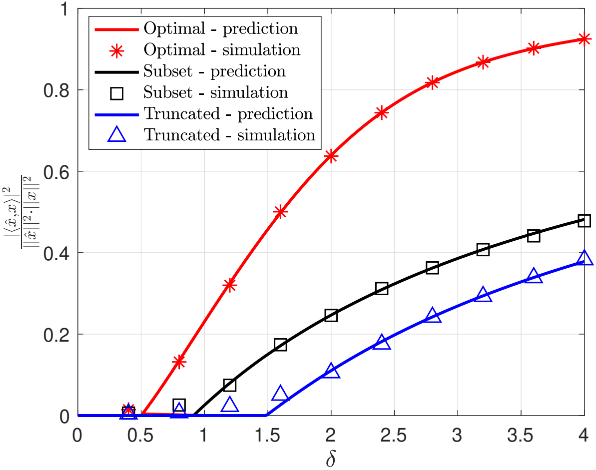
Assume $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \sim_{\text{i.i.d.}} \mathrm{N}(0, \boldsymbol{I}_d/d)$. Let $\mathcal{T}$ be Lipschitz and satisfy some mild regularity conditions. Then, as $n, d \to \infty$ with $n/d \to \delta$,

$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x} \rangle|}{\|\hat{\boldsymbol{x}}^{\mathrm{s}}\| \, \|\boldsymbol{x}\|} = \begin{cases} \rho(\mathcal{T}, \delta) > 0 & \text{if } \delta > \delta^*(\mathcal{T}), \\ 0 & \text{otherwise} \end{cases}$$

Optimal $\mathcal{T}^*$: Minimizes threshold δ^* *and* maximizes overlap [Mondelli-Montanari '19], [Luo et al. '20]

Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2, \quad \text{i.i.d. Gaussian } \boldsymbol{A}, \quad d = 4096$$



Spectral estimator attains the **optimal** weak recovery threshold!

[Mondelli, Montanari '19]

Iterative methods for GLMs

$$\boldsymbol{x} \longrightarrow \boxed{\boldsymbol{A}} \xrightarrow{\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}} \boxed{q(\boldsymbol{z}, \varepsilon)} \longrightarrow \boldsymbol{y} = q(\boldsymbol{z}, \varepsilon)$$

$$\varepsilon \downarrow q(\boldsymbol{z}, \varepsilon)$$

GOAL: To maximize $\frac{|\langle \boldsymbol{x}, \boldsymbol{x} \rangle|}{\|\boldsymbol{x}\| \, \|\boldsymbol{x}\|}$

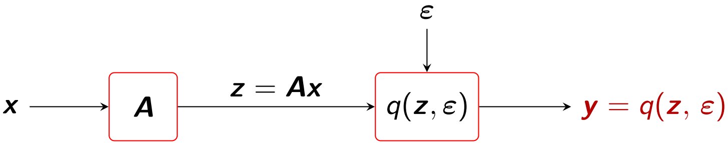
(or to minimize the MSE $\frac{1}{d}\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2$)

Spectral estimator provides warm start for iterative methods:

Gradient descent, alternating miniminization ...

- Generic, can incorporate certain constraints like sparsity
- But not well-equipped to exploit specific structural info about signal, e.g., known prior

Bayes-optimal estimation



$$\mathrm{MMSE}_d := \frac{1}{d}\mathbb{E}\{\|\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A}, \boldsymbol{y}\}\|^2\}.$$

$\lim_{d\to\infty} \mathrm{MMSE}_d$ for a fixed $\delta = \frac{n}{d}$

- Characterized for i.i.d. Gaussian designs [Barbier et al.'19]
- Replica formula for rotationally invariant designs [Kabashima '08, . . .]

Bayes-optimal estimation

$$\begin{tikzcd} & & \varepsilon \arrow[d] & \\ \boldsymbol{x} \arrow[r] & \boxed{\boldsymbol{A}} \arrow[r, "\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}"] & \boxed{q(\boldsymbol{z}, \varepsilon)} \arrow[r] & \boldsymbol{y} = q(\boldsymbol{z}, \varepsilon) \end{tikzcd}$$

$$\mathrm{MMSE}_d := \frac{1}{d}\mathbb{E}\{\|\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A}, \boldsymbol{y}\}\|^2\}.$$

$\lim_{d\to\infty} \mathrm{MMSE}_d$ for a fixed $\delta = \frac{n}{d}$

- Characterized for i.i.d. Gaussian designs [Barbier et al.'19]
- Replica formula for rotationally invariant designs [Kabashima '08, ...]

Bayes estimator infeasible in general

How well can we do with polynomial-time algorithms?

Approximate Message Passing for GLMs

GENERALIZED APPROXIMATE MESSAGE PASSING 1

Generalized Approximate Message Passing for Estimation with Random Linear Mixing

Sundeep Rangan, *Member, IEEE*

***Abstract*—We consider the estimation of an i.i.d. random vector observed through a linear transform followed by a componentwise, probabilistic (possibly nonlinear) measurement channel. A novel algorithm, called generalized approximate message passing (GAMP), is presented that provides computationally efficient approximate implementations of max-sum and sum-problem loopy belief propagation for such problems. The algorithm extends earlier approximate message passing methods to incorporate arbitrary distributions on both the input and output of the transform and can be applied to a wide range of problems in nonlinear compressed sensing and learning.**

Extending an analysis by Bayati and Montanari, we argue that the asymptotic componentwise behavior of the GAMP

from the system input vector $\mathbf{q}$, system output $\mathbf{y}$ and linear transform $\mathbf{A}$.

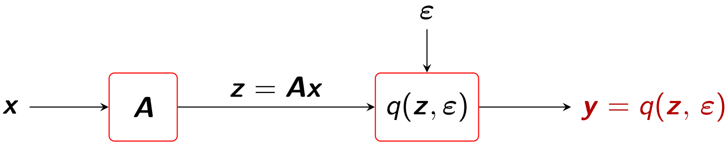
The formulation is general and has wide applicability – we will present several applications in Section II. However, for many non-Gaussian instances of the problem, exact computations of quantities such as the posterior mode or mean of $\mathbf{x}$ is computationally prohibitive. The primary difficulty is that the matrix $\mathbf{A}$ "couples" or "mixes" the coefficients of $\mathbf{x}$ into $\mathbf{z}$. If the transform matrix were the identity matrix (i.e. $m = n$ and $\mathbf{A} = I$), then the estimation problem would *decouple* into $m = n$ scalar estimation problems, each defined by the

Generalized Approximate Message Passing (**GAMP**)

- Can be tailored to take advantage of prior info about signal
- Rigorous performance characterization via **state evolution**

  Allows us to precisely compute asymptotic overlap, MSE

GAMP iteration

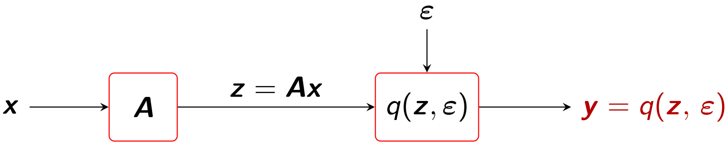


$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}}\boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathsf{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y})$$

$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}^{\mathsf{T}} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathsf{c}_t f_t(\boldsymbol{v}^t)$$

- f_t and g_t Lipschitz and act component-wise
- $\mathsf{b}_t = \frac{1}{n}\sum_{i=1}^{d} f_t'(v_i^t)$, $\mathsf{c}_t = \frac{1}{n}\sum_{i=1}^{n} g_t'(u_i^t; y_i)$

GAMP iteration



$$
\begin{aligned}
\boldsymbol{u}^t &= \frac{1}{\sqrt{\delta}} \boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathrm{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y}) \\
\boldsymbol{v}^{t+1} &= \frac{1}{\sqrt{\delta}} \boldsymbol{A}^{\mathsf{T}} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathrm{c}_t f_t(\boldsymbol{v}^t)
\end{aligned}
$$

- f_t and g_t Lipschitz and act component-wise
- $\mathrm{b}_t = \frac{1}{n} \sum_{i=1}^{d} f_t'(v_i^t)$, $\mathrm{c}_t = \frac{1}{n} \sum_{i=1}^{n} g_t'(u_i^t; y_i)$
- $\boldsymbol{v}^t$ an estimate of $\boldsymbol{x}$ and $\boldsymbol{u}^t$ an estimate of $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$

GAMP Asymptotics

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}}\boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathsf{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y})$$
$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}^{\mathsf{T}} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathsf{c}_t f_t(\boldsymbol{v}^t)$$

Theorem [Rangan '11, Javanmard-Montanari '13]

Assume i.i.d. Gaussian $\boldsymbol{A}$ and that the empirical distribution of $\boldsymbol{x}$ converges to the law of X. Then as $n, d \to \infty$, the empirical joint distribution of $(\boldsymbol{x}, \boldsymbol{v}^t)$ converges as

$$(\boldsymbol{x}, \boldsymbol{v}^t) \to (X, \mu_{V,t} X + W_{V,t}), \quad \text{where} \quad W_{V,t} \sim \mathsf{N}(0, \sigma^2_{V,t}) \perp X$$

GAMP Asymptotics

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}}\boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathrm{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y})$$
$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}^\mathsf{T} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathrm{c}_t f_t(\boldsymbol{v}^t)$$

Theorem [Rangan '11, Javanmard-Montanari '13]

Assume i.i.d. Gaussian $\boldsymbol{A}$ and that the empirical distribution of $\boldsymbol{x}$ converges to the law of X. Then as $n, d \to \infty$, the empirical joint distribution of $(\boldsymbol{x}, \boldsymbol{v}^t)$ converges as

$$(\boldsymbol{x}, \boldsymbol{v}^t) \to (X, \mu_{V,t} X + W_{V,t}), \quad \text{where} \quad W_{V,t} \sim \mathrm{N}(0, \sigma^2_{V,t}) \perp X$$

The empirical joint distribution of $(\boldsymbol{z}, \boldsymbol{u}^t)$ converges as

$$(\boldsymbol{z}, \boldsymbol{u}^t) \to (G, \mu_{U,t} G + W_{U,t}), \quad W_{U,t} \sim \mathrm{N}(0, \sigma^2_{U,t}) \perp G \sim \mathrm{N}(0,1)$$

State Evolution

$$(\boldsymbol{x}, \boldsymbol{v}^t) \to (X,\, \mu_{V,t} X + W_{V,t}), \quad W_{V,t} \sim \mathrm{N}(0, \sigma^2_{V,t}) \perp X$$

$$(\boldsymbol{z}, \boldsymbol{u}^t) \to (G,\, \mu_{U,t} G + W_{U,t}), \quad W_{U,t} \sim \mathrm{N}(0, \sigma^2_{U,t}) \perp G \sim \mathrm{N}(0, 1)$$

- Parameters $\mu_{V,t}, \sigma_{V,t}, \mu_{U,t}, \sigma_{U,t}$ recursively computed

$$(\mu_{V,t}, \sigma_{V,t}, \mu_{U,t}, \sigma_{U,t}) \;\Rightarrow\; (\mu_{V,t+1}, \sigma_{V,t+1}, \mu_{U,t+1}, \sigma_{U,t+1})$$

  Covariances of $(W_{V,t})_{t\geq 0}$ and $(W_{U,t})_{t\geq 0}$ also available

State Evolution

$$(\boldsymbol{x}, \boldsymbol{v}^t) \to (X,\, \mu_{V,t} X + W_{V,t}), \quad W_{V,t} \sim \mathrm{N}(0, \sigma_{V,t}^2) \perp X$$

$$(\boldsymbol{z}, \boldsymbol{u}^t) \to (G,\, \mu_{U,t} G + W_{U,t}), \quad W_{U,t} \sim \mathrm{N}(0, \sigma_{U,t}^2) \perp G \sim \mathrm{N}(0,1)$$

- Parameters $\mu_{V,t}, \sigma_{V,t}, \mu_{U,t}, \sigma_{U,t}$ recursively computed

  $$(\mu_{V,t}, \sigma_{V,t}, \mu_{U,t}, \sigma_{U,t}) \;\Rightarrow\; (\mu_{V,t+1}, \sigma_{V,t+1}, \mu_{U,t+1}, \sigma_{U,t+1})$$

  Covariances of $(W_{V,t})_{t\geq 0}$ and $(W_{U,t})_{t\geq 0}$ also available

- Exact asymptotic formulas for MSE, overlap etc.

  $$\lim_{d\to\infty} \frac{1}{d}\|\boldsymbol{x} - f_t(\boldsymbol{v}^t)\|^2 = \mathbb{E}\left[(X - f_t(\mu_{V,t} X + W_{V,t}))^2\right]$$

State Evolution

$$(\boldsymbol{x}, \boldsymbol{v}^t) \to (X,\, \mu_{V,t} X + W_{V,t}), \quad W_{V,t} \sim \mathsf{N}(0, \sigma_{V,t}^2) \perp X$$
$$(\boldsymbol{z}, \boldsymbol{u}^t) \to (G,\, \mu_{U,t} G + W_{U,t}), \quad W_{U,t} \sim \mathsf{N}(0, \sigma_{U,t}^2) \perp G \sim \mathsf{N}(0,1)$$

- Parameters $\mu_{V,t}, \sigma_{V,t}, \mu_{U,t}, \sigma_{U,t}$ recursively computed

  $$(\mu_{V,t}, \sigma_{V,t}, \mu_{U,t}, \sigma_{U,t}) \;\Rightarrow\; (\mu_{V,t+1}, \sigma_{V,t+1}, \mu_{U,t+1}, \sigma_{U,t+1})$$

  Covariances of $(W_{V,t})_{t\geq 0}$ and $(W_{U,t})_{t\geq 0}$ also available

- Exact asymptotic formulas for MSE, overlap etc.

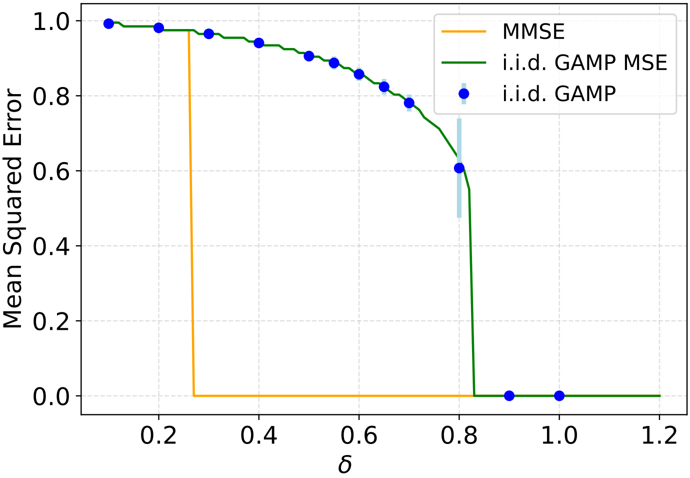
  $$\lim_{d\to\infty} \frac{1}{d} \|\boldsymbol{x} - f_t(\boldsymbol{v}^t)\|^2 = \mathbb{E}\left[(X - f_t(\mu_{V,t} X + W_{V,t}))^2\right]$$

- Can tailor f_t, g_t to signal prior and model. E.g.

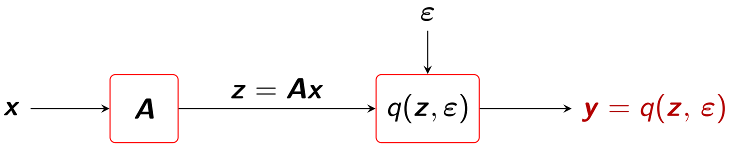
  $$f_t^*(v) = \mathbb{E}[X \mid \mu_{V,t} X + W_{V,t} = v]$$
  $$g_t^*(u, y) = \ldots \qquad \text{(formula based on } q(\cdot,\cdot))$$

Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2, \quad \text{i.i.d. Gaussian } \boldsymbol{A}, \quad \text{Prior} : P_X(-a) = 0.4,\ P_X(a) = 0.6$$



GAMP: Two Appealing Features



1. Asymptotic performance characterized by **state evolution**
2. **Conjectured to be optimal** among all poly-time algorithms [Celentano, Montanari, Wu '20]
3. Bayes-optimal with **spatially coupled** designs [Pascual Cobo, Hsieh, V '24]

GAMP: One Major Issue

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}}\boldsymbol{A}f_t(\boldsymbol{v}^t) - \mathrm{b}_t g_{t-1}(\boldsymbol{u}^{t-1};\boldsymbol{y})$$
$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}^\mathsf{T} g_t(\boldsymbol{u}^t;\boldsymbol{y}) - \mathrm{c}_t f_t(\boldsymbol{v}^t)$$

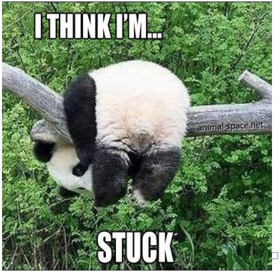
☹ For models such as phase retrieval, requires **initialization correlated** with $\boldsymbol{x}$ and **independent** of $\{\boldsymbol{A}\}$

GAMP: One Major Issue

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}} \boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathrm{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y})$$

$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}} \boldsymbol{A}^\mathsf{T} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathrm{c}_t f_t(\boldsymbol{v}^t)$$

- ☹ For models such as phase retrieval, requires **initialization correlated** with $\boldsymbol{x}$ and **independent** of $\{\boldsymbol{A}\}$
- ☹ With random intialization $\mu_{V,t} = 0$ for all t
  $\Rightarrow \boldsymbol{v}^t$ independent of $\boldsymbol{x}$



GAMP with Spectral Initialization

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}} \boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathrm{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y})$$
$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}} \boldsymbol{A}^\mathsf{T} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathrm{c}_t f_t(\boldsymbol{v}^t)$$

Spectral estimator provides natural initialization: $\boldsymbol{v}^1 = \hat{\boldsymbol{x}}^{\mathrm{s}}$

Recall $\hat{\boldsymbol{x}}^{\mathrm{s}}$ principal eigenvector of

$$\boldsymbol{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^\mathsf{T}$$

$\hat{\boldsymbol{x}}^{\mathrm{s}}$ and $\boldsymbol{A}$ are correlated $\Rightarrow$ standard state evolution analysis does not apply!

GAMP with Spectral Initialization

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}}\boldsymbol{A}f_t(\boldsymbol{v}^t) - \mathsf{b}_t g_{t-1}(\boldsymbol{u}^{t-1};\boldsymbol{y})$$
$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}^{\mathsf{T}}g_t(\boldsymbol{u}^t;\boldsymbol{y}) - \mathsf{c}_t f_t(\boldsymbol{v}^t)$$

Theorem [Mondelli-V '21]

Assume i.i.d. Gaussian $\boldsymbol{A}$, and $\mathcal{T}$ Lipschitz and satisfying mild regularity conditions. Then, if $\delta > \delta^*(\mathcal{T})$, state evolution is valid for GAMP initialized with $\boldsymbol{v}^1 = \hat{\boldsymbol{x}}^{\mathrm{s}}$. That is, for $t \geq 1$,

$$(\boldsymbol{x}, \boldsymbol{v}^t) \to (X,\, \mu_{V,t}X + W_{V,t}), \quad \text{where} \quad W_{V,t} \sim \mathsf{N}(0, \sigma^2_{V,t}) \perp X$$
$$(\boldsymbol{z}, \boldsymbol{u}^t) \to (G,\, \mu_{U,t}G + W_{U,t}), \quad G \sim \mathsf{N}(0,1) \perp W_{U,t} \sim \mathsf{N}(0, \sigma^2_{U,t})$$

$\mu_{V,1}, \sigma_{V,1}$ determined by asymptotic overlap between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}^{\mathrm{s}}$

Proof Idea: Artificial AMP iteration

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}} \boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathrm{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y})$$
$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}} \boldsymbol{A}^\mathsf{T} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathrm{c}_t f_t(\boldsymbol{v}^t)$$

$$\boldsymbol{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^\mathsf{T}$$

[Phase 1]

Artificial AMP simulates **power method** and iterates approach $\hat{\boldsymbol{x}}^{\mathrm{s}}$

Initialization can depend on unknown signal $\boldsymbol{x}$!

Proof Idea: Artificial AMP iteration

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}}\boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathsf{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y})$$
$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}^\mathsf{T} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathsf{c}_t f_t(\boldsymbol{v}^t)$$

$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^\mathsf{T}$$

[Phase 1]

Artificial AMP simulates **power method** and iterates approach $\hat{\boldsymbol{x}}^{\mathrm{s}}$

Initialization can depend on unknown signal $\boldsymbol{x}$!

[Phase 2]

Artificial AMP iterates mimic iterates of true GAMP

Proof Idea: Artificial AMP iteration

$$\boldsymbol{u}^t = \frac{1}{\sqrt{\delta}}\boldsymbol{A} f_t(\boldsymbol{v}^t) - \mathsf{b}_t g_{t-1}(\boldsymbol{u}^{t-1}; \boldsymbol{y})$$

$$\boldsymbol{v}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}^\mathsf{T} g_t(\boldsymbol{u}^t; \boldsymbol{y}) - \mathsf{c}_t f_t(\boldsymbol{v}^t)$$

$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^\mathsf{T}$$

[Phase 1]

Artificial AMP simulates **power method** and iterates approach $\hat{\boldsymbol{x}}^{\mathrm{s}}$

Initialization can depend on unknown signal $\boldsymbol{x}$!

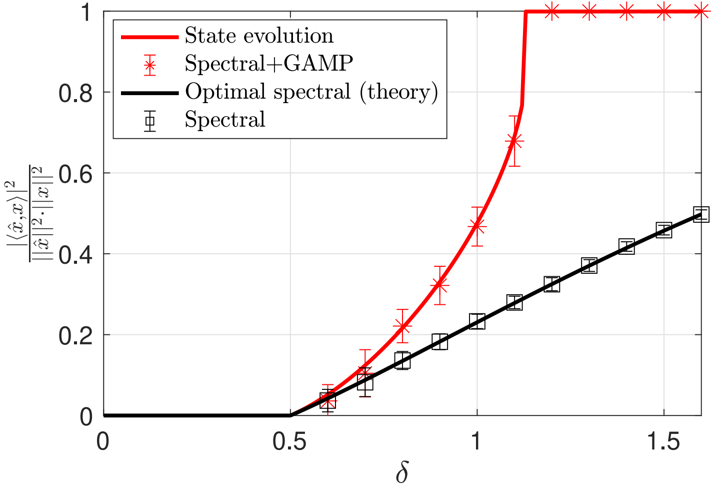
[Phase 2]

Artificial AMP iterates mimic iterates of true GAMP

- Phases 1 and 2 differ only in choice of f_t, g_t
- Can analyze via state evolution **if**:
  D_n has a **spectral gap**. Guaranteed by $\delta > \delta^*(\mathcal{T})$

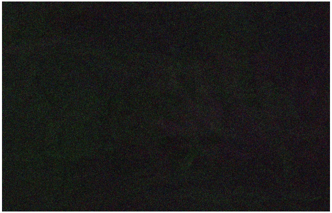
Phase Retrieval with synthetic data

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2, \quad \text{i.i.d. Gaussian } \boldsymbol{A}, \quad \text{Prior : N}(0,1), \quad d = 8000$$

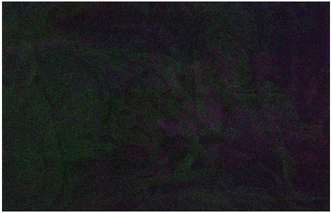


Phase Retrieval with natural images

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2, \quad \boldsymbol{A} \text{ from coded diffraction patterns}$$

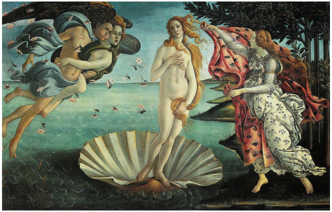


$\delta = 2.2$, spectral



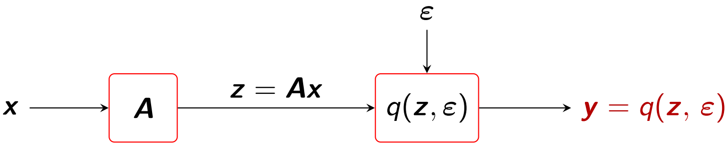
$\delta = 2.4$, spectral

$\delta = 2.2$, **spectral+AMP**



$\delta = 2.4$, **spectral+AMP**

So far . . .



With i.i.d. Gaussian $\boldsymbol{A}$, precise asymptotics for

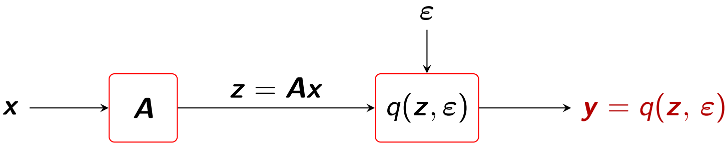
- Spectral estimator

  Provides warm start for iterative algorithms

  Optimal weak recovery threshold for phase retrieval

- AMP with spectral initialization

So far . . .



With i.i.d. Gaussian $\boldsymbol{A}$, precise asymptotics for

- Spectral estimator

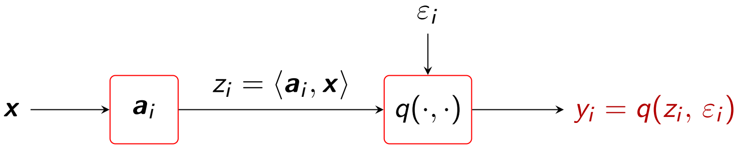
  Provides warm start for iterative algorithms

  Optimal weak recovery threshold for phase retrieval

- AMP with spectral initialization

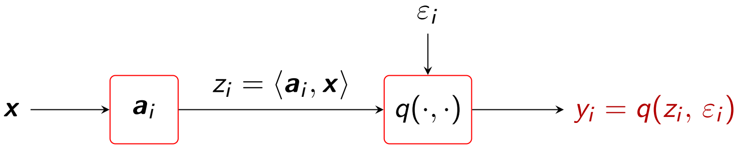
Spectral estimators and AMP for more general designs?

GLMs with correlated Gaussian design



- Covariate vectors $\{\boldsymbol{a}_i\}_{1 \le i \le n} \sim_{\text{i.i.d.}} \mathsf{N}(\mathbf{0}_d, \boldsymbol{\Sigma}/n)$
- Covariance $\boldsymbol{\Sigma}$ typically **unknown**

GLMs with correlated Gaussian design



- Covariate vectors $\{\boldsymbol{a}_i\}_{1 \le i \le n} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{\Sigma}/n)$
- Covariance $\boldsymbol{\Sigma}$ typically **unknown**

Spectral estimator

$$\boldsymbol{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{T}}, \quad \hat{\boldsymbol{x}}^{\mathrm{s}} = \text{ top eigenvector of } \boldsymbol{D}_n$$

- $\mathcal{T} : \mathbb{R} \to \mathbb{R}$ pre-processing function

Precise asymptotics

Theorem [ZJVM25]

Assume empirical spectral distribution of $\boldsymbol{\Sigma}$ converges to Law(Σ).

Let $\mathcal{T}$ be Lipschitz and satisfy some mild regularity conditions.

If $F(\delta, \Sigma, \mathcal{T}) > 0$, then as $n, d \to \infty$ with $n/d \to \delta$ we have:

1. **Spectral gap:** the limits of the top two eigenvalues of $\boldsymbol{D}_n$ are $\lambda_1(\delta, \Sigma, \mathcal{T}) > \lambda_2(\delta, \Sigma, \mathcal{T})$.
2. **Spectral estimator works:**
$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x} \rangle|}{\|\hat{\boldsymbol{x}}^{\mathrm{s}}\| \cdot \|\boldsymbol{x}\|} = \rho(\delta, \Sigma, \mathcal{T}) > 0.$$

- Explicit expressions for $\rho(\delta, \Sigma, \mathcal{T}), F(\delta, \Sigma, \mathcal{T}), \lambda_1(\delta, \Sigma, \mathcal{T}), \lambda_2(\delta, \Sigma, \mathcal{T})$.

Y. Zhang, H.C. Ji, R. Venkataramanan, M. Mondelli, *Spectral Estimators for Structured Generalized Linear Models via Approximate Message Passing*, Mathematical Statistics and Learning, 2025

Spectral threshold

Spectral estimator works if $F(\delta, \Sigma, \mathcal{T}) > 0$

Spectral threshold $=$ smallest δ s.t. $F(\delta, \Sigma, \mathcal{T}) > 0$

Spectral threshold

Spectral estimator works if $F(\delta, \Sigma, \mathcal{T}) > 0$

Spectral threshold $=$ smallest δ s.t. $F(\delta, \Sigma, \mathcal{T}) > 0$

Theorem [ZJVM25]

If $\delta > \delta^*(\Sigma)$, then there is $\mathcal{T}^*(\Sigma)$ s.t. spectral estimator works. Otherwise, under an additional technical assumption, there is no $\mathcal{T}$ s.t. $F(\delta, \Sigma, \mathcal{T}) > 0$.

- Explicit expressions for $\delta^*(\Sigma)$ and $\mathcal{T}^*(\Sigma)$.
- $\mathcal{T}^*(\Sigma)$ depends on Σ only via $\mathbb{E}[\Sigma]$

Spectral threshold

Spectral estimator works if $F(\delta, \Sigma, \mathcal{T}) > 0$

Spectral threshold $=$ smallest δ s.t. $F(\delta, \Sigma, \mathcal{T}) > 0$

Theorem [ZJVM25]

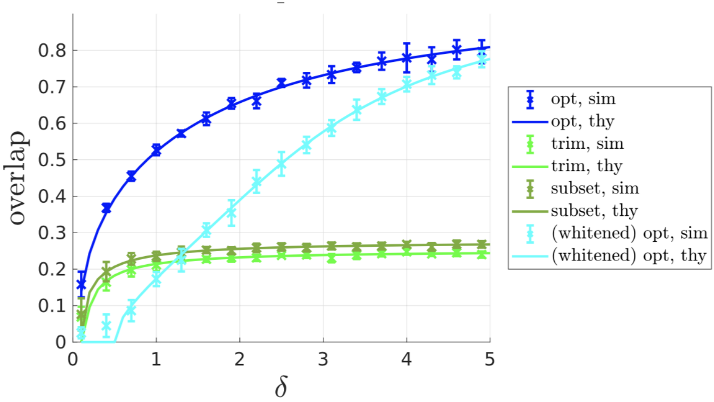
If $\delta > \delta^*(\Sigma)$, then there is $\mathcal{T}^*(\Sigma)$ s.t. spectral estimator works. Otherwise, under an additional technical assumption, there is no $\mathcal{T}$ s.t. $F(\delta, \Sigma, \mathcal{T}) > 0$.

- Explicit expressions for $\delta^*(\Sigma)$ and $\mathcal{T}^*(\Sigma)$.
- $\mathcal{T}^*(\Sigma)$ depends on Σ only via $\mathbb{E}[\Sigma]$

No need to estimate $\boldsymbol{\Sigma}$, but only $\widehat{\mathbb{E}[\Sigma]} = \mathrm{Tr}(\boldsymbol{\Sigma})/d$!

Example: Phase Retrieval

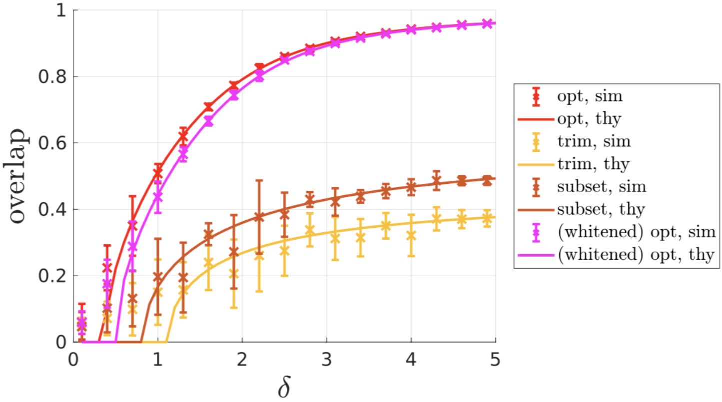
$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2, \quad \boldsymbol{\Sigma} \text{ Toeplitz}$$



Significant improvement over heuristic choices of $\mathcal{T}$

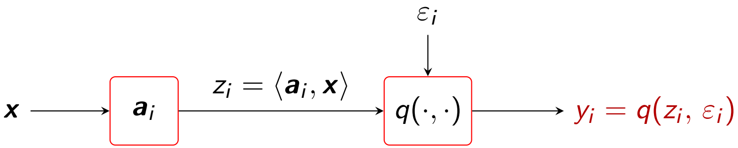
Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2, \quad \boldsymbol{\Sigma} \text{ circulant}$$



Significant improvement over heuristic choices of $\mathcal{T}$

GLMs with correlated Gaussian design



- Covariate vectors $\{\boldsymbol{a}_i\}_{1 \le i \le n} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{\Sigma}/n)$
- Covariance $\boldsymbol{\Sigma}$ typically **unknown**
- Initial estimate via spectral estimator:

$$\boldsymbol{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{T}}, \quad \hat{\boldsymbol{x}}^{\mathrm{s}} = \text{ top eigenvector of } \boldsymbol{D}_n$$

Can run AMP to improve performance but "optimal" AMP requires knowledge of $\boldsymbol{\Sigma}$

Rotationally Invariant Design

$$
\begin{tikzcd}
 & & & \boldsymbol{\varepsilon} \arrow[d] & \\
\boldsymbol{x} \arrow[r] & \boxed{\boldsymbol{A}} \arrow[rr, "\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}"] & & \boxed{q(\boldsymbol{z}, \varepsilon)} \arrow[r] & \boldsymbol{y} = q(\boldsymbol{z}, \varepsilon)
\end{tikzcd}
$$

Rotationally invariant $\boldsymbol{A}$

- SVD of $\boldsymbol{A} = \boldsymbol{O}\boldsymbol{\Lambda}\boldsymbol{Q}^T$
- $\boldsymbol{O}, \boldsymbol{Q}$ uniformly random orthogonal matrices
- Arbitrary singular values $\boldsymbol{\Lambda}$
- More general than Gaussian $\boldsymbol{A}$, can capture complex correlation structures in the data (in both rows and columns)
- i.i.d Gaussian a special case, but not correlated Gaussian

Spectral Estimator

$$\boldsymbol{A} = \begin{bmatrix} \longleftarrow & \boldsymbol{a}_1^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \boldsymbol{a}_n^T & \longrightarrow \end{bmatrix} \in \mathbb{R}^{n \times d}, \qquad \boldsymbol{y} = q(\boldsymbol{z}, \varepsilon), \quad \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$$

$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}}, \quad \hat{\boldsymbol{x}}^{\mathrm{s}} = \text{ top eigenvector of } \boldsymbol{D}_n$$

$\mathcal{T}$: pre-processing function of the form

$$\mathcal{T}(y) = \frac{\overline{g}(y)}{\overline{g}(y) + \gamma}, \quad \text{where } \gamma := \left(\frac{\kappa_4}{\kappa_2^2} + \delta\right) \mathbb{E}\{\overline{g}(Y)Z^2\}$$

Precise Asymptotics

Theorem [ZJVM26]

Assume empirical distribution of $\boldsymbol{\Lambda}$ converges to $\mathrm{Law}(\Lambda)$.

Let $\overline{g}$ satisfy some mild regularity conditions.

If $\delta > \delta_{\mathrm{Thresh}}(\overline{g}, \Lambda)$, then as $n, d \to \infty$ with $n/d \to \delta$ we have:

1. **Spectral gap:** the limits of the top two eigenvalues of $\boldsymbol{D}_n$ are $\lambda_1(\delta, \Lambda, \overline{g}) > \lambda_2(\delta, \Lambda, \overline{g})$.
2. **Spectral estimator works:** $\lim_{n\to\infty} \dfrac{|\langle \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x}\rangle|}{\|\hat{\boldsymbol{x}}^{\mathrm{s}}\| \, \|\boldsymbol{x}\|} = \rho(\delta, \Lambda, \overline{g}) > 0.$

- Explicit expressions for $\rho(\delta, \Lambda, \overline{g}), \lambda_1(\delta, \Lambda, \overline{g}), \lambda_2(\delta, \Lambda, \overline{g})$.

Y. Zhang, H. C. Ji, R.Venkataramanan, M. Mondelli, "Optimal Estimation in Orthogonally Invariant Generalized Linear Models: Spectral Initialization and Approximate Message Passing", 2026

Spectral threshold

Spectral threshold: Smallest δ for which there is $\overline{g}$ such that $\delta > \delta_{\mathrm{Thresh}}(\overline{g}, \Lambda)$

Spectral threshold

Spectral threshold: Smallest δ for which there is $\overline{g}$ such that $\delta > \delta_{\mathrm{Thresh}}(\overline{g}, \Lambda)$

Theorem [ZJVM26]

If $\delta > \delta^*(\Lambda)$, then there is $\overline{g}^*(\Lambda)$ s.t. spectral estimator works.
Otherwise, there is no $\overline{g}$ satisfying $\delta > \delta_{\mathrm{Thresh}}(\overline{g}, \Lambda)$.

Spectral threshold

Spectral threshold: Smallest δ for which there is $\overline{g}$ such that $\delta > \delta_{\mathrm{Thresh}}(\overline{g}, \Lambda)$

Theorem [ZJVM26]

If $\delta > \delta^*(\Lambda)$, then there is $\overline{g}^*(\Lambda)$ s.t. spectral estimator works. Otherwise, there is no $\overline{g}$ satisfying $\delta > \delta_{\mathrm{Thresh}}(\overline{g}, \Lambda)$.

- $\overline{g}^*(\Lambda)$ depends on Λ only via the second moment, can be easily estimated
- Matches conjectured optimal weak recovery threshold for polynomial-time algorithms [Kabashima '08, Maillard et al. '20]

Spectral threshold

Spectral threshold: Smallest δ for which there is $\overline{g}$ such that $\delta > \delta_{\text{Thresh}}(\overline{g}, \Lambda)$

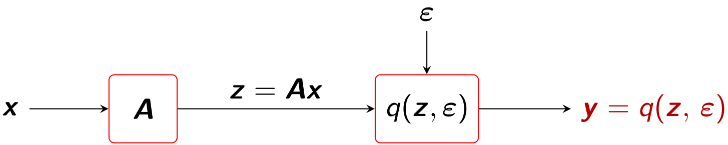
Theorem [ZJVM26]

If $\delta > \delta^*(\Lambda)$, then there is $\overline{g}^*(\Lambda)$ s.t. spectral estimator works. Otherwise, there is no $\overline{g}$ satisfying $\delta > \delta_{\text{Thresh}}(\overline{g}, \Lambda)$.

- $\overline{g}^*(\Lambda)$ depends on Λ only via the second moment, can be easily estimated
- Matches conjectured optimal weak recovery threshold for polynomial-time algorithms [Kabashima '08, Maillard et al. '20]

Analysis of spectral via Generalized Vector AMP (GVAMP)

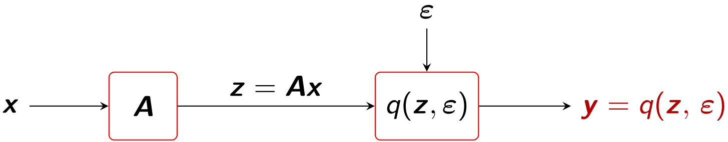
Generalized Vector AMP



$$
\begin{aligned}
\boldsymbol{u}^{t+1} &= \Phi_t(\boldsymbol{A}^\top \boldsymbol{A}) f_t(\boldsymbol{u}^t) + \widetilde{\Phi}_t(\boldsymbol{A})^\top g_t(\boldsymbol{v}^t; \boldsymbol{y}), \\
\boldsymbol{v}^{t+1} &= \Psi_t(\boldsymbol{A}\boldsymbol{A}^\top) g_t(\boldsymbol{v}^t; \boldsymbol{y}) + \widetilde{\Psi}_t(\boldsymbol{A}) f_t(\boldsymbol{u}^t),
\end{aligned}
$$

- Scalar functions Φ_t, Ψ_t applied to eigenvalues and
  Trace-free: $\mathbb{E}[\Phi_t(\Lambda_n)] = \mathbb{E}[\Phi_t(\Lambda_d)] = 0$
- Scalar functions $\widetilde{\Phi}_t, \widetilde{\Psi}_t$ applied to singular values
- Functions f_t, g_t *divergence free*: $\mathbb{E}[f_t'] = E[g_t'] = 0$

Generalized Vector AMP



$$
\begin{aligned}
\boldsymbol{u}^{t+1} &= \Phi_t(\boldsymbol{A}^\top \boldsymbol{A}) f_t(\boldsymbol{u}^t) + \widetilde{\Phi}_t(\boldsymbol{A})^\top g_t(\boldsymbol{v}^t; \boldsymbol{y}),\\
\boldsymbol{v}^{t+1} &= \Psi_t(\boldsymbol{A}\boldsymbol{A}^\top) g_t(\boldsymbol{v}^t; \boldsymbol{y}) + \widetilde{\Psi}_t(\boldsymbol{A}) f_t(\boldsymbol{u}^t),
\end{aligned}
$$

- Scalar functions Φ_t, Ψ_t applied to eigenvalues and
  Trace-free: $\mathbb{E}[\Phi_t(\Lambda_n)] = \mathbb{E}[\Phi_t(\Lambda_d)] = 0$
- Scalar functions $\widetilde{\Phi}_t, \widetilde{\Psi}_t$ applied to singular values
- Functions f_t, g_t *divergence free*: $\mathbb{E}[f_t'] = E[g_t'] = 0$

Special case: Vector AMP [Rangan-Schniter-Fletcher '19] , Orthogonal AMP [Ma-Ping '17] for linear model

State Evolution

Theorem [ZJVM26]

Assume rotationally invariant $\boldsymbol{A}$, and that the empirical distributions of $\boldsymbol{x}, \boldsymbol{\Lambda}$ converges to the laws of X, Λ. Then as $n, d \to \infty$, the empirical joint distribution of $(\boldsymbol{x}, \boldsymbol{v}^t)$ converges as

$$(\boldsymbol{x}, \boldsymbol{v}^t) \to (X, \mu_{V,t} X + W_{V,t}), \quad \text{where} \quad W_{V,t} \sim \mathrm{N}(0, \sigma_{V,t}^2) \perp X$$

The empirical joint distribution of $(\boldsymbol{z}, \boldsymbol{u}^t)$ converges as

$$(\boldsymbol{z}, \boldsymbol{u}^t) \to (G, \mu_{U,t} G + W_{U,t}), \quad W_{U,t} \sim \mathrm{N}(0, \sigma_{U,t}^2) \perp G \sim \mathrm{N}(0, 1)$$

State Evolution

Theorem [ZJVM26]

Assume rotationally invariant $\boldsymbol{A}$, and that the empirical distributions of $\boldsymbol{x}, \boldsymbol{\Lambda}$ converges to the laws of X, Λ. Then as $n, d \to \infty$, the empirical joint distribution of $(\boldsymbol{x}, \boldsymbol{v}^t)$ converges as

$$(\boldsymbol{x}, \boldsymbol{v}^t) \to (X, \mu_{V,t} X + W_{V,t}), \quad \text{where} \quad W_{V,t} \sim \mathrm{N}(0, \sigma_{V,t}^2) \perp X$$

The empirical joint distribution of $(\boldsymbol{z}, \boldsymbol{u}^t)$ converges as

$$(\boldsymbol{z}, \boldsymbol{u}^t) \to (G, \mu_{U,t} G + W_{U,t}), \quad W_{U,t} \sim \mathrm{N}(0, \sigma_{U,t}^2) \perp G \sim \mathrm{N}(0, 1)$$

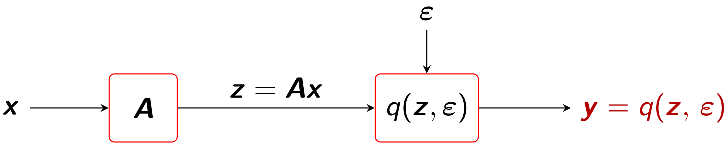
State evolution parameters depend on X, Λ and functions Φ_t, Ψ_t, $\widetilde{\Phi}_t, \widetilde{\Psi}_t$, f_t, g_t

$$\begin{tikzcd} & & & \varepsilon \arrow[d] & \\ \boldsymbol{x} \arrow[r] & \boxed{\boldsymbol{A}} \arrow[rr, "\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}"] & & \boxed{q(\boldsymbol{z}, \varepsilon)} \arrow[r] & \boldsymbol{y} = q(\boldsymbol{z}, \varepsilon) \end{tikzcd}$$

$$\begin{aligned} \boldsymbol{u}^{t+1} &= \Phi_t(\boldsymbol{A}^\top \boldsymbol{A}) f_t(\boldsymbol{u}^t) + \widetilde{\Phi}_t(\boldsymbol{A})^\top g_t(\boldsymbol{v}^t; \boldsymbol{y}), \\ \boldsymbol{v}^{t+1} &= \Psi_t(\boldsymbol{A}\boldsymbol{A}^\top) g_t(\boldsymbol{v}^t; \boldsymbol{y}) + \widetilde{\Psi}_t(\boldsymbol{A}) f_t(\boldsymbol{u}^t), \end{aligned}$$

Spectral Estimator

- Power method via linearized GVAMP
- Needs spectral gap: shown via random matrix theory + state evolution of GVAMP



$$\boldsymbol{u}^{t+1} = \Phi_t(\boldsymbol{A}^\top \boldsymbol{A}) f_t(\boldsymbol{u}^t) + \widetilde{\Phi}_t(\boldsymbol{A})^\top g_t(\boldsymbol{v}^t; \boldsymbol{y}),$$
$$\boldsymbol{v}^{t+1} = \Psi_t(\boldsymbol{A}\boldsymbol{A}^\top) g_t(\boldsymbol{v}^t; \boldsymbol{y}) + \widetilde{\Psi}_t(\boldsymbol{A}) f_t(\boldsymbol{u}^t),$$

Spectral Estimator

- Power method via linearized GVAMP
- Needs spectral gap: shown via random matrix theory + state evolution of GVAMP

Bayes GVAMP

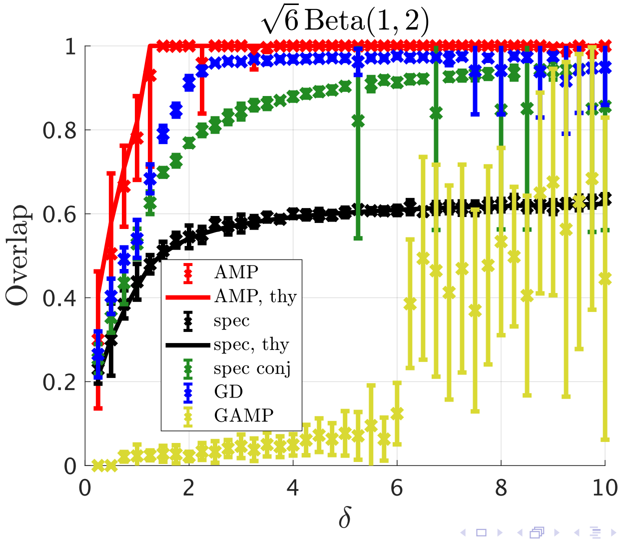
"Optimal" choice of functions $\Phi_t, \Psi_t, \widetilde{\Phi}_t, \widetilde{\Psi}_t, f_t, g_t$

Can be initialized with spectral estimator $\boldsymbol{v}^1 = \hat{\boldsymbol{x}}^{\mathrm{s}}$

State evolution fixed points match conjectured Bayes-optimal

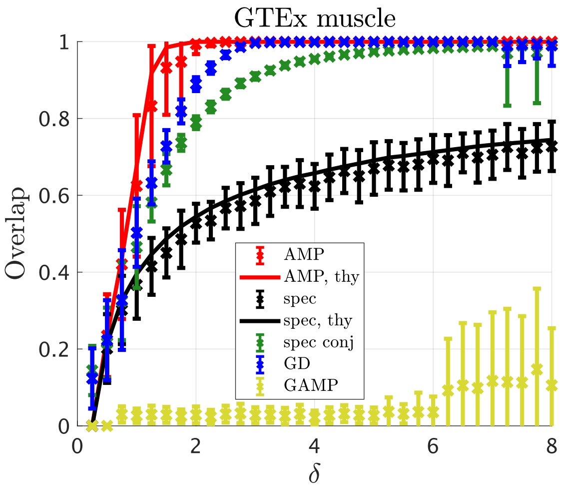
Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$$

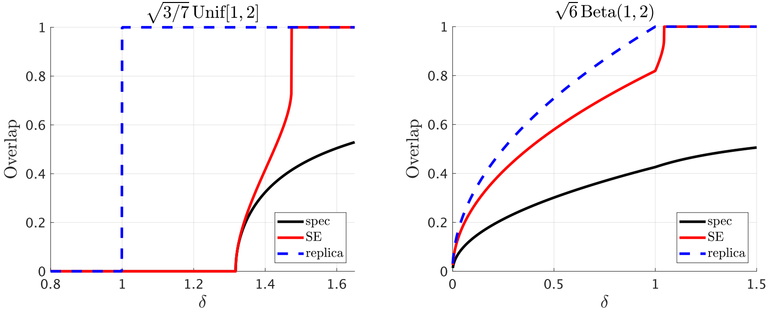


Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$$



Information-Computation Gap



. . . although state evolution fixed points satisfy replica formula!

Takeaways

AMP a valuable tool for:

1. **High-dimensional estimation in non-linear models**

2. **High-dimensional asymptotics**

Takeaways

AMP a valuable tool for:

1. **High-dimensional estimation in non-linear models**
   - Conjectured best among poly-time algorithms
   - Can initialize with spectral estimators
   - **Minimal model assumptions**

2. **High-dimensional asymptotics**

Takeaways

AMP a valuable tool for:

1. **High-dimensional estimation in non-linear models**
   - Conjectured best among poly-time algorithms
   - Can initialize with spectral estimators
   - **Minimal model assumptions**

2. **High-dimensional asymptotics**
   - Spectral estimators
   - Penalized regression, e.g. LASSO, logistic regression
   - General first-order methods

Applications of G-VAMP in communications?