# From Spectral Estimators to Approximate Message Passing... And Back

Marco Mondelli

Institute of Science and Technology Austria (ISTA)

Banff International Research Station, March 13, 2024

Institute of Science and Technology Austria


SOMETIMES
THEY COME
BACK

# Credits

Andrea Montanari (Stanford)

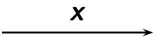
Ramji Venkataramanan (Cambridge)

Yihan Zhang (ISTA)

Hong Chang Ji (ISTA)

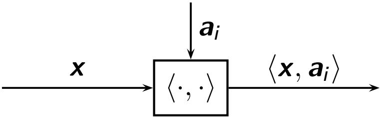
# From **Spectral Estimators** to Approximate Message Passing. . . And Back

# Generalized linear models



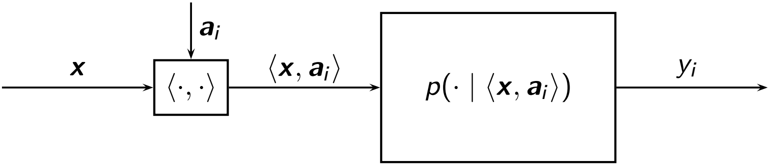

- Signal to **recover** $\boldsymbol{x} \in \mathbb{R}^d$.

# Generalized linear models



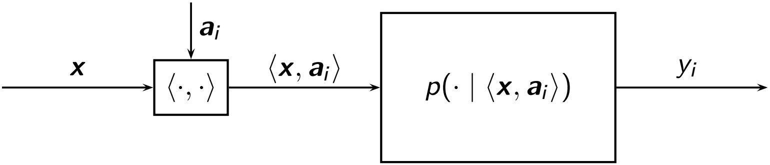

- Signal to **recover** $\boldsymbol{x} \in \mathbb{R}^d$.
- **Known** sensing vector $\boldsymbol{a}_i$, $i \in [n]$.

# Generalized linear models



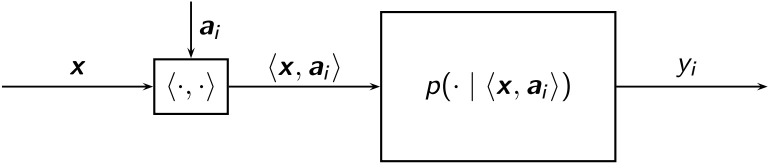

- Signal to **recover** $\boldsymbol{x} \in \mathbb{R}^d$.
- **Known** sensing vector $\boldsymbol{a}_i$, $i \in [n]$.
- Measurement $y_i \sim p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle)$, $i \in [n]$.

# Generalized linear models




- Signal to **recover** $\boldsymbol{x} \in \mathbb{R}^d$.
- **Known** sensing vector $\boldsymbol{a}_i$, $i \in [n]$.
- Measurement $y_i \sim p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle)$, $i \in [n]$.
- **High-dimensional** regime: $n, d \to \infty$ and $n/d \to \delta \in (0, \infty)$.

# Spectral initialization

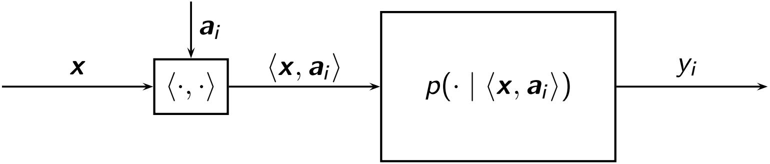



Most algorithms are iterative and require an **initialization**, often given by a spectral estimator:

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

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

# Spectral initialization




Most algorithms are iterative and require an **initialization**, often given by a spectral estimator:

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

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

**Key questions**

- What's the performance of $\hat{\boldsymbol{x}}^{\mathrm{s}}$ (e.g., in terms of $\frac{|\langle \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x} \rangle|}{\|\hat{\boldsymbol{x}}^{\mathrm{s}}\| \cdot \|\boldsymbol{x}\|}$)?
- What's the optimal $\mathcal{T}$?

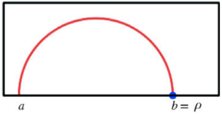
# Phase transition for random matrices

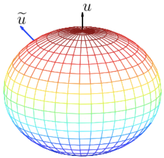
$$\boldsymbol{D}_n = \boldsymbol{X}_n + \theta \boldsymbol{u}\boldsymbol{u}^{\mathsf{T}}, \qquad \text{with } \lambda_1(\boldsymbol{X}_n) = b.$$

# Phase transition for random matrices

$$\boldsymbol{D}_n = \boldsymbol{X}_n + \theta \boldsymbol{u}\boldsymbol{u}^{\mathsf{T}}, \qquad \text{with } \lambda_1(\boldsymbol{X}_n) = b.$$

- $\theta < \theta_c$

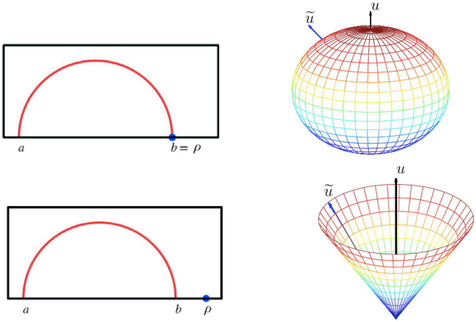


# Phase transition for random matrices

$$\boldsymbol{D}_n = \boldsymbol{X}_n + \theta \boldsymbol{u}\boldsymbol{u}^\mathsf{T}, \qquad \text{with } \lambda_1(\boldsymbol{X}_n) = b.$$

- $\theta < \theta_c$
- $\theta > \theta_c$



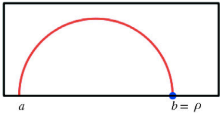

[BGN11]

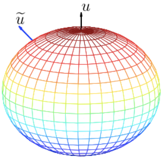
**Eigenvalue gap $\Rightarrow$ eigenvector correlation**

# Phase transition for spectral algorithm

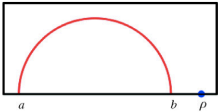
$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}}, \quad \hat{\boldsymbol{x}}^{\mathrm{s}} = \text{ principal eigenvector of } \boldsymbol{D}_n.$$

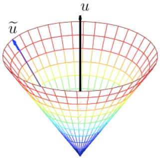
- $F(\delta, \mathcal{T}) < 0$




- $F(\delta, \mathcal{T}) > 0$




**Reduction to a rank-1 perturbation**

# Precise asymptotics for spectral estimators

**Theorem**

Let $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$. If $F(\delta,\mathcal{T}) > 0$, then:

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

If $F(\delta,\mathcal{T}) < 0$, then $\lambda_1(\delta,\mathcal{T}) = \lambda_2(\delta,\mathcal{T})$ and $\rho(\delta,\mathcal{T}) = 0$.

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

# Precise asymptotics for spectral estimators

**Theorem**

Let $\{\boldsymbol{a}_i\}_{1\leq i\leq n} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$. If $F(\delta, \mathcal{T}) > 0$, then:

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

If $F(\delta, \mathcal{T}) < 0$, then $\lambda_1(\delta, \mathcal{T}) = \lambda_2(\delta, \mathcal{T})$ and $\rho(\delta, \mathcal{T}) = 0$.

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

- Precise asymptotics allow to optimize $\mathcal{T}$ [LL17, MM18, LAL19]

[MM18] M. Mondelli and A. Montanari, "Fundamental Limits of Weak Recovery with Applications to Phase Retrieval", *COLT 2018 & FoCM 2019*.

# Precise asymptotics for spectral estimators

**Theorem**

Let $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$. If $F(\delta, \mathcal{T}) > 0$, then:

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

If $F(\delta, \mathcal{T}) < 0$, then $\lambda_1(\delta, \mathcal{T}) = \lambda_2(\delta, \mathcal{T})$ and $\rho(\delta, \mathcal{T}) = 0$.

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

- Precise asymptotics allow to optimize $\mathcal{T}$ [LL17, MM18, LAL19]
- [DMM20, DBMM20, MDX+21] consider a Haar matrix $\boldsymbol{A}$.

[MM18] M. Mondelli and A. Montanari, "Fundamental Limits of Weak Recovery with Applications to Phase Retrieval", *COLT 2018 & FoCM 2019*.

# From Spectral Estimators to **Approximate Message Passing**. . . And Back

# How to solve phase retrieval?

Most algorithms are iterative and require an **initialization**:

- Approximate message passing [Ran11, SR15]
- Alternating minimization [NJS13]
- Wirtinger flow [CLS15]
- Iterative projections [LGL15]
- Kaczmarz method [Wei15]
- Many many more. . . [FS20]

# How to solve phase retrieval?

Most algorithms are iterative and require an **initialization**:

- **Approximate message passing** [Ran11, SR15]
- Alternating minimization [NJS13]
- Wirtinger flow [CLS15]
- Iterative projections [LGL15]
- Kaczmarz method [Wei15]
- Many many more. . . [FS20]

# Generalized Approximate Message Passing (GAMP)

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

# Generalized Approximate Message Passing (GAMP)

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

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

# Generalized Approximate Message Passing (GAMP)

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

## Theorem [Ran11, JM13]

The empirical joint distribution of $(\boldsymbol{u}^t, \boldsymbol{v}^t)$ converges to the law of

$$(U_t, V_t) \triangleq (\mu_{U,t} G + \sigma_{U,t} W_{U,t}, \mu_{V,t} X + \sigma_{V,t} W_{V,t}),$$

with $G \sim \mathrm{N}(0,1) \perp W_{U,t} \sim \mathrm{N}(0,1)$ and $X \perp W_{V,t} \sim \mathrm{N}(0,1)$.

# Generalized Approximate Message Passing (GAMP)

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

**Theorem [Ran11, JM13]**

The empirical joint distribution of $(\boldsymbol{u}^t, \boldsymbol{v}^t)$ converges to the law of

$$(U_t, V_t) \triangleq (\mu_{U,t} G + \sigma_{U,t} W_{U,t}, \mu_{V,t} X + \sigma_{V,t} W_{V,t}),$$

with $G \sim \mathsf{N}(0,1) \perp W_{U,t} \sim \mathsf{N}(0,1)$ and $X \perp W_{V,t} \sim \mathsf{N}(0,1)$.

- **Deterministic scalar** recursion for $\{\mu_{U,t}, \mu_{V,t}, \sigma_{U,t}, \sigma_{V,t}\}_{t\geq 1}$.
- **Bayes-optimal** (unless statistical-to-computational barrier) [BKM$^+$19].

# Spectral initialization of AMP

**Key difficulty**

Spectral initialization depends on the design matrix $\boldsymbol{A}$: AMP and SE need to be changed accordingly.

- Low-rank matrix estimation in [MV21a] (different approach)
- Heuristic argument in [MXM18]

# Spectral initialization of AMP

**Key difficulty**

Spectral initialization depends on the design matrix $\boldsymbol{A}$: AMP and SE need to be changed accordingly.

- Low-rank matrix estimation in [MV21a] (different approach)
- Heuristic argument in [MXM18]

**AMP correction and provable state evolution in [MV21b]**

[MV21b] M. Mondelli and R. Venkataramanan, "Approximate Message Passing with Spectral Initialization for Generalized Linear Models", *AISTATS 2021 & JSTAT 2022*.

# Spectral initialization of AMP

**Key difficulty**

Spectral initialization depends on the design matrix $\boldsymbol{A}$: AMP and SE need to be changed accordingly.

- Low-rank matrix estimation in [MV21a] (different approach)
- Heuristic argument in [MXM18]

**AMP correction and provable state evolution in [MV21b]**

Design and analyze an **artificial AMP**

[MV21b] M. Mondelli and R. Venkataramanan, "Approximate Message Passing with Spectral Initialization for Generalized Linear Models", *AISTATS 2021 & JSTAT 2022*.

# Artificial AMP

**[Phase #1]** Iterates of artificial AMP approach $\hat{\boldsymbol{x}}^{\rm s}$ via **power method**.

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

# Artificial AMP

**[Phase #1]** Iterates of artificial AMP approach $\hat{\boldsymbol{x}}^{\mathrm{s}}$ via **power method**.

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

**[Phase #2]** Iterates of artificial AMP **mimic** iterates of true AMP.

# Artificial AMP as a power method

**[Phase #1]** Iterates of artificial AMP approach $\hat{\boldsymbol{x}}^{\mathrm{s}}$ via **power method**.

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

Choose $f_t, g_t$ so that

$$\boldsymbol{v}^{t+1} \propto \boldsymbol{D}_n \boldsymbol{v}^t, \qquad \text{with } \boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}}$$

- As $t \to \infty$, $\boldsymbol{v}^t$ aligned with $\mathrm{eig}(\boldsymbol{D}_n)$

# Artificial AMP as a power method

**[Phase #1]** Iterates of artificial AMP approach $\hat{\boldsymbol{x}}^{\mathrm{s}}$ via **power method**.

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

Choose $f_t, g_t$ so that

$$\boldsymbol{v}^{t+1} \propto \boldsymbol{D}_n \boldsymbol{v}^t, \qquad \text{with } \boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}}$$

- As $t \to \infty$, $\boldsymbol{v}^t$ aligned with $\mathrm{eig}(\boldsymbol{D}_n)$ if $\boldsymbol{D}_n$ has a **spectral gap**

# Artificial AMP as a power method

**[Phase #1]** Iterates of artificial AMP approach $\hat{\boldsymbol{x}}^{\mathrm{s}}$ via **power method**.

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

Choose $f_t, g_t$ so that

$$\boldsymbol{v}^{t+1} \propto \boldsymbol{D}_n \boldsymbol{v}^t, \qquad \text{with } \boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^\mathsf{T}$$

- As $t \to \infty$, $\boldsymbol{v}^t$ aligned with $\mathrm{eig}(\boldsymbol{D}_n)$ if $\boldsymbol{D}_n$ has a **spectral gap**

Spectral gap proved in [MM18]!


SOMETIMES
THEY COME
BACK

# From Spectral Estimators to Approximate Message Passing. . . **And Back**

# Towards heterogeneous and correlated data

So far, estimation of single signal $\boldsymbol{x}$ via design matrix $\boldsymbol{A}$ i.i.d. Gaussian

- In practice, data are heterogeneous $\Longrightarrow$ **mixed GLMs**
- In practice, data have correlations $\Longrightarrow$ **structured GLMs**

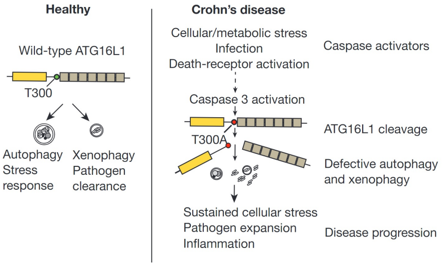
# Towards heterogeneous and correlated data

So far, estimation of single signal $\boldsymbol{x}$ via design matrix $\boldsymbol{A}$ i.i.d. Gaussian

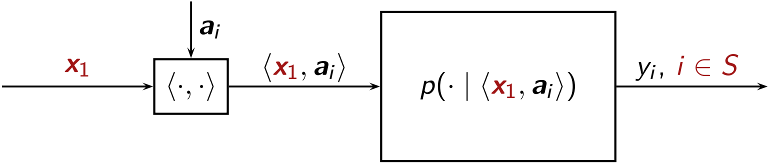
- In practice, data are heterogeneous $\Longrightarrow$ **mixed GLMs**
- In practice, data have correlations $\Longrightarrow$ **structured GLMs**

Example: Genome-Wide Association Studies (GWAS)

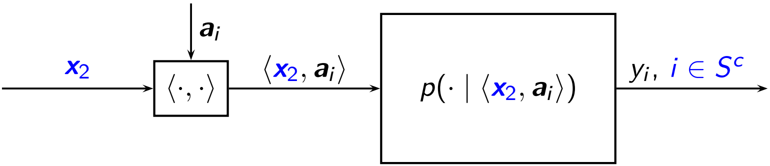
- Discovering novel biological mechanisms [MLP+14]
- Advancement in clinical care (validating new disease biomarkers, personalized medicine)



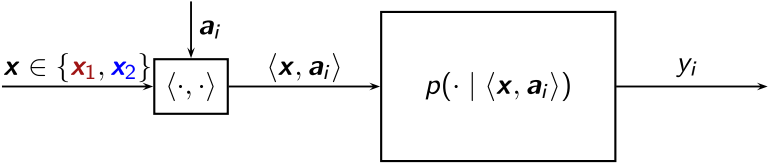
# Mixed generalized linear models



# Mixed generalized linear models



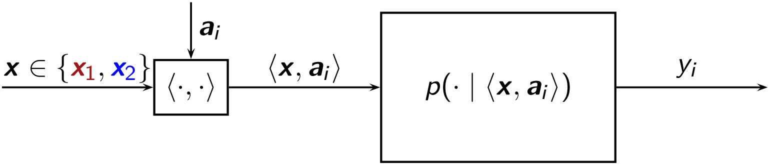
# Mixed generalized linear models




- Signals to **recover** $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$.
- **Known** sensing vector $\boldsymbol{a}_i$, $i \in [n]$.
- **Unknown** (latent) variables $\eta_i$, $i \in [n]$.
- Measurement $y_i \sim p(\cdot \mid \eta_i \langle \boldsymbol{x}_1, \boldsymbol{a}_i \rangle + (1 - \eta_i) \langle \boldsymbol{x}_2, \boldsymbol{a}_i \rangle)$, $i \in [n]$.
- **High dimensional** regime: $n, d \to \infty$ and $n/d \to \delta \in (0, \infty)$.

Model data **heterogeneity**, with applications in biology, physics, and economics [MP04, GL07, LSL19, DGP20].

# Mixed generalized linear models




- $\boldsymbol{x}_1, \boldsymbol{x}_2$ i.i.d. and uniform on the sphere with radius $\sqrt{d}$.
- $\{\boldsymbol{a}_i\}_{1 \le i \le n} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$.
- High dimensional regime: $n, d \to \infty$ and $n/d \to \delta \in (0, \infty)$.

**Spectral estimator:**

$$\boldsymbol{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{T}}, \quad \hat{\boldsymbol{x}}_1^{\mathrm{s}} / \hat{\boldsymbol{x}}_2^{\mathrm{s}} = \text{ first/second top eigenvector of } \boldsymbol{D}_n$$

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

# Precise asymptotics for spectral estimators

**Theorem [ZMV22]**

Let $\boldsymbol{x}_1, \boldsymbol{x}_2$ be i.i.d. and uniform on the sphere, and $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$. If $F_1(\delta, \mathcal{T}) > 0$, then

$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}_1^{\mathrm{s}}, \boldsymbol{x}\rangle|}{||\hat{\boldsymbol{x}}_1^{\mathrm{s}}|| \cdot ||\boldsymbol{x}||} = \rho_1(\delta, \mathcal{T}).$$

Explicit expressions for $\rho_1(\delta, \mathcal{T}), F_1(\delta, \mathcal{T})$

[ZMV22] Y. Zhang, M. Mondelli, and R. Venkataramanan, "Precise Asymptotics for Spectral Methods in Mixed Generalized Linear Models", *arXiv:2211.11368*, 2022.

# Precise asymptotics for spectral estimators

**Theorem [ZMV22]**

Let $\boldsymbol{x}_1, \boldsymbol{x}_2$ be i.i.d. and uniform on the sphere, and $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{\text{i.i.d.}} \mathsf{N}(\mathbf{0}_d, \boldsymbol{I}_d/d)$. If $F_1(\delta, \mathcal{T}) > 0$, then

$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}_1^{\mathrm{s}}, \boldsymbol{x}\rangle|}{||\hat{\boldsymbol{x}}_1^{\mathrm{s}}|| \cdot ||\boldsymbol{x}||} = \rho_1(\delta, \mathcal{T}).$$

If $F_2(\delta, \mathcal{T}) > 0$, then

$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}_2^{\mathrm{s}}, \boldsymbol{x}\rangle|}{||\hat{\boldsymbol{x}}_2^{\mathrm{s}}|| \cdot ||\boldsymbol{x}||} = \rho_2(\delta, \mathcal{T}).$$

Explicit expressions for $\rho_1(\delta, \mathcal{T}), F_1(\delta, \mathcal{T})$, $\rho_2(\delta, \mathcal{T}), F_2(\delta, \mathcal{T})$.

[ZMV22] Y. Zhang, M. Mondelli, and R. Venkataramanan, "Precise Asymptotics for Spectral Methods in Mixed Generalized Linear Models", *arXiv:2211.11368*, 2022.

# Precise asymptotics for spectral estimators

**Theorem [ZMV22]**

Let $\boldsymbol{x}_1, \boldsymbol{x}_2$ be i.i.d. and uniform on the sphere, and $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$. If $F_1(\delta, \mathcal{T}) > 0$, then

$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}_1^{\mathrm{s}}, \boldsymbol{x}\rangle|}{||\hat{\boldsymbol{x}}_1^{\mathrm{s}}|| \cdot ||\boldsymbol{x}||} = \rho_1(\delta, \mathcal{T}).$$

If $F_2(\delta, \mathcal{T}) > 0$, then

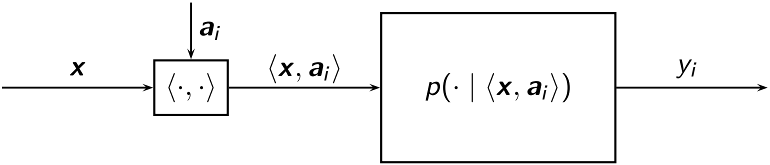
$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}_2^{\mathrm{s}}, \boldsymbol{x}\rangle|}{||\hat{\boldsymbol{x}}_2^{\mathrm{s}}|| \cdot ||\boldsymbol{x}||} = \rho_2(\delta, \mathcal{T}).$$

Explicit expressions for $\rho_1(\delta, \mathcal{T}), F_1(\delta, \mathcal{T})$, $\rho_2(\delta, \mathcal{T}), F_2(\delta, \mathcal{T})$.

- **Optimize** $\mathcal{T}$ both in terms of spectral threshold and overlap.

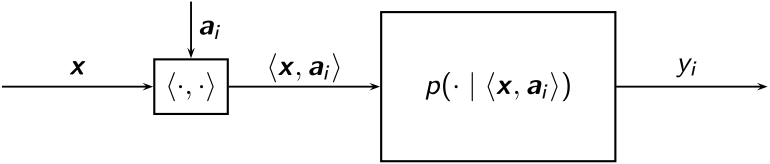
[ZMV22] Y. Zhang, M. Mondelli, and R. Venkataramanan, "Precise Asymptotics for Spectral Methods in Mixed Generalized Linear Models", *arXiv:2211.11368*, 2022.

# Generalized linear models with general Gaussian design




- $\boldsymbol{x}$ with i.i.d. zero-mean unit-variance components.
- $\{\boldsymbol{a}_i\}_{1 \le i \le n} \sim_{\text{i.i.d.}} \mathsf{N}(\mathbf{0}_d, \boldsymbol{\Sigma}/n)$.
- High dimensional regime: $n, d \to \infty$ and $n/d \to \delta \in (0, \infty)$.

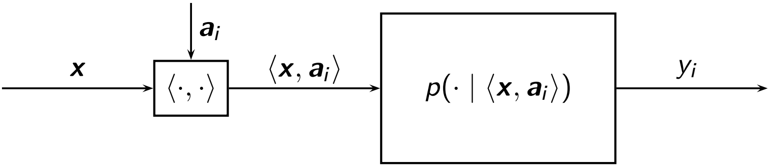
# Generalized linear models with general Gaussian design




- $\boldsymbol{x}$ with i.i.d. zero-mean unit-variance components.
- $\{\boldsymbol{a}_i\}_{1 \le i \le n} \sim_{\text{i.i.d.}} \mathsf{N}(\mathbf{0}_d, \boldsymbol{\Sigma}/n)$.
- High dimensional regime: $n, d \to \infty$ and $n/d \to \delta \in (0, \infty)$.

Anisotropic covariates commonly seen in practice, but existing work mostly focuses on penalized regression [W09, GBRD14, JM14, ZZ14, JM18, ZSC22].

# Generalized linear models with general Gaussian design




- $\boldsymbol{x}$ with i.i.d. zero-mean unit-variance components.
- $\{\boldsymbol{a}_i\}_{1 \le i \le n} \sim_{\text{i.i.d.}} \mathsf{N}(\mathbf{0}_d, \boldsymbol{\Sigma}/n)$.
- High dimensional regime: $n, d \to \infty$ and $n/d \to \delta \in (0, \infty)$.

**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 for spectral estimators

**Theorem [ZJVM23]**

Let $\boldsymbol{x}$ have i.i.d. zero-mean unit-variance components, and $\{\boldsymbol{a}_i\}_{1\leq i\leq n} \sim_{\text{i.i.d.}} \mathrm{N}(\mathbf{0}_d, \boldsymbol{\Sigma}/n)$. Assume $\mathcal{T}$ is Lipschitz and satisfies some mild regularity conditions. If $F(\delta, \Sigma, \mathcal{T}) > 0$, then:

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

[ZJVM23] Y. Zhang, H. C. Ji, R. Venkataramanan, and M. Mondelli, "Spectral Estimators for Structured Generalized Linear Models via Approximate Message Passing", *arXiv:2308.14507*, 2023.

# Optimal spectral methods for general Gaussian designs

**Theorem [ZJVM23]**

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 $\mathcal{T}^*(\Sigma)$.

# Optimal spectral methods for general Gaussian designs

**Theorem [ZJVM23]**

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 $\Sigma$ only via its **trace** (which can be estimated consistently).

# Optimal spectral methods for general Gaussian designs

**Theorem [ZJVM23]**

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 $\Sigma$ only via its **trace** (which can be estimated consistently).
- This **proves conjecture** of [MKLZ22] for class of spectral distributions of $\boldsymbol{A}$.

# Optimal spectral methods for general Gaussian designs

**Theorem [ZJVM23]**

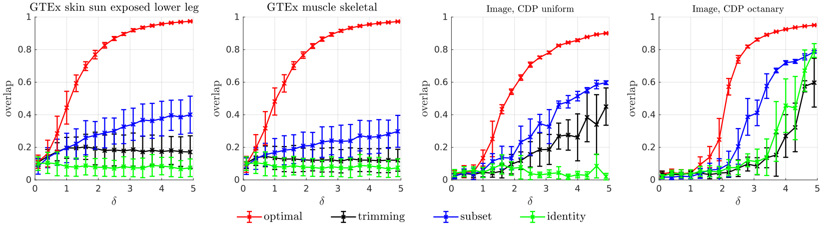
If $\delta > \delta^*(\Sigma)$, then there is $\mathcal{T}^*(\Sigma)$ s.t. spectral estimator works. Otherwise, under an additional technical assumption, there is no $\mathcal{T}$ s.t. $F(\delta, \Sigma, \mathcal{T}) > 0$.

Explicit expressions for $\delta^*(\Sigma)$ and $\mathcal{T}^*(\Sigma)$.

- $\mathcal{T}^*(\Sigma)$ depends on $\Sigma$ only via its **trace** (which can be estimated consistently).
- This **proves conjecture** of [MKLZ22] for class of spectral distributions of $\boldsymbol{A}$.
- $\delta^*(\Sigma)$ **meets information-theoretic** weak recovery **limit** conjectured in [MLKZ20].

# Universality of the optimal $\mathcal{T}^*(\Sigma)$

- $\boldsymbol{x}$ uniform on the sphere.
- $\boldsymbol{A}$ taken from datasets popular in quantative genetics (GTEx) and computational imaging (CDP).
- $y_i = |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle|$.




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

# Challenges

Mixed GLMs: we characterize eigenvalues with free probability tools, but unclear how to study eigenvectors...

Structured GLMs: unclear even how to characterize eigenvalues...

# Challenges and ideas

Mixed GLMs: we characterize eigenvalues with free probability tools, but unclear how to study eigenvectors...

Structured GLMs: unclear even how to characterize eigenvalues...

**New strategy** to analyze spectral methods based on **Approximate Message Passing (AMP)**

# AMP as a power method

Choose $f_t, g_t$ so that

$$\boldsymbol{v}^{t+1} \propto \boldsymbol{D}_n \boldsymbol{v}^t, \qquad \text{with } \boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}}$$

- As $t \to \infty$, $\boldsymbol{v}^t$ aligned with $\mathrm{eig}(\boldsymbol{D}_n)$ if $\boldsymbol{D}_n$ has a **spectral gap**

Mixed GLMs: we characterize eigenvalues with free probability tools, but unclear how to study eigenvectors...

# AMP as a power method

Choose $f_t, g_t$ so that

$$\boldsymbol{v}^{t+1} \propto \boldsymbol{D}_n \boldsymbol{v}^t, \qquad \text{with } \boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}}$$

- As $t \to \infty$, $\boldsymbol{v}^t$ aligned with $\mathrm{eig}(\boldsymbol{D}_n)$ if $\boldsymbol{D}_n$ has a **spectral gap**

Mixed GLMs: we characterize eigenvalues ($\Rightarrow$ spectral gap) with free probability tools and **eigenvectors via AMP**

# AMP as a power method

Choose $f_t, g_t$ so that

$$\boldsymbol{v}^{t+1} \propto \boldsymbol{D}_n \boldsymbol{v}^t, \qquad \text{with } \boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i \boldsymbol{a}_i^\mathsf{T}$$

- As $t \to \infty$, $\boldsymbol{v}^t$ aligned with $\mathrm{eig}(\boldsymbol{D}_n)$ if $\boldsymbol{D}_n$ has a **spectral gap**

Mixed GLMs: we characterize eigenvalues ($\Rightarrow$ spectral gap) with free probability tools and **eigenvectors via AMP**

Structured GLMs: unclear even how to characterize eigenvalues...

# AMP as a power method

Choose $f_t, g_t$ so that

$$\boldsymbol{v}^{t+1} \propto \boldsymbol{D}_n \boldsymbol{v}^t, \qquad \text{with } \boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}}$$

- As $t \to \infty$, $\boldsymbol{v}^t$ aligned with $\mathrm{eig}(\boldsymbol{D}_n)$ if $\boldsymbol{D}_n$ has a **spectral gap**

Mixed GLMs: we characterize eigenvalues ($\Rightarrow$ spectral gap) with free probability tools and **eigenvectors via AMP**

Structured GLMs: we characterize $\ell_2$-**norm of AMP iterates** to unveil spectral gap

# AMP iterates

$$\boldsymbol{v}^{t+1} = \frac{\boldsymbol{D}_n}{\gamma}\boldsymbol{v}^t + \boldsymbol{e}^t$$

- $\lim_{t\to\infty}\lim_{d\to\infty}\|\boldsymbol{e}^t\|^2/d = 0$

# AMP iterates

$$\boldsymbol{v}^{t+1} = \frac{\boldsymbol{D}_n}{\gamma}\boldsymbol{v}^t$$

# AMP iterates

$$\boldsymbol{v}^{t+t'} = \left( \frac{\boldsymbol{D}_n}{\gamma} \right)^{t'} \boldsymbol{v}^t$$

# $\ell_2$-norm of AMP iterates

$$\frac{1}{d}\|\boldsymbol{v}^{t+t'}\|^2 = \frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{v}^t\right\|^2$$

# $\ell_2$-norm of AMP iterates

$$\frac{1}{d}\|\boldsymbol{v}^{t+t'}\|^2 = \frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{v}^t\right\|^2$$

$$= \left(\lambda_1\left(\frac{\boldsymbol{D}_n}{\gamma}\right)\right)^{2t'}\frac{\langle \boldsymbol{v}_1(\boldsymbol{D}_n), \boldsymbol{v}^t\rangle^2}{d}$$

$$+ \frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{\Pi}\boldsymbol{v}^t\right\|^2$$

- $\boldsymbol{\Pi}$ projector orthogonal to top eigenvector

# $\ell_2$-norm of AMP iterates

$$\begin{aligned}\frac{1}{d}\|\boldsymbol{v}^{t+t'}\|^2 &= \frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{v}^t\right\|^2\\ &= \left(\lambda_1\left(\frac{\boldsymbol{D}_n}{\gamma}\right)\right)^{2t'}\frac{\langle \boldsymbol{v}_1(\boldsymbol{D}_n), \boldsymbol{v}^t\rangle^2}{d}\\ &\quad + \frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{\Pi}\boldsymbol{v}^t\right\|^2\end{aligned}$$

$$\frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{\Pi}\boldsymbol{v}^t\right\|^2 \le \left(\lambda_2\left(\frac{\boldsymbol{D}_n}{\gamma}\right)\right)^{2t'}\frac{1}{d}\|\boldsymbol{v}^t\|^2$$

- $\boldsymbol{\Pi}$ projector orthogonal to top eigenvector

# $\ell_2$-norm of AMP iterates

$$\lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{d}\|\boldsymbol{v}^{t+t'}\|^2 = \frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{v}^t\right\|^2$$

$$= \lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\left(\lambda_1\left(\frac{\boldsymbol{D}_n}{\gamma}\right)\right)^{2t'}\frac{\langle \boldsymbol{v}_1(\boldsymbol{D}_n), \boldsymbol{v}^t\rangle^2}{d}$$

$$+ \lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{\Pi}\boldsymbol{v}^t\right\|^2$$

$$\frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{\Pi}\boldsymbol{v}^t\right\|^2 \le \left(\lambda_2\left(\frac{\boldsymbol{D}_n}{\gamma}\right)\right)^{2t'}\frac{1}{d}\|\boldsymbol{v}^t\|^2$$

- $\lim_{d\to\infty}$ allows to apply **state evolution**
- $\lim_{t\to\infty}$ gives the **fixed point**
- $\lim_{t'\to\infty}$ boosts the **spectral gap**

# $\ell_2$-norm of AMP iterates

$$\lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{d}\|\boldsymbol{v}^{t+t'}\|^2 = \frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{v}^t\right\|^2$$

$$= \lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\left(\lambda_1\left(\frac{\boldsymbol{D}_n}{\gamma}\right)\right)^{2t'}\frac{\langle \boldsymbol{v}_1(\boldsymbol{D}_n), \boldsymbol{v}^t\rangle^2}{d}$$

$$+ \lim_{t'\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{\Pi}\boldsymbol{v}^t\right\|^2$$

$$\frac{1}{d}\left\|\left(\frac{\boldsymbol{D}_n}{\gamma}\right)^{t'}\boldsymbol{\Pi}\boldsymbol{v}^t\right\|^2 \le \left(\lambda_2\left(\frac{\boldsymbol{D}_n}{\gamma}\right)\right)^{2t'}\frac{1}{d}\|\boldsymbol{v}^t\|^2 \to 0$$

Provided that $\lim_{d\to\infty}\lambda_2(\boldsymbol{D}_n) < \gamma$

# Proof strategy

1. Guess the correct value of $\gamma$

# Proof strategy

1. Guess the correct value of $\gamma$
2. Compute edge of the bulk and verify that $\lim_{d\to\infty} \lambda_2(\boldsymbol{D}_n) < \gamma$

# Proof strategy

1. Guess the correct value of $\gamma$
2. Compute edge of the bulk and verify that $\lim_{d\to\infty} \lambda_2(\boldsymbol{D}_n) < \gamma$
3. Deduce that

$$\lim_{t'\to\infty} \lim_{t\to\infty} \lim_{d\to\infty} \left(\lambda_1\left(\frac{\boldsymbol{D}_n}{\gamma}\right)\right)^{2t'} \frac{\langle \boldsymbol{v}_1(\boldsymbol{D}_n), \boldsymbol{v}^t\rangle^2}{d} = \lim_{t'\to\infty} \lim_{t\to\infty} \lim_{d\to\infty} \frac{1}{d}\|\boldsymbol{v}^{t+t'}\|^2 := \rho^2$$

4. Conclude that

$$\lim_{d\to\infty} \lambda_1(\boldsymbol{D}_n) = \gamma \text{ (outlier)}$$

$$\lim_{t\to\infty} \lim_{d\to\infty} \frac{\langle \boldsymbol{v}_1(\boldsymbol{D}_n), \boldsymbol{v}^t\rangle^2}{d} = \rho^2 \text{ (overlap)}$$

# Conclusions

Analysis based on AMP **broadly applicable**:

- Rotationally invariant designs
- Matrix estimation with heteroscedastic/correlated noise
- Universality of spectral estimators
- . . .

# Conclusions

Analysis based on AMP **broadly applicable**:

- Rotationally invariant designs
- Matrix estimation with heteroscedastic/correlated noise
- Universality of spectral estimators
- ...


Thank you for your attention!