Phase retrieval with random matrices

phase transitions and optimal spectral methods

Antoine Maillard

Joint work with Bruno Loureiro, Florent Krzakala, Yue M. Lu & Lenka Zdeborová

Matrices et Graphes Aléatoires (MEGA) – 16 Décembre 2022

Phase retrieval

Generalized Linear Model (GLM)

Real / Complex

$\beta = 1$ $\beta = 2$

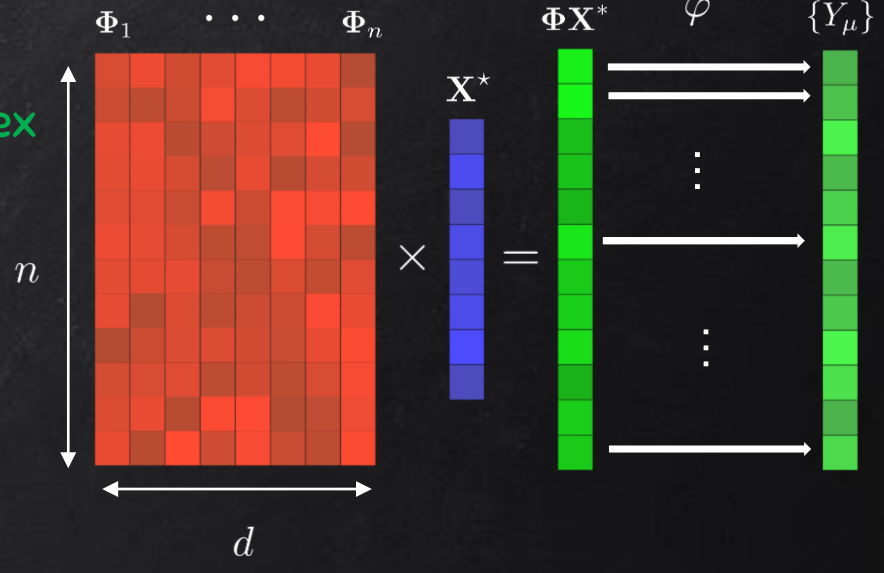
Observations

$$Y_\mu \sim P_{\text{out}}\left(\cdot \middle| \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \Phi_{\mu i} X_i^\star\right) \quad \mu \in \{1, \cdots, n\}$$

(Probabilistic) channel with possible noise.

Random sensing matrix

Signal



Phase retrieval: $P_{\text{out}}(y|z) = P_{\text{out}}(y||z|)$ E.g.

- Noiseless $Y_\mu = \frac{1}{d}|(\mathbf{\Phi X}^\star)_\mu|^2$
- Poisson-noise $Y_\mu \sim \text{Pois}(\Lambda|(\mathbf{\Phi X}^\star)_\mu|^2/d)$

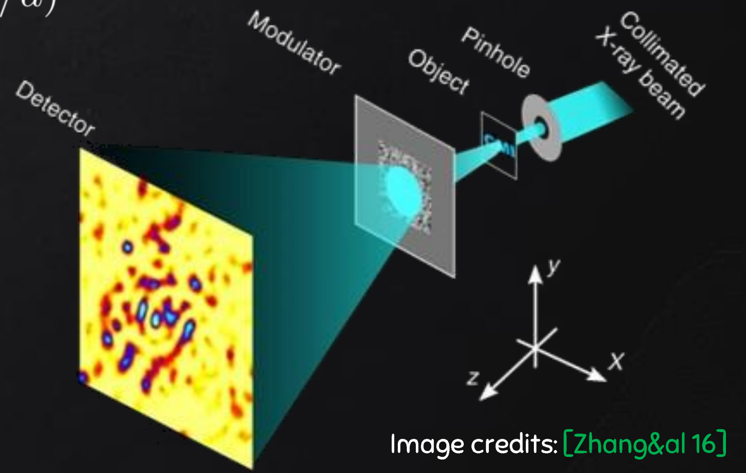


Image credits: [Zhang&al 16]

Random $\mathbf{\Phi}$

- Imaging in complex media
- Randomness can be implemented in practice !
- Deep analytical understanding

Goal : Fundamental limits of random phase retrieval in high dimension. $n, d \to \infty$

Setting

$$\left\{ Y_\mu \sim P_{\text{out}}\left(\cdot \,\Big|\, \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \Phi_{\mu i} X_i^\star \right) \right\}_{\mu=1}^{n}$$

- $\mathbf{\Phi}$ is right-orthogonally (unitarily) invariant: $\forall \mathbf{U},\ \mathbf{\Phi} \overset{d}{=} \mathbf{\Phi}\mathbf{U}$
- (a.s.) $\frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i(\mathbf{\Phi}^\dagger \mathbf{\Phi}/d)} \xrightarrow[d\to\infty]{\text{weakly}} \nu$ with compact support.
- $\lambda_{\max}(\mathbf{\Phi}^\dagger \mathbf{\Phi}/d) \leq C$ with probability $1 - o_d(1)$.

➢ Products of i.i.d. Gaussians matrices

➢ Haar-distributed column-unitary $\mathbf{\Phi}$

➢ Any $\mathbf{\Phi} \equiv \mathbf{U}\mathbf{S}\mathbf{V}^\dagger$ with $S_i^2 \overset{\text{i.i.d.}}{\sim} \nu$.

In the limit $d, n \to \infty$ with $\alpha = n/d = \Theta(1)$, what is the smallest α needed to recover $\mathbf{X}^\star$…

➢ Better than a random guess ? Weak recovery

➢ Perfectly ? Full recovery

➢ With which efficient algorithms ?

"Fundamental limits of phase retrieval"

Part I : Optimal errors

[M., Loureiro, Krzakala, Zdeborová '20]

$$\mathbb{P} = P_1^{\otimes d} \qquad \int P_1(\mathrm{d}x)\, x^2 = \rho$$

$$P_{\mathrm{out}}\Big(Y_\mu \Big| \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \Phi_{\mu i} X_i^\star\Big)$$

Bayes' law $\quad \mathbb{P}(\mathbf{x}|\mathbf{Y}, \boldsymbol{\Phi}) = \dfrac{\mathbb{P}(\mathbf{x})\mathbb{P}(\mathbf{Y}|\mathbf{x}, \boldsymbol{\Phi})}{\mathbb{P}(\mathbf{Y}|\boldsymbol{\Phi})}$

Random distribution

Prop: Minimal Mean Squared Error estimator $\hat{\mathbf{X}}_{\mathrm{MMSE}}(\mathbf{Y}, \boldsymbol{\Phi}) \equiv \underset{\hat{\mathbf{X}}(\mathbf{Y}, \boldsymbol{\Phi})}{\arg\min}\ \mathbb{E}\|\hat{\mathbf{X}}(\mathbf{Y}, \boldsymbol{\Phi}) - \mathbf{X}^\star\|^2 = \mathbb{E}[\mathbf{x}|\mathbf{Y}, \boldsymbol{\Phi}]$

Conjecture ("Replica formula"): There is a scalar var. formula

$$f = \sup_{q_x, q_z} \Big[\overbrace{\underbrace{I_0^{(\beta)}(q_x)}_{P_1} + \underbrace{I_{\mathrm{out}}^{(\beta)}(q_x, q_z)}_{P_{\mathrm{out}}} + \underbrace{\beta I_{\mathrm{int}}(q_x, q_z)}_{\nu}}^{\text{"Replica-symmetric" potential } f(q_x, q_z)}\Big]$$

ν (asymptotic spectrum of $\boldsymbol{\Phi}$)

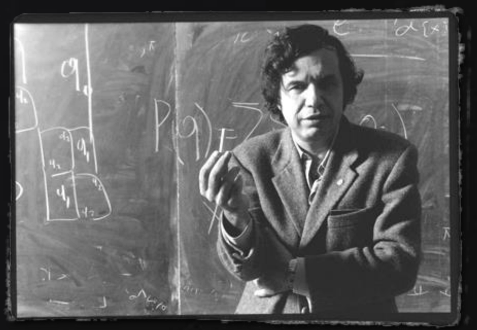
such that the MMSE is $\quad \lim_{d \to \infty} \dfrac{1}{d} \mathbb{E}\, \|\hat{\mathbf{X}}_{\mathrm{MMSE}} - \mathbf{X}^\star\|^2 = \rho - q_x^\star$

- $(I_0, I_{\mathrm{out}}, I_{\mathrm{int}})$ are fully explicit.
- I_{int} is related to a rank-one spherical integral.

$$\lim_{d \to \infty} \frac{1}{d} \log \int_{\mathcal{S}^{n-1} \times \mathcal{S}^{d-1}} \mu_n(\mathrm{d}\mathbf{e})\, \mu_d(\mathrm{d}\mathbf{f})\, \exp\{\theta \sqrt{d}\, \mathbf{e}^\dagger \boldsymbol{\Phi} \mathbf{f}\}$$

Obtained with the heuristic replica method of statistical physics.

2021



Rigorous statistical limits

Conjecture ("Replica formula") : $f = \sup_{q_x, q_z} f(q_x, q_z) \implies \lim_{d\to\infty} \frac{1}{d} \mathbb{E}\,\|\hat{\mathbf{X}}_{\mathrm{MMSE}} - \mathbf{X}^\star\|^2 = \rho - q_x^\star$

Theorem (informal) : If either

a) $\Phi_{\mu i}$ independent and $\mathbb{E}[\Phi_{\mu i}] = 0 \;\; \mathbb{E}[|\Phi_{\mu i}|^2] = 1 \;\; \mathbb{E}[|\Phi_{\mu i}|^3] \leq C$

b) P_1 is Gaussian and $\Phi = \mathbf{W}\mathbf{B}$

i.i.d. Gaussian matrix "Any" matrix

, the replica conjecture stands.

- Previously known for real i.i.d. matrices [Barbier&al '19] $\implies$ We tackle correlated data $\mathbf{\Phi}$!
- Main proof idea: Use the replica conjecture to build a "smart" Gaussian interpolation path.

Algorithmic limits

$$f = \sup_{q_x, q_z} f(q_x, q_z) \Longrightarrow \lim_{d\to\infty} \frac{1}{d} \mathbb{E}\,\|\hat{\mathbf{X}}_{\mathrm{MMSE}} - \mathbf{X}^\star\|^2 = \rho - q_x^\star$$

Statistical limits ✓

What about efficient (polynomial-time) algorithms ?

Best-known class of algorithms: AMP *(Approximate Message Passing)* [Donoho&al '11]

In rotationally-invariant models: G-VAMP (Generalized Vector AMP) [Schniter&al '16]

Theorem: AMP is optimal among the class of "generalized first order methods" for i.i.d. Gaussian $\mathbf{\Phi}$.

[Celentano & al '22]

Conjecture: AMP optimality holds
- among all polynomial-time algorithms
- for rotationally-invariant models

Strong numerical and analytical evidence
[Krzakala & Zdeborova '16, M.&al '19, ...]

Key theorem (informal): For any rotationally-invariant $\mathbf{\Phi}$: $\lim_{d\to\infty} \mathrm{MSE}_{\mathrm{AMP}} = \rho - q_x^{(\mathrm{local})}$

$q_x^{(\mathrm{local})}$ obtained by local optimization of $f(q_x, q_z)$ starting from $q_x = q_z = 0$.

[Fletcher, Rangan & Schniter '18]

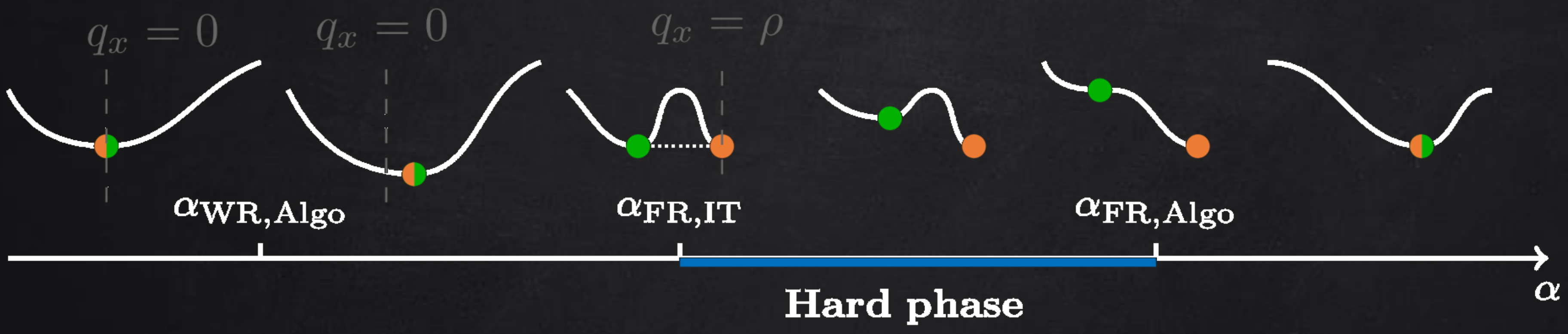
Statistical vs algorithmic

$$f = \sup_{q_x, q_z} f(q_x, q_z) \Longrightarrow \lim_{d\to\infty} \frac{1}{d} \mathbb{E}\, \|\hat{\mathbf{X}}_{\mathrm{MMSE}} - \mathbf{X}^\star\|^2 = \rho - q_x^\star$$

Key theorem (informal): For any rotationally-invariant Φ : $\lim_{d\to\infty} \mathrm{MSE}_{\mathrm{AMP}} = \rho - q_x^{(\mathrm{local})}$

$q_x^{(\mathrm{local})}$ obtained by local optimization of $f(q_x, q_z)$ starting from $q_x = q_z = 0$.

We can investigate "computational-to-statistical" gaps by studying the landscape of $f(q_x, q_z)$



: global minimum of $-f$

: local minimum closest to $q_x = 0$

Parameters of the problem

- Noise
- Structure of Φ
- ...

Analytical toolbox

Influences

Statistical and algorithmic limits of phase retrieval

Application: fundamental thresholds

$n/d \to \alpha \qquad \frac{1}{d}\mathrm{Tr}\left[f\left(\frac{\Phi^\dagger \Phi}{d}\right)\right] \to \langle f(\lambda)\rangle_\nu$

Algorithmic weak-recovery

wlog $\langle \lambda \rangle_\nu = \alpha$

Stability analysis of the replica-symmetric potential.

$$\alpha_{\mathrm{WR,Algo}} = \langle \lambda^2 \rangle_\nu \times \frac{Z}{1+Z}$$

$$Z \equiv \int_{\mathbb{R}} \mathrm{d}y \frac{\left(\int_{\mathbb{K}} \mathcal{D}_\beta z (|z|^2 - 1) P_{\mathrm{out}}(y|\sqrt{\rho} z)\right)^2}{\int_{\mathbb{K}} \mathcal{D}_\beta z \, P_{\mathrm{out}}(y|\sqrt{\rho} z)}$$

This is a
- Theorem for AMP.
- Conjecture for polytime algorithms.

For any noise, the highest weak recovery threshold is reached by column-unitary matrices.

Information-theoretic full recovery

$$\mathbf{Y}_\mu = \frac{1}{d}|(\Phi \mathbf{X}^\star)_\mu|^2 \quad + \quad P_1 = \mathcal{N}(0,1)$$

Global maximum of the replica-symmetric potential

$$\alpha_{\mathrm{FR,IT}} = \beta(1 - \nu(\{0\}))$$

"Asymptotic rank of Φ "

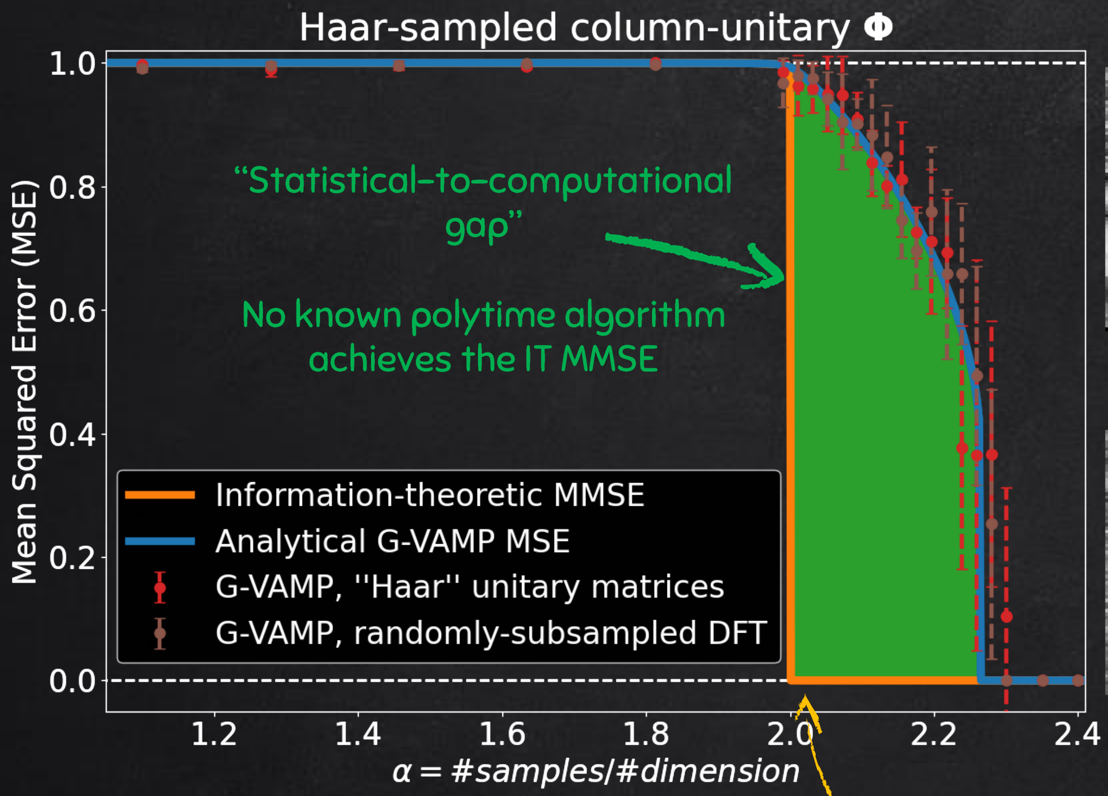
This is a
- Theorem for i.i.d. Φ or $\Phi = \mathbf{WB}$
- Conjecture for rotationally-invariant Φ .

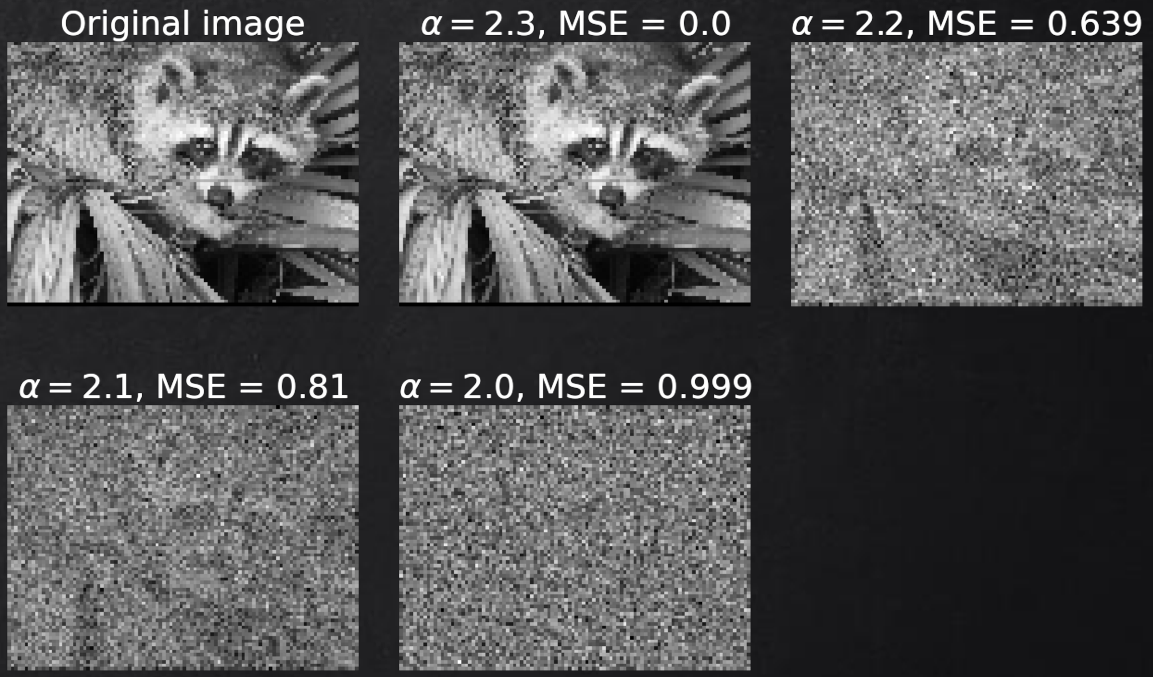
Does not depend on the precise statistics of Φ

Proving / interpreting these predictions ?

Numerical application (I)

$$\mathbf{Y}_\mu = \frac{1}{d}|(\mathbf{\Phi}\mathbf{X}^\star)_\mu|^2 \quad + \quad P_1 = \mathcal{N}(0,1)$$



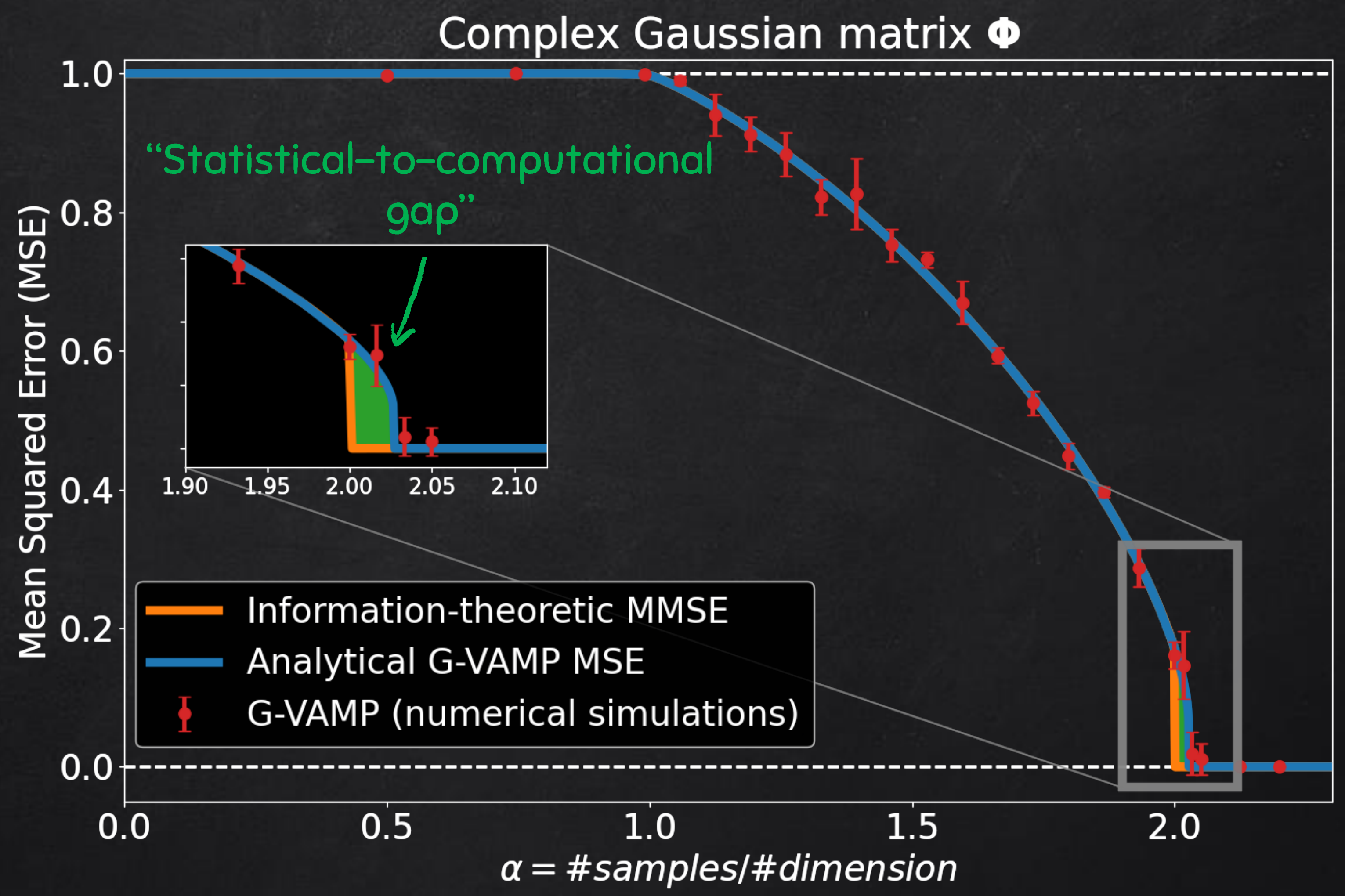


$\alpha_{\mathrm{WR,IT}} = \alpha_{\mathrm{FR,IT}} = 2$: "all-or-nothing" IT transition.

✓ AMP matches analytical predictions, even with a natural image !

✓ Matrices with controlled structure still perform very well !

Numerical application (II)



Numerical application (III)

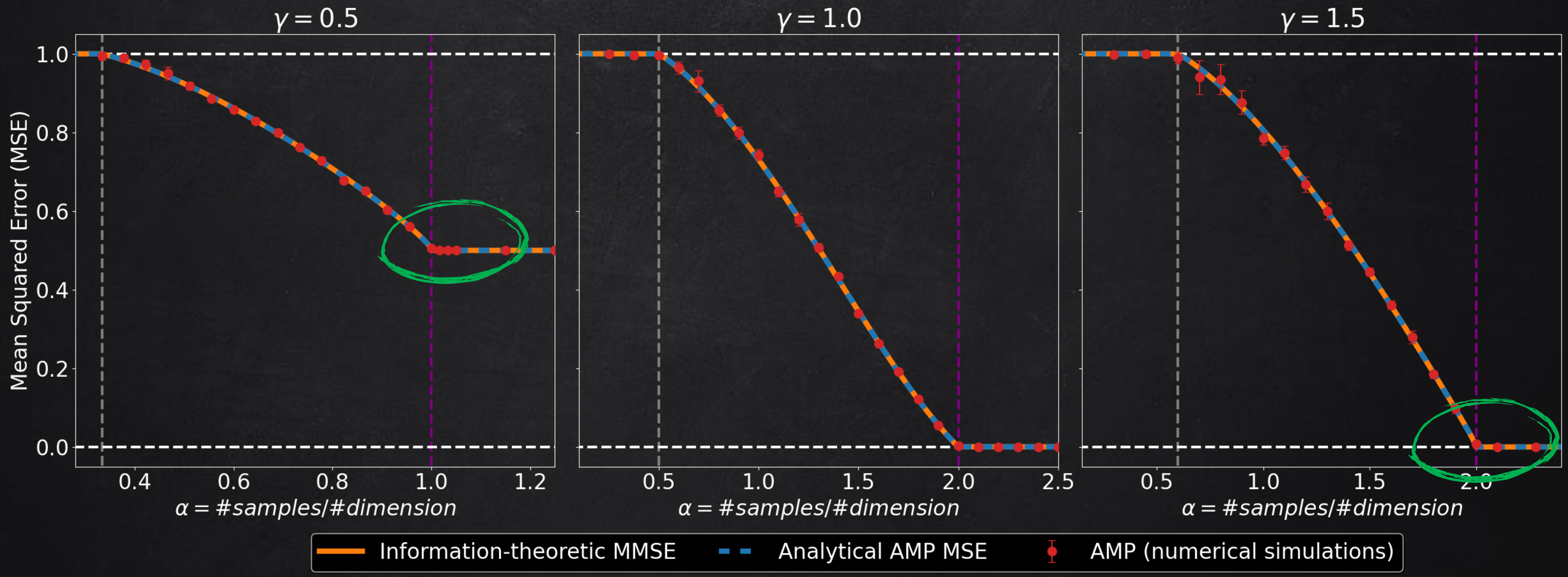
Product of i.i.d. Gaussians

$$\Phi = \mathbf{W}_1 \mathbf{W}_2$$

$n \times p$ $\quad$ $p \times d$

$n/d \to \alpha$

$p/d \to \gamma$



Small "statistical-to-computational gap" for $\gamma \neq 1$, disappears for $\gamma = 1$.

Zoology of statistical-to-computational gaps as a function of the spectrum

Part II : Cheaper algorithms?

[M., Lu, Krzakala, Zdeborová '21]

- Semidefinite relaxations
- Non-convex optimization procedures
- Approximate Message-Passing (Part I)

[Candès&al '15, Waldspurger&al'15, ...]

Computationally heavy / Need informed initialization → Spectral methods

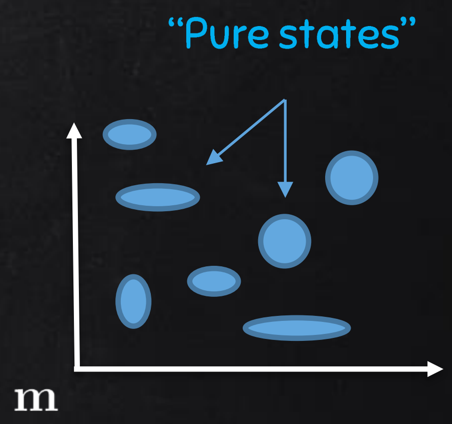
[Mondelli&al '18, Luo&al '18, Dudeja&al '19,...]

This talk: Two strategies related to the statistical physics approach to high-dimensional inference.

- Method I : Linearization of message-passing (AMP) algorithms.
- Method II : Bethe Hessian analysis from the Thouless-Anderson-Palmer [TAP77] free energy.

TAP free energy

Thouless–Anderson–Palmer decomposition [TAP77]



- $\mathbb{P}(\mathbf{x}|\mathbf{Y})$ decomposes along "pure states".
- Pure states found by "tilting" the measure, imposing $m_i = \mathbb{E}[x_i|\mathbf{Y}]$ and $\sigma_{ij} = \mathrm{Cov}[x_i, x_j|\mathbf{Y}]$.

  They are the maxima of the free entropy of this constrained measure, as a function of $(\mathbf{m}, \boldsymbol{\sigma})$.

Conjectured TAP free entropy for rotationally-invariant GLMs [Parisi&Potters '95], [M.&al '19]

$$f_{\mathrm{TAP}}(\mathbf{m}) = \sup_{\sigma \geq 0} \sup_{\substack{\mathbf{g} \in \mathbb{K}^n \\ r \geq 0}} \mathrm{extr}_{\substack{\boldsymbol{\omega} \in \mathbb{K}^n \\ b \geq 0}} \mathrm{extr}_{\substack{\boldsymbol{\lambda} \in \mathbb{K}^d \\ \gamma \geq 0}} \Bigg[\frac{\beta}{d} \sum_{i=1}^{d} \lambda_i \cdot m_i + \frac{\beta\gamma}{2d} \Big(d\sigma^2 + \sum_{i=1}^{d} |m_i|^2\Big) - \frac{\beta}{d} \sum_{\mu=1}^{n} \omega_\mu \cdot g_\mu - \frac{\beta b}{2d} \Big(\sum_{\mu=1}^{n} |g_\mu|^2 - \alpha d r\Big) + \frac{1}{d} \sum_{i=1}^{d} \ln \int_{\mathbb{K}} P_0(\mathrm{d}x) e^{-\frac{\beta\gamma}{2}|x|^2 - \beta\lambda_i \cdot x}$$

$$+ \frac{\alpha}{n} \sum_{\mu=1}^{n} \ln \int_{\mathbb{K}} \frac{\mathrm{d}h}{\left(\frac{2\pi b}{\beta}\right)^{\beta/2}} P_{\mathrm{out}}(y_\mu|h) e^{-\frac{\beta|h-\omega_\mu|^2}{2b}} + \frac{\beta}{d} \sum_{i=1}^{d} \sum_{\mu=1}^{n} g_\mu \cdot \Big(\frac{\Phi_{\mu i}}{\sqrt{d}} m_i\Big) + \beta F(\sigma^2, r) \Bigg].$$

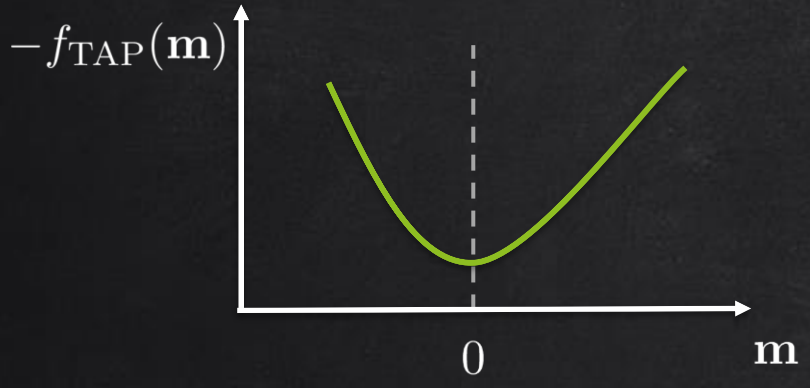
$$F(x,y) \equiv \inf_{\zeta_x, \zeta_y > 0} \left[\frac{\zeta_x x}{2} + \frac{\alpha \zeta_y y}{2} - \frac{\alpha - 1}{2} \ln \zeta_y - \frac{1}{2} \langle \ln(\zeta_x \zeta_y + \lambda) \rangle_\nu \right] - \frac{1}{2} \ln x - \frac{\alpha}{2} \ln y - \frac{1+\alpha}{2}.$$

Involved but explicit!

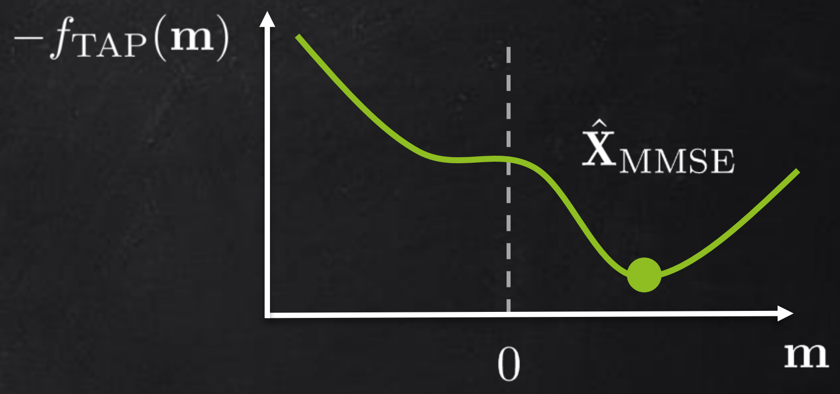
The Bethe Hessian

Landscape of the TAP free energy

Recovery impossible



Recovery possible



$$\underbrace{\text{Spectral methods can only use the local information available around the "starting" point } \mathbf{m} = 0 \text{ .}}_{\hat{\mathbf{X}}_{\mathrm{TAP}} \triangleq \mathbf{v}_{\max}\{\nabla^2 f_{\mathrm{TAP}}(\mathbf{m}=0)\}}$$

"The Bethe Hessian represents the local information available before doing any inference."

Our conjecture: $\hat{\mathbf{X}}_{\mathrm{TAP}} = \underset{\hat{\mathbf{X}}=\hat{\mathbf{X}}_{\mathrm{spectral}}(\mathbf{Y},\mathbf{\Phi})}{\arg\min} \mathbb{E}\|\hat{\mathbf{X}} - \mathbf{X}^\star\|^2$

Similar to previous strategies in community detection. [Saade&al '14]

Main conjecture on spectral methods

wlog $\langle \lambda \rangle_\nu = \alpha$

Computation of the Bethe Hessian

$$\mathbf{M}_{\mathrm{TAP}} = \mathbf{M}(\mathcal{T}^\star) = \frac{1}{d}\sum_{\mu=1}^{n} \mathcal{T}^\star(Y_\mu)\mathbf{\Phi}_\mu \mathbf{\Phi}_\mu^\dagger \qquad \mathcal{T}^\star(y) = \rho^{-1} - \left(\frac{\int_{\mathbb{K}} \mathrm{d}z |z|^2 \, e^{-\frac{\beta|z^2|}{2\rho}} P_{\mathrm{out}}(y|z)}{\int_{\mathbb{K}} \mathrm{d}z \, e^{-\frac{\beta|z^2|}{2\rho}} P_{\mathrm{out}}(y|z)}\right)^{-1}$$

$\mathbf{M}_{\mathrm{TAP}}$ does not depend on the spectrum of the sensing matrix !

Practical consequence: Only needs to know the observation channel to construct the optimal method!

Coherent with previous literature: Theorem [Lu&al '19] For i.i.d. Gaussian $\mathbf{\Phi}$, $\mathcal{T}^\star = \arg\max_{\mathcal{T}}[\mathrm{MSE}(\mathbf{M}(\mathcal{T}))]$

E.g. Optimal method for all noiseless phase retrieval models with rot-inv matrices:

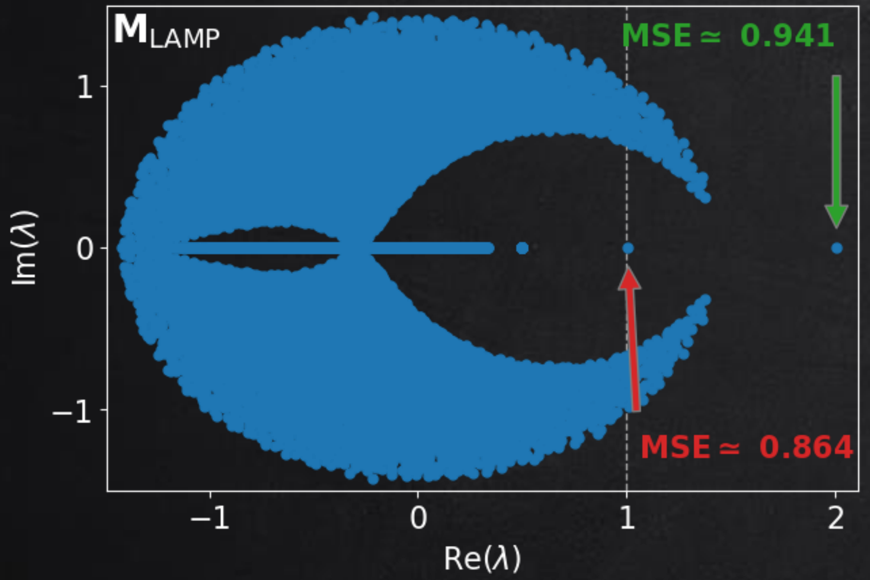
$$P_{\mathrm{out}}(y|z) = \delta(y - |z|^2) \implies \mathcal{T}^\star(y) = 1 - y^{-1}$$

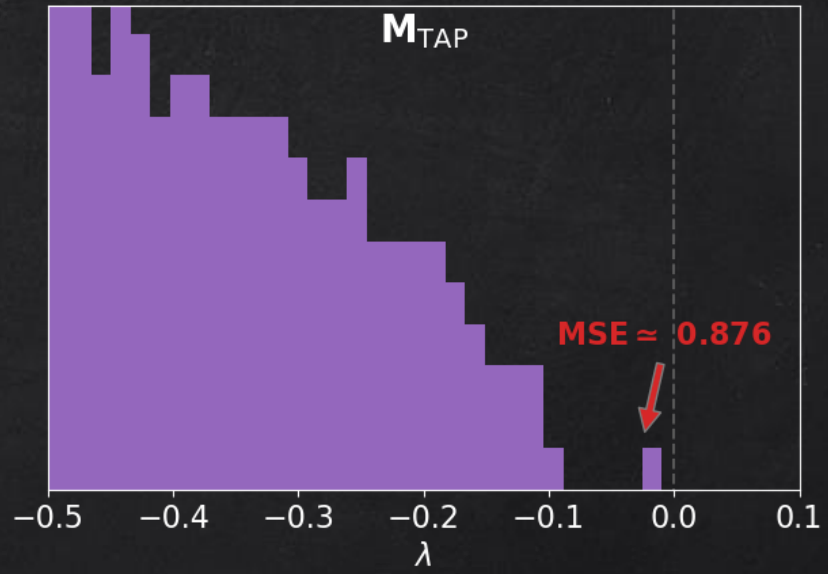
Linearized Approximate Message Passing

Linearization of AMP around initialization

$$\mathbf{M}_{\mathrm{LAMP}} = \left(\frac{\mathbf{\Phi}\mathbf{\Phi}^\dagger}{d} - \mathrm{I}_d\right)\mathrm{Diag}(\{\mathcal{T}^\star(Y_\mu)/[\rho^{-1} + \mathcal{T}^\star(Y_\mu)]\})$$

$$\mathbf{M}_{\mathrm{TAP}} = \mathbf{M}(\mathcal{T}^\star) = \frac{1}{d}\sum_{\mu=1}^{n} \mathcal{T}^\star(Y_\mu)\mathbf{\Phi}_\mu\mathbf{\Phi}_\mu^\dagger$$





Optimal MSE : marginal stability

Theorem: as $d \to \infty$,

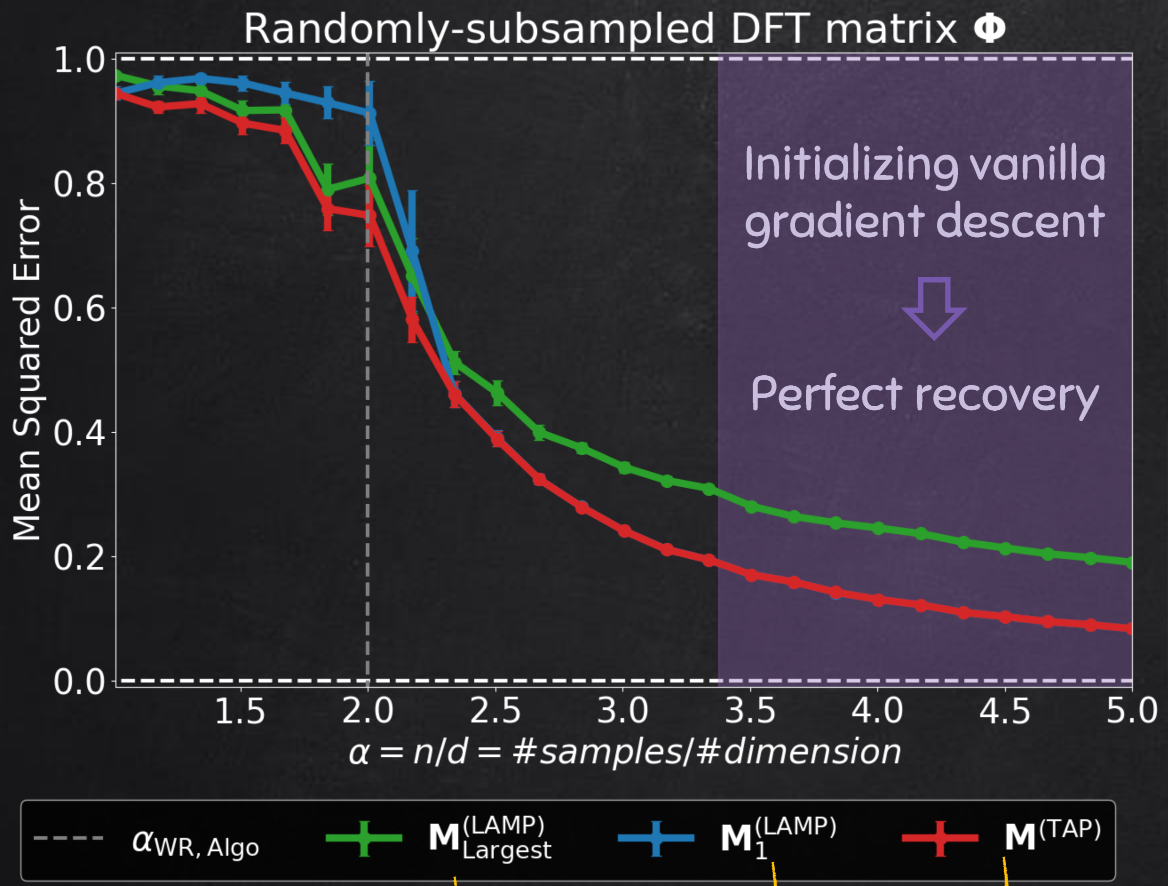
- $1 \in \mathrm{Sp}(\mathbf{M}_{\mathrm{LAMP}}) \Leftrightarrow 0 \in \mathrm{Sp}(\mathbf{M}_{\mathrm{TAP}})$
- These estimators are equivalent.

Puzzle: why is the dominant eigenvector of $\mathrm{M}_{\mathrm{LAMP}}$ a suboptimal estimator ?

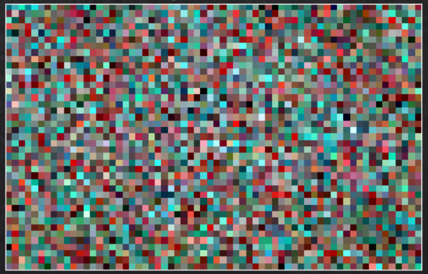
AMP and TAP are fundamentally equivalent in rotationally-invariant models [M.&al '19]

Similar remarks in community detection [Dall'Amico&al ' 19, 21].

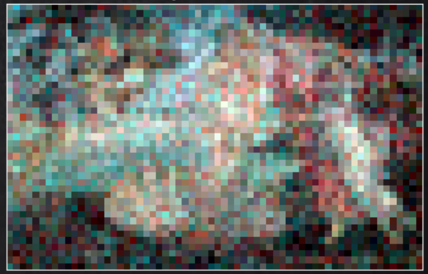
Spectral methods performance



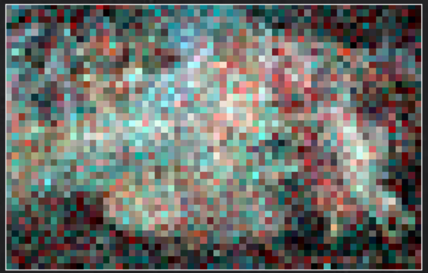
TAP, $\alpha = 1.5$



TAP, $\alpha = 3.5$



LAMP, $\alpha = 3.5$



Original image

Suboptimal

Two equivalent estimators

- Achieves <u>optimal weak recovery</u>
- Combined with gradient descent: efficient and cheap procedure!

Conclusion

(new results in red)

I

Fundamental limits of phase retrieval

(Complex) matrix ensemble	Fundamental limits
i.i.d. $\mathbf{\Phi}$, symm. P_{out}	Theorem
$\mathbf{\Phi} = \mathbf{WB}$, Gauss P_1, symm. P_{out}	Theorem
Rot-inv $\mathbf{\Phi}$, symm. P_{out}	Conjecture (stat. limits) / Theorem (alg. limits)

II

Spectral methods

- Constructive derivation of a conjecturally optimal spectral method in generic phase retrieval.
- Results apply to randomly subsampled DFT matrices and to real image recovery.

Universality of rot-inv AMP & spectral methods under sign and permutation invariance [Dudeja&al '20, Dudeja&al '22, Wang&al '22]

THANK YOU !