Phase retrieval in high dimensions : Statistical and computational phase transitions

A.M, Bruno Loureiro, Florent Krzakala, Lenka Zdeborová

arXiv:2006.05228

ENS | PSL

FONDATION CFM POUR LA RECHERCHE

IPhT SACLAY

1 Phase retrieval as a GLM

Generalized Linear Model (GLM)

Observations $Y_\mu \in \mathbb{R}$

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

(Probabilistic) channel

Sensing matrix (real/complex)

Signal (real/complex), n-dimensional

In phase retrieval, one only measures the modulus : $P_{\rm out}(\cdot|z) = P_{\rm out}(\cdot||z|)$

Classical problem, non-trivial even in the noiseless case $Y_\mu = |(\mathbf{\Phi X}^\star)_\mu|^2/n$, many algorithms :

- SDP relaxations [Candès&al '11, Candès&al '12, Waldspurger&al '12, Goldstein&al '16, ...]
- Non-convex optimization procedures [Netrapalli&al '13, Candès&al '14, Gerchberg 1972, ...]
- Spectral methods [Mondelli&al '18, Luo&al '18, Dudeja&al '19, ...]

Goal : Fundamental limits of phase retrieval with random sensing matrices and random signal in the typical case and in high dimensions.

⚠ Different from the injectivity studies of the "worst-case" [Bandeira&al '13]

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

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

- Better than a random guess ?
- Perfectly ? (up to the possible rank deficiency of Φ)
- With which (polynomial-time) algorithm ?

Our model : i) $\mathbf{X}^\star$ and $\mathbf{\Phi}$ can be real ($\beta = 1$) or complex ($\beta = 2$)

ii) The signal $\mathbf{X}^\star$ is generated using a (known) i.i.d. prior distribution P_0 and $\mathrm{Var}_{P_0}(X^\star) = \rho > 0$

iii) The matrix $\mathbf{\Phi}$ is right-orthogonally (unitarily) invariant : $\forall \mathbf{U},\ \mathbf{\Phi} \overset{d}{=} \mathbf{\Phi U}$

iv) The empirical spectral distribution of $\mathbf{\Phi}^\dagger \mathbf{\Phi}/n$ converges :

$$\nu_n \equiv \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i\left(\frac{\mathbf{\Phi}^\dagger \mathbf{\Phi}}{n}\right)} \underset{n\to\infty}{\longrightarrow} \nu \in \mathcal{M}_1^+(\mathbb{R}_+).$$

Encompasses many models : Gaussians, product of Gaussians, random column-orthogonal/unitary, any $\mathbf{\Phi} \equiv \mathbf{USV}^\dagger$ with $S_i^2 \overset{\text{i.i.d.}}{\sim} \nu$

2 Optimal error in GLMs

We consider any channel (not necessarily phase retrieval)

Conjecture : Consider the following scalar optimization problem $f = \sup_{q_x, q_z} [\underbrace{I_0^{(\beta)}(q_x)}_{P_0} + \underbrace{I_{\text{out}}^{(\beta)}(q_z)}_{P_{\text{out}}} + \underbrace{\beta I_{\text{int}}(q_x, q_z)}_{\nu}]$

Then the *information-theoretic Minimal Mean Squared Error* is : $\text{MMSE} = \rho - q_x$

(The functions involved in the optimization problem are fully explicit)

Theorem : If either

a) $\Phi_{\mu i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_\beta(0, 1)$ (standard Gaussian distribution)

b) P_0 is Gaussian and $\Phi = \mathbf{W}\mathbf{B}$

Gaussian matrix   Any matrix

, the conjecture above stands.

- Conjecture obtained with the replica method of statistical physics [Parisi&al 1987, Takahashi&al '20]
- Proven using probabilistic interpolation methods [Guerra '03, Talagrand '07, Barbier&al '18, Barbier&al '19]

2' Optimal error in GLMs

We consider any channel (not necessarily phase retrieval)

$$f = \sup_{q_x, q_z} \underbrace{[I_0^{(\beta)}(q_x) + I_{\text{out}}^{(\beta)}(q_z) + \beta I_{\text{int}}(q_x, q_z)]}$$

"Replica-symmetric" potential $f(q_x, q_z)$

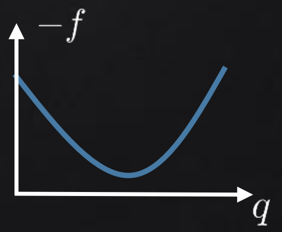
Strong conjecture : For GLMs, the optimal polynomial-time algorithm is an explicit iterative algorithm :

Approximate Message Passing. Called here G-VAMP *(Generalized Vector Approximate Message Passing).*

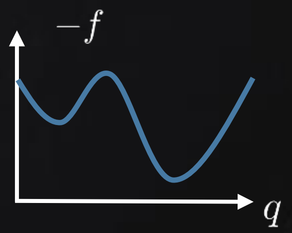
[Mézard '89, Donoho&al '09, Montanari&al '10 , Krzakala&al '11, Rangan&al '16, Schniter&al '16, ...]

Important result [Schniter&al '16] : The MSE of G-VAMP in the large n limit is given by running *gradient ascent* on the Replica-symmetric potential starting from $q_x = q_z = 0$ (random initialization).

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



No gap



Gap : "Hard" phase

3 (Algorithmic) Weak recovery

We consider phase retrieval : $P_{\mathrm{out}}(\cdot|z) = P_{\mathrm{out}}(\cdot||z|)$

What is the minimal number of measurements $\alpha = m/n$ necessary to beat a random guess in polynomial time ?

This threshold $\alpha_{\mathrm{WR,Algo}}$ is a solution of :

$$\boxed{\alpha = \frac{\langle\lambda\rangle_\nu^2}{\langle\lambda^2\rangle_\nu}\left[1 + \left\{\int_{\mathbb{R}} \mathrm{d}y \frac{\left(\int_{\mathbb{K}} \mathcal{D}_\beta z \, (|z|^2 - 1) \; P_{\mathrm{out}}\left[y\middle|\sqrt{\frac{\rho\langle\lambda\rangle_\nu}{\alpha}}z\right]\right)^2}{\int_{\mathbb{K}} \mathcal{D}_\beta z \; P_{\mathrm{out}}\left[y\middle|\sqrt{\frac{\rho\langle\lambda\rangle_\nu}{\alpha}}z\right]}\right\}^{-1}\right]}$$

- This is an implicit equation ⚠
- Derived by a stability analysis of the replica-symmetric potential.

For any phase retrieval channel and prior, the highest weak recovery threshold is reached by random column-orthogonal/unitary matrices (up to a scaling).

For noiseless phase retrieval : $\boxed{\alpha = \left(1 + \frac{\beta}{2}\right)\frac{\langle\lambda\rangle_\nu^2}{\langle\lambda^2\rangle_\nu}}$

- Gaussian matrices : $\alpha_{\mathrm{WR,Algo}} = \frac{\beta}{2}$ [Barbier&al '18, Mondelli &al '18]
- Random column-orthogonal/unitary matrices : $\alpha_{\mathrm{WR,Algo}} = 1 + \frac{\beta}{2}$ [Dudeja&al '19] for $\beta = 2$

4 Strong (full) recovery

We consider noiseless phase retrieval : $P_{\text{out}}(y|z) = \delta(y - |z|^2)$ and a Gaussian prior $P_0 = \mathcal{N}_\beta(0,1)$

How many measurements are necessary to be able to information-theoretically achieve the best possible recovery ?

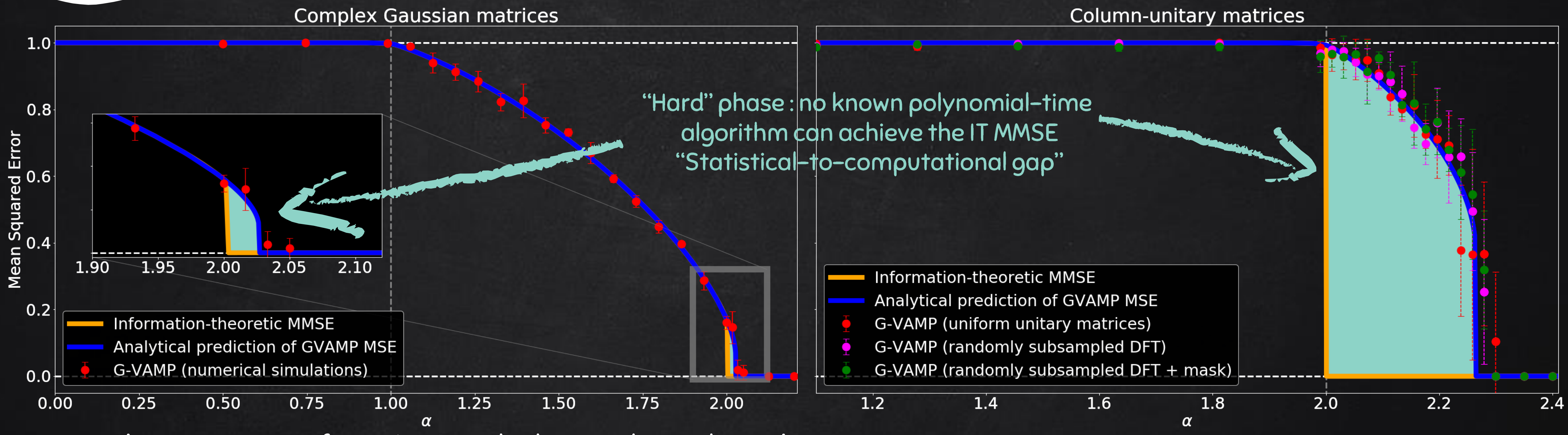
If $\frac{1}{n}\mathrm{rk}\left(\frac{\mathbf{\Phi}^\dagger \mathbf{\Phi}}{n}\right) \to r \in [0,1]$: $\boxed{\alpha_{\mathrm{FR,IT}} = \beta r}$ Does not depend on the precise statistics of $\mathbf{\Phi}$

For $\alpha \geq \alpha_{\mathrm{FR,IT}}$, the MMSE reaches a plateau with $\mathrm{MMSE}_{\mathbf{x}} = 1 - r$; $\mathrm{MMSE}_{\mathbf{\Phi x}} = 0$

- The real case $\alpha_{\mathrm{FR,IT}} = r$ can be derived by a counting argument [Candès&al, '05]
- The complex case $\alpha_{\mathrm{FR,IT}} = 2r$ can (as far as we know) only be derived by the replica-symmetric potential !

5 Applications

We consider noiseless phase retrieval : $P_{\mathrm{out}}(y|z) = \delta(y - |z|^2)$ and a Gaussian prior $P_0 = \mathcal{N}_\beta(0,1)$



Very good agreement of G-VAMP with the analytical predictions.

Some (funny) remarks :

- For column-unitary matrices $\alpha_{\mathrm{FR,IT}} = \alpha_{\mathrm{WR,Algo}} = 2$: "all-or-nothing" IT transition.
- For all other full-rank complex matrices $\alpha_{\mathrm{WR,Algo}} < \alpha_{\mathrm{FR,IT}}$
- For real matrices, there can be a large gap ! Ex : column-orthogonal matrices $\alpha_{\mathrm{FR,IT}} = 1 < \alpha_{\mathrm{WR,Algo}} = 3/2$
- Matrices with controlled structure (e.g. randomly subsampled DFT) still perform very well with G-VAMP !

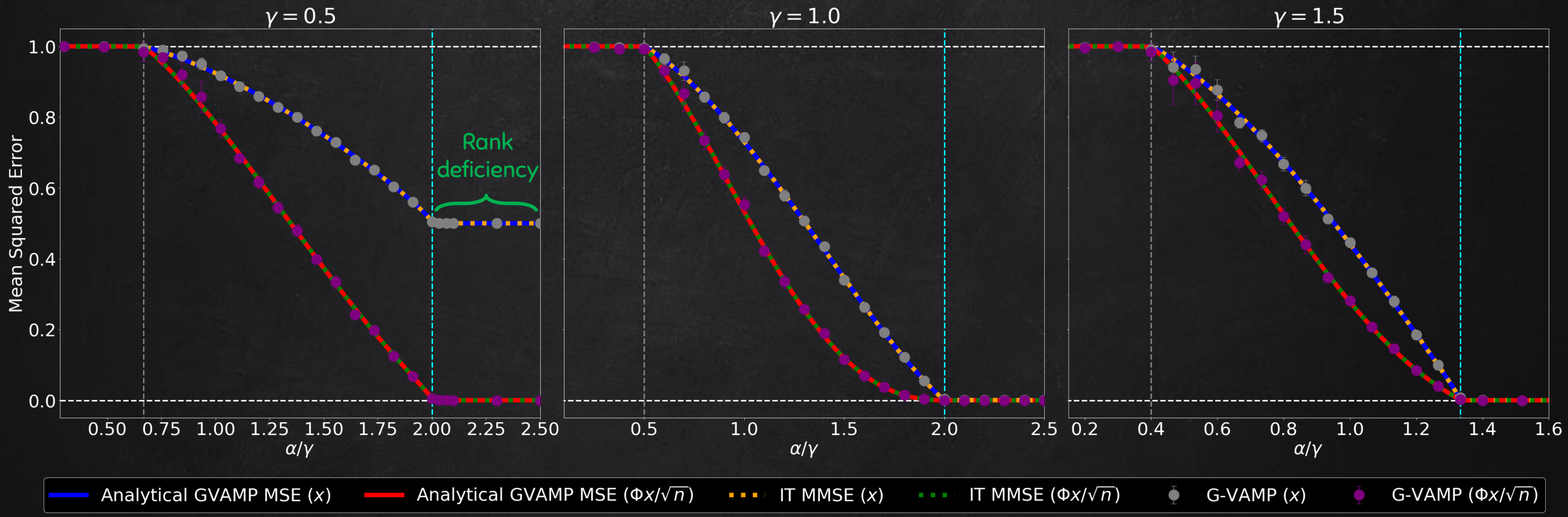
Conclusion / Summary (new results in red)

	Matrix ensemble and value of β	$\alpha_{\mathrm{WR,Algo}}$	$\alpha_{\mathrm{FR,IT}}$	$\alpha_{\mathrm{FR,Algo}}$
Noiseless phase retrieval with Gaussian prior	Real Gaussian $\mathbf{\Phi}$ ($\beta = 1$)	0.5	1	$\simeq 1.12$
	Complex Gaussian $\mathbf{\Phi}$ ($\beta = 2$)	1	2	$\simeq 2.027$
	Real column-orthogonal $\mathbf{\Phi}$ ($\beta = 1$)	1.5	1	$\simeq 1.584$
	Complex column-unitary $\mathbf{\Phi}$ ($\beta = 2$)	2	2	$\simeq 2.265$
	$\mathbf{\Phi} = \mathbf{W}_1 \mathbf{W}_2$ ($\beta = 1$, aspect ratio γ)	$\gamma/(2(1+\gamma))$	$\min(1, \gamma)$	Theorem
	$\mathbf{\Phi} = \mathbf{W}_1 \mathbf{W}_2$ ($\beta = 2$, aspect ratio γ)	$\gamma/(1+\gamma)$	$\min(2, 2\gamma)$	Theorem
	$\mathbf{\Phi}$, $\beta \in \{1, 2\}$, $\mathrm{rk}[\mathbf{\Phi}^\dagger \mathbf{\Phi}]/n = r$	Analytical expression	βr	Conjecture
Generic phase retrieval with any prior	Gauss. $\mathbf{\Phi}$, $\beta \in \{1, 2\}$, symm. P_0, P_{out}	Analytical expression	Theorem	Theorem
	$\mathbf{\Phi} = \mathbf{W}\mathbf{B}$, $\beta \in \{1, 2\}$, Gauss. P_0, symm. P_{out}	Analytical expression	Theorem	Theorem
	$\mathbf{\Phi}$, $\beta \in \{1, 2\}$, symm. P_0, P_{out}	Analytical expression	Conjecture	Conjecture

THANK YOU !

Many numerical simulations were performed using the open-source TrAMP package [Baker&al, '20]

Product of two complex Gaussian matrices $\Phi = \mathbf{W}_1\mathbf{W}_2$, with $\mathbf{W}_1 \in \mathbb{C}^{m\times p}$, $\mathbf{W}_2 \in \mathbb{C}^{p\times n}$ and $\gamma = p/n$



- Very good agreement of G-VAMP with the analytical predictions.
- We recover the two thresholds $\alpha_{\mathrm{WR,Algo}} = \gamma/(1+\gamma)$ and $\alpha_{\mathrm{FR,IT}} = \min(2, 2\gamma)$
- (Very small) computational-to-statistical gap $\alpha_{\mathrm{FR,Algo}} > \alpha_{\mathrm{FR,IT}}$ for $\gamma \neq 1$

Optimal algorithms (real case)