Phase retrieval in high dimensions: Statistical and computational phase transitions

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

arXiv:2006.05228
Phase retrieval as a GLM

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

Generalized Linear Model (GLM)

$Y_\mu \sim P_{\text{out}} \left( \cdot \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi_{\mu i} X_i^* \right) \right)$ $\mu \in \{1, \ldots, m\}$

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

Classical problem, non-trivial even in the noiseless case $Y_\mu = |(\Phi X^*)_\mu|^2 / n$, 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]
\[ Y_\mu \sim P_{\text{out}} \left( \cdot \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Phi_{\mu i} X_i^* \right) \right) \quad \mu \in \{1, \cdots, m\} \]

In the limit \( m, n \to \infty \) with \( \alpha = m/n = \Theta(1) \), what is the smallest \( \alpha \) needed to recover \( X^* \)...

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

**Our model:**

i) \( X^* \) and \( \Phi \) can be real (\( \beta = 1 \)) or complex (\( \beta = 2 \))

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

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

iv) The empirical spectral distribution of \( \frac{1}{n} \Phi^\dagger \Phi / n \) converges:

\[
\nu_n \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i \left( \frac{\Phi^\dagger \Phi}{n} \right)} \xrightarrow{n \to \infty} \nu \in \mathcal{M}_1^+(\mathbb{R}^+) .
\]

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

We consider any channel (not necessarily phase retrieval)

**Conjecture**: Consider the following scalar optimization problem

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

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) \( \Phi_{\mu i} \sim \mathcal{N}_{\beta}(0, 1) \) (standard Gaussian distribution)

b) \( P_0 \) is Gaussian and \( \Phi = WB \)

\( \) Gaussian matrix \( \) Any matrix

\}

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]
Optimal error in GLMs

We consider any channel (not necessarily phase retrieval)

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

“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 \( f(q_x, q_z) \)!
3 (Algorithmic) Weak recovery

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

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

This threshold \( \alpha_{\text{WR,Algo}} \) is a solution of:

\[
\alpha = \left( \frac{\langle \lambda \rangle^2}{\langle \lambda^2 \rangle} \right) \left[ 1 + \left\{ \int_{\mathbb{R}} dy \frac{D_{\beta z} (|z|^2 - 1) P_{\text{out}} \left[ y \left| \sqrt{\frac{\langle \lambda \rangle^2}{\alpha n}} z \right] \right]}{\int_{\mathbb{R}} D_{\beta z} P_{\text{out}} \left[ y \left| \sqrt{\frac{\langle \lambda \rangle^2}{\alpha n}} z \right] \right]} \right\}^{-1} \right]
\]

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:

\[
\alpha = \left( 1 + \frac{\beta}{2} \right) \frac{\langle \lambda \rangle^2}{\langle \lambda^2 \rangle}
\]

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

• This is an implicit equation
• Derived by a stability analysis of the replica-symmetric potential.
4 Strong (full) recovery

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

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

If
$$\frac{1}{n} \text{rk} \left( \frac{\Phi^\dagger \Phi}{n} \right) \to r \in [0, 1] : \alpha_{\text{FR, IT}} = \beta r$$

Does not depend on the precise statistics of $\Phi$

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

- The real case $\alpha_{\text{FR, IT}} = r$ can be derived by a counting argument [Candès & al., '05]
- The complex case $\alpha_{\text{FR, IT}} = 2r$ can (as far as we know) only be derived by the replica-symmetric potential!
We consider noiseless phase retrieval: \( P_{\text{out}}(y,z) = \delta(y - |z|^2) \) and a Gaussian prior \( P_0 = N_2(0, 1) \)

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

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

<table>
<thead>
<tr>
<th>Matrix ensemble and value of $\beta$</th>
<th>$\alpha_{WR, Algo}$</th>
<th>$\alpha_{FR, IT}$</th>
<th>$\alpha_{FR, Algo}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Gaussian $\Phi$ ($\beta = 1$)</td>
<td>0.5</td>
<td>1</td>
<td>$\simeq 1.12$</td>
</tr>
<tr>
<td>Complex Gaussian $\Phi$ ($\beta = 2$)</td>
<td>1</td>
<td>2</td>
<td>$\simeq 2.027$</td>
</tr>
<tr>
<td>Real column-orthogonal $\Phi$ ($\beta = 1$)</td>
<td>1.5</td>
<td>1</td>
<td>$\simeq 1.584$</td>
</tr>
<tr>
<td>Complex column-unitary $\Phi$ ($\beta = 2$)</td>
<td>2</td>
<td>2</td>
<td>$\simeq 2.265$</td>
</tr>
<tr>
<td>$\Phi = W_1 W_2$ ($\beta = 1$, aspect ratio $\gamma$)</td>
<td>$\gamma/(2(1 + \gamma))$</td>
<td>$\gamma/(1 + \gamma)$</td>
<td>Theorem</td>
</tr>
<tr>
<td>$\Phi = W_1 W_2$ ($\beta = 2$, aspect ratio $\gamma$)</td>
<td>Analytical expression</td>
<td>min$(1, \gamma)$</td>
<td>Theorem</td>
</tr>
<tr>
<td>$\Phi, \beta \in {1, 2}$, $r k[\Phi^\dagger \Phi]/n = r$</td>
<td>Analytical expression</td>
<td>min$(2, 2\gamma)$</td>
<td>Conjecture</td>
</tr>
<tr>
<td>Gauss. $\Phi$, $\beta \in {1, 2}$, symm. $P_0, P_{out}$</td>
<td>$\beta r$</td>
<td>$\beta r$</td>
<td>Theorem</td>
</tr>
<tr>
<td>$\Phi = WB$, $\beta \in {1, 2}$, Gauss. $P_0$, symm. $P_{out}$</td>
<td>Analytical expression</td>
<td>Analytical expression</td>
<td>Theorem</td>
</tr>
<tr>
<td>$\Phi, \beta \in {1, 2}$, symm. $P_0, P_{out}$</td>
<td>Analytical expression</td>
<td>Analytical expression</td>
<td>Conjecture</td>
</tr>
</tbody>
</table>

**Thank you!**

Many numerical simulations were performed using the open-source TrAMP package [Baker et al., ‘20].
Product of two complex Gaussian matrices $\Phi = W_1W_2$, with $W_1 \in \mathbb{C}^{m \times p}$, $W_2 \in \mathbb{C}^{p \times n}$ and $\gamma = p/n$.

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