Phase retrieval in high dimensions: Statistical and computational phase transitions

arXiv:2006.05228, *NeurIPS '20.*

Antoine Maillard & Bruno Loureiro

Joint work with Florent Krzakala & Lenka Zdeborová

INRIA – December 16th 2020

[A.M., Loureiro, Krzakala, Zdeborová *NeurIPS '20*]

Generalized Linear Models

Generalized Linear Model (GLM)

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

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

Real / Complex
$\beta = 1$ $\beta = 2$

(Probabilistic) channel

Sensing matrix (real/complex) Signal (real/complex), d-dimensional

In phase retrieval, one only measures the modulus, e.g. noiseless $Y_\mu = \frac{1}{d}|(\mathbf{\Phi X}^\star)_\mu|^2$; Poisson-noise $Y_\mu \sim \mathrm{Pois}(\Lambda|(\mathbf{\Phi X}^\star)_\mu|^2/d)$

Classical problem, non-trivial even in the noiseless case, many algorithms :

- SDP relaxations [Candès&al '15a, Candès&al '15b, Waldspurger&al '15, Goldstein&al '18, ...]
- Non-convex optimization procedures [Netrapalli&al '15, Candès&al '15c, ...]
- Spectral methods [Mondelli&al '18, Luo&al '18, Dudeja&al '19, A.M., Lu, Krzakala, Zdeborová '20 ...]

Goal : Fundamental limits of GLMs with random sensing matrices and random signal in the typical case and in high dimension.

⚠ Different from the injectivity studies of the "worst-case" (e.g. [Bandeira&al '14] for phase retrieval)

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 ?
- Perfectly ? (up to the possible rank deficiency of $\mathbf{\Phi}$)
- With which (polynomial-time) algorithm ?

[A.M., Loureiro, Krzakala, Zdeborová *NeurIPS '20*]

Our model: i) The signal $\mathbf{X}^\star$ is generated using a (known) i.i.d. zero-mean prior distribution P_0 and $\mathbb{E}_{P_0}[|X|^2] = \rho$

ii) The matrix $\mathbf{\Phi}$ is right-orthogonally (unitarily) invariant : $\forall \mathbf{U},\ \mathbf{\Phi} \stackrel{d}{=} \mathbf{\Phi}\mathbf{U}$ and the empirical spectral distribution of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/d$ weakly converges (a.s.) to a compactly-supported measure: $\nu_d \equiv \frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i\left(\frac{\mathbf{\Phi}^\dagger\mathbf{\Phi}}{d}\right)} \xrightarrow[n\to\infty]{\text{weakly, a.s.}} \nu \in \mathcal{M}_1^+(\mathbb{R}_+)$

e.g. Gaussian matrices, product of Gaussians, random column-orthogonal/unitary, any $\mathbf{\Phi} \equiv \mathbf{U}\mathbf{S}\mathbf{V}^\dagger$ with $S_i^2 \stackrel{\text{i.i.d.}}{\sim} \nu$.

The information-theoretic Bayes-optimal estimator can be found as the first moment of the posterior distribution:

$$P(\mathrm{d}\mathbf{x}|\mathbf{Y},\mathbf{\Phi}) \equiv \frac{1}{\mathcal{Z}_d(\mathbf{Y},\mathbf{\Phi})} \prod_{i=1}^{d} P_0(\mathrm{d}x_i) \prod_{\mu=1}^{n} P_{\text{out}}\left(Y_\mu \Big| \frac{1}{\sqrt{d}}\sum_{i=1}^{d}\Phi_{\mu i}x_i\right)$$

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

Conjecture ("Replica formula") : Under the above hypotheses,

$$\lim_{d\to\infty} \frac{1}{d}\mathbb{E}\ln \mathcal{Z}_d(\mathbf{Y},\mathbf{\Phi}) = \sup_{q_x,q_z} \overbrace{[\underbrace{I_0^{(\beta)}(q_x)}_{P_0} + \underbrace{I_{\text{out}}^{(\beta)}(q_z)}_{P_{\text{out}}} + \underbrace{\beta I_{\text{int}}(q_x,q_z)}_{\nu}]}^{f(q_x,q_z)}$$

Then the *information-theoretic Minimal Mean Squared Error* is : $\lim_{d\to\infty} \mathbb{E}\|\mathbf{X}^\star - \hat{\mathbf{X}}_{\text{opt}}\|^2/d = \rho - q_x$.

- The functions involved in the optimization problem are fully explicit.
- The log-partition (or *free entropy*) is related to the *mutual information* $\mathrm{I}(\mathbf{X}^\star;\mathbf{Y}|\mathbf{\Phi}) = \mathbb{E}\ln\mathcal{Z}_d - n\mathbb{E}\ln P_{\text{out}}\left(Y_1 \Big| \frac{(\mathbf{\Phi}\mathbf{X}^\star)_1}{\sqrt{d}}\right)$
- Conjecture obtained with the heuristic replica method of statistical physics. [Mézard&al 1987, Takahashi&al '20] $\mathbb{E}\ln\mathcal{Z} = \lim_{r\to 0^+} \frac{\mathbb{E}\mathcal{Z}^r - 1}{r}$

Proving the replica formula

[A.M., Loureiro, Krzakala, Zdeborová *NeurIPS* '20]

Theorem (informal) : 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 replica conjecture stands.

The replica formula for GLMs was so far only proven for real Gaussian matrices [Barbier&al '19], we tackle for the first time heavily correlated data !

Theorem (formal) : Assume that the channel is "well-behaved" (i.e. regular enough, with bounded derivatives) and let:

- (H1) P_0 is a centered Gaussian distribution.
- (H2) $\mathbf{\Phi}$ is distributed as $\mathbf{\Phi} \overset{\text{d}}{=} \mathbf{W}\mathbf{B}/\sqrt{p}$, with $\mathbf{W} \in \mathbb{K}^{n\times p}$ i.i.d. Gaussian, and $\mathbf{B} \in \mathbb{K}^{p\times d}$ an arbitrary matrix (random or deterministic) independent of $\mathbf{W}$. Moreover, as $d \to \infty$, $p/d \to \delta > 0$.
- (H3) The empirical spectral distribution of $\mathbf{B}^\dagger\mathbf{B}/d$ weakly converges (a.s.) to a compactly-supported $\nu_B \neq \delta_0$. Moreover, there exists $\lambda_{\max}$ such that a.s. $\lambda_{\max}(\mathbf{B}^\dagger\mathbf{B}/d) \underset{d\to\infty}{\longrightarrow} \lambda_{\max}$.
- (H') P_0 has a finite second moment, and $\Phi_{\mu i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_\beta(0,1)$.

Assume that (H1)-(H2)-(H3) hold, or that (H') holds. Then the replica conjecture stands.

- We use probabilistic interpolation methods [Guerra '03, Talagrand '06], specifically an adaptive interpolation [Barbier&al '19].
- The regularity and boundedness hypotheses on the channel can be relaxed following the lines of [Barbier&al '19], including e.g. noiseless phase retrieval.

Algorithmic performance in GLMs

[A.M., Loureiro, Krzakala, Zdeborová *NeurIPS '20*]

$$\lim_{d\to\infty} \frac{1}{d}\mathbb{E}\ln \mathcal{Z}_d(\mathbf{Y},\mathbf{\Phi}) = \sup_{q_x,q_z} \underbrace{\left[I_0^{(\beta)}(q_x) + I_{\text{out}}^{(\beta)}(q_z) + \beta I_{\text{int}}(q_x,q_z)\right]}_{\text{"Replica-symmetric" potential } f(q_x,q_z)}$$

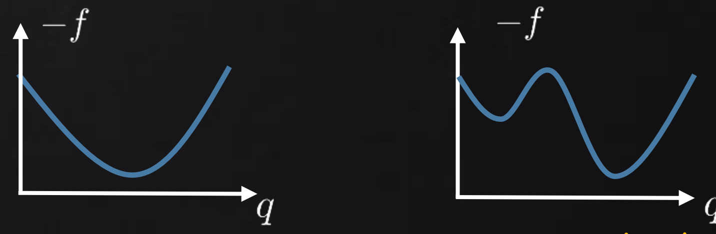
<u>Strong conjecture :</u> For GLMs, the optimal polynomial-time algorithm is an explicit iterative algorithm :

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).

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



No gap

Gap : "Hard" phase

Application: thresholds in phase retrieval

[A.M., Loureiro, Krzakala, Zdeborová *NeurIPS '20*]

$$P_{\rm out}(y|z) = P_{\rm out}(y||z|)$$

Weak-recovery

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

This threshold $\alpha_{\rm WR, Algo}$ is the only solution to :

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

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

Derived by a stability analysis of the replica-symmetric potential around the uninformative point.

Strong recovery

How many measurements $\alpha = n/d$ are necessary to be able to information-theoretically achieve the best possible recovery ?

Noiseless phase retrieval $P_{\rm out}(y|z) = \delta(y - |z|^2)$ and Gaussian prior

If (a.s.) $\frac{1}{d}\mathrm{rk}\left(\frac{\Phi^\dagger \Phi}{d}\right) \to r \in [0,1]$ then $\alpha_{\rm FR,IT} = \beta r$

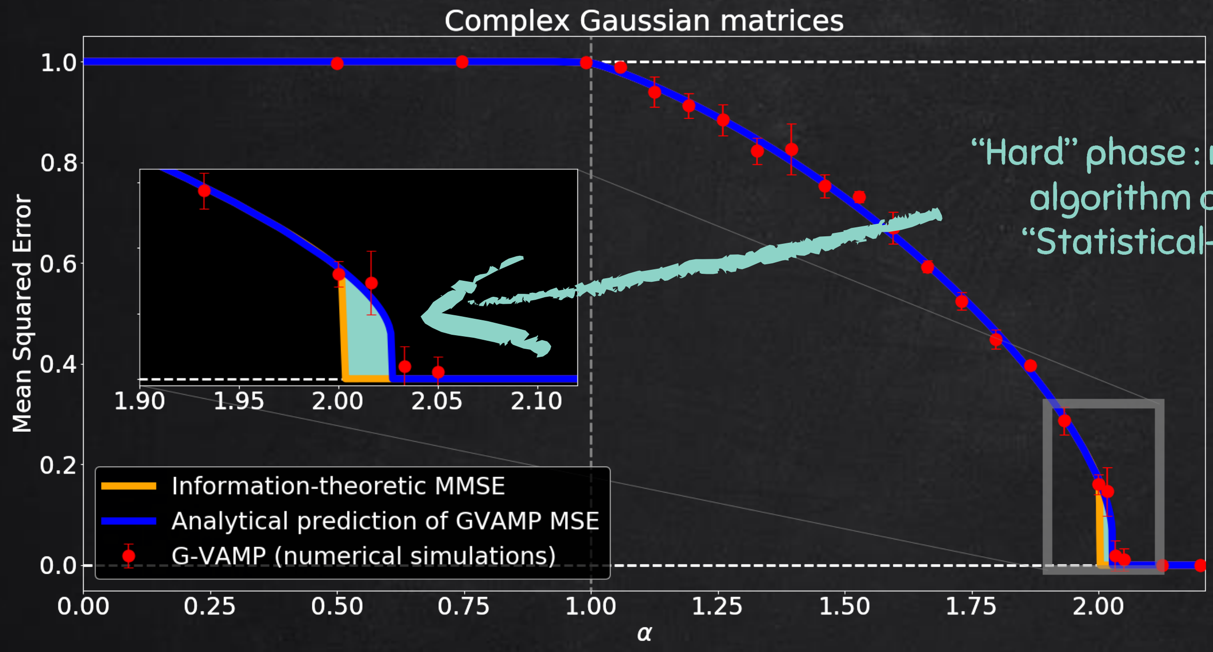
Derived by analyzing under which condition is the "perfect" recovery point a global maximum of the RS potential.

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

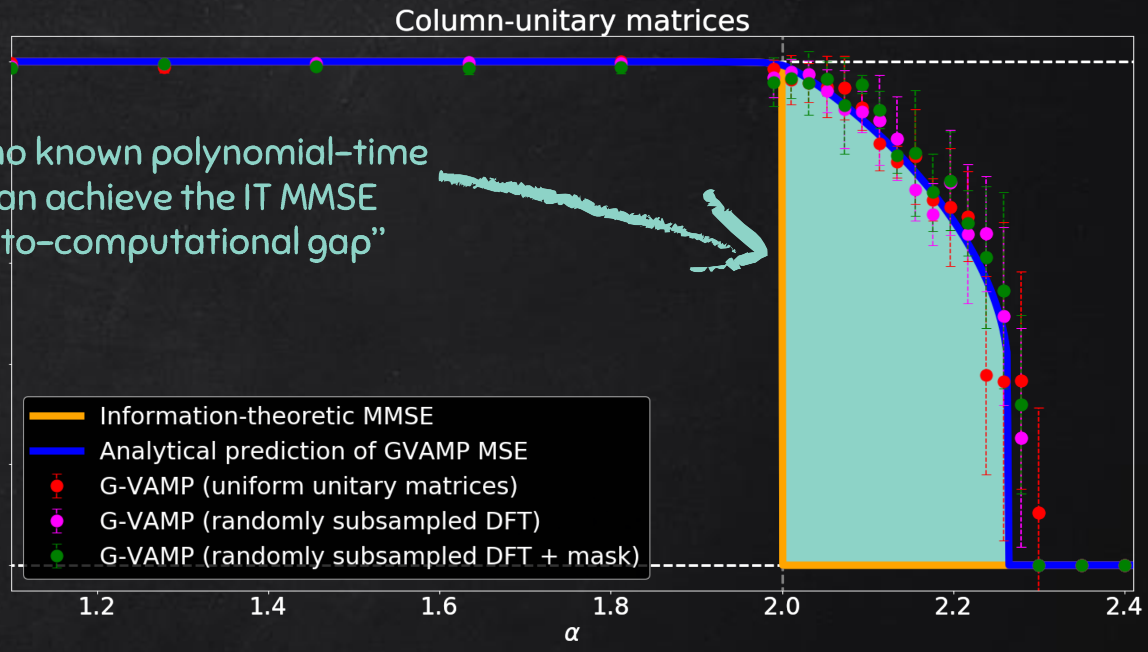
Numerical applications (1)

[A.M., Loureiro, Krzakala, Zdeborová *NeurIPS '20*]

We consider noiseless complex phase retrieval: $P_{\mathrm{out}}(y|z) = \delta(y - |z|^2)$ and a Gaussian prior



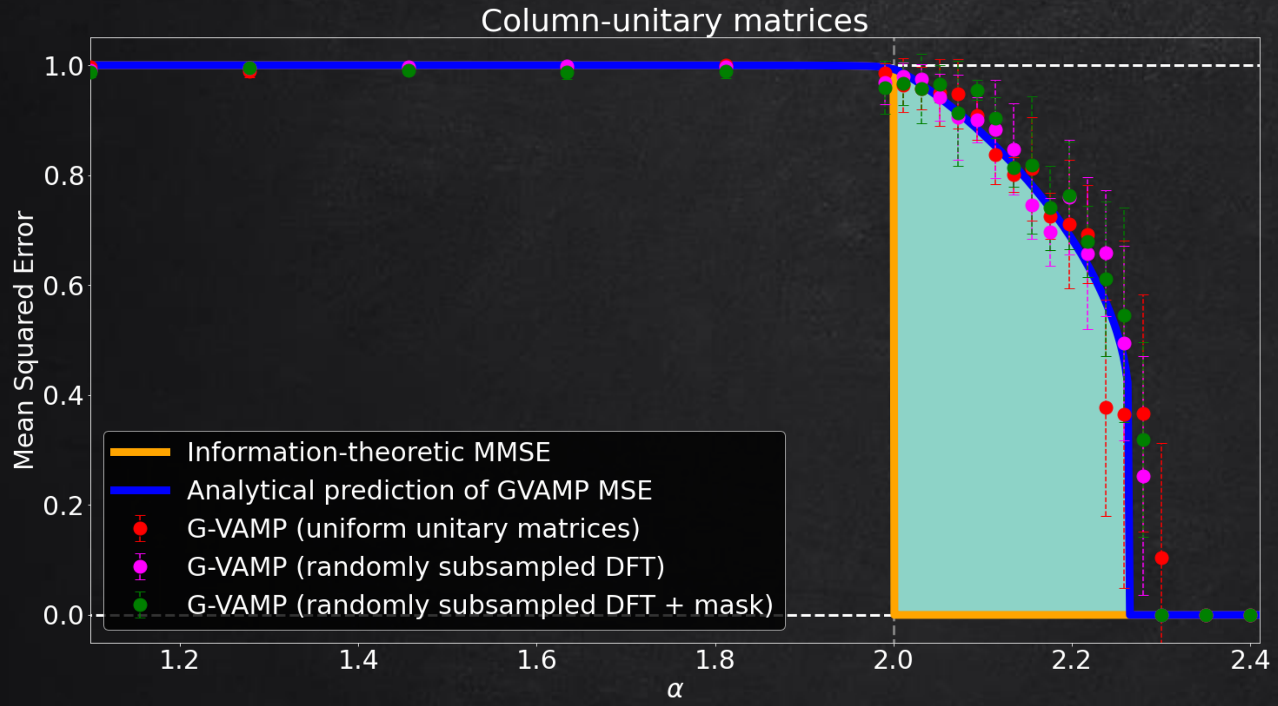
"Hard" phase: no known polynomial-time algorithm can achieve the IT MMSE
"Statistical-to-computational gap"

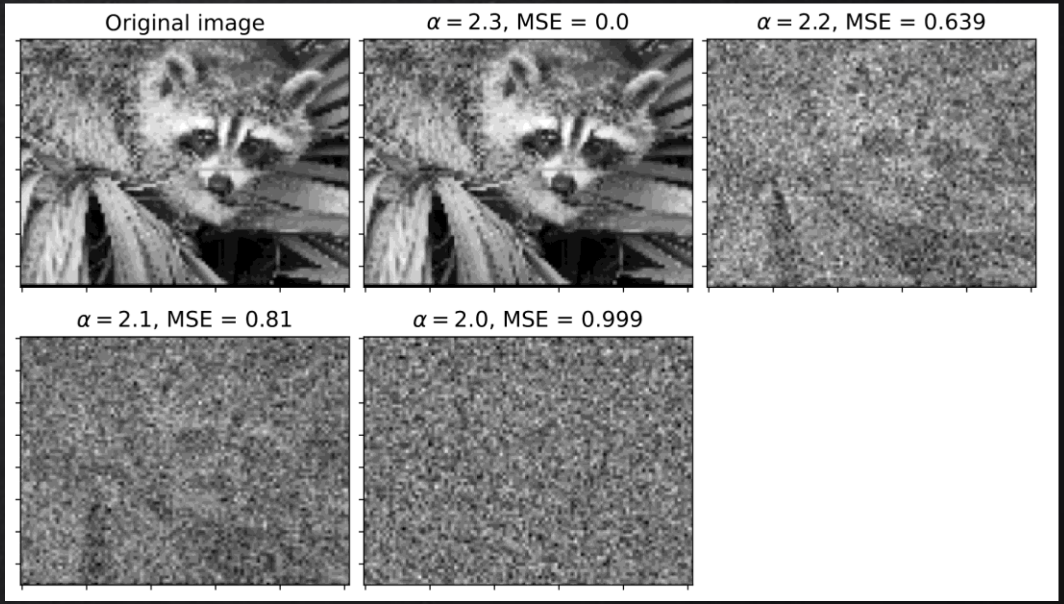


Very good agreement of G-VAMP with the analytical predictions.

Numerical applications (2)

[A.M., Loureiro, Krzakala, Zdeborová *NeurIPS '20*]





- We consider uniformly sampled column-unitary sensing matrices.
- Very good agreement of G-VAMP with the analytical predictions, even a with natural image (i.e. a very structured signal) !
- Matrices with controlled structure (e.g. randomly subsampled DFT) still perform very well !
- For column-unitary matrices we have $\alpha_{\mathrm{FR,IT}} = \alpha_{\mathrm{WR,Algo}} = 2$: "all-or-nothing" IT transition.
- For all other full-rank complex matrices $\alpha_{\mathrm{WR,Algo}} < \alpha_{\mathrm{FR,IT}}$.

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 $\gamma)$	$\gamma/(2(1+\gamma))$	$\min(1,\gamma)$	Theorem
	$\mathbf{\Phi} = \mathbf{W}_1\mathbf{W}_2$ $(\beta = 2$, aspect ratio $\gamma)$	$\gamma/(1+\gamma)$	$\min(2, 2\gamma)$	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

The theory is still far from complete

- All current proofs of replica formulas need some Gaussianity/iid behavior. Go beyond ?
- What if we do not know how the data was generated ?

THANK YOU !

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

Appendix: The replica-symmetric potential

$$\boxed{\lim_{d\to\infty}\frac{1}{d}\mathbb{E}_{\mathbf{Y},\mathbf{\Phi}}\ln \mathcal{Z}_d(\mathbf{Y},\mathbf{\Phi}) = \sup_{q_x\in[0,\rho]}\sup_{q_z\in[0,Q_z]}\left[I_0(q_x)+\alpha I_{\rm out}(q_z)+I_{\rm int}(q_x,q_z)\right]}$$

$$I_0(q_x) \equiv \inf_{\hat{q}_x\geq 0}\left[-\frac{\beta\hat{q}_x q_x}{2}+\mathbb{E}_\xi \mathcal{Z}_0(\sqrt{\hat{q}_x}\xi,\hat{q}_x)\log\mathcal{Z}_0(\sqrt{\hat{q}_x}\xi,\hat{q}_x)\right]$$

$$I_{\rm out}(q_z) \equiv \inf_{\hat{q}_z\geq 0}\left[-\frac{\beta\hat{q}_z q_z}{2}-\frac{\beta}{2}\ln(\hat{Q}_z+\hat{q}_z)+\frac{\beta\hat{q}_z}{2\hat{Q}_z}\quad+\mathbb{E}_\xi\int_{\mathbb{R}}\mathrm{d}y\ \mathcal{Z}_{\rm out}\left(y;\sqrt{\frac{\hat{q}_z}{\hat{Q}_z(\hat{Q}_z+\hat{q}_z)}}\xi,\frac{1}{\hat{Q}_z+\hat{q}_z}\right)\log\mathcal{Z}_{\rm out}\left(y;\sqrt{\frac{\hat{q}_z}{\hat{Q}_z(\hat{Q}_z+\hat{q}_z)}}\xi,\frac{1}{\hat{Q}_z+\hat{q}_z}\right)\right]$$

$$I_{\rm int}(q_x,q_z) \equiv \inf_{\gamma_x,\gamma_z\geq 0}\left[\frac{\beta}{2}(\rho-q_x)\gamma_x+\frac{\alpha\beta}{2}(Q_z-q_z)\gamma_z-\frac{\beta}{2}\langle\ln(\rho^{-1}+\gamma_x+\lambda\gamma_z)\rangle_\nu\right]-\frac{\beta}{2}\ln(\rho-q_x)-\frac{\beta q_x}{2\rho}-\frac{\alpha\beta}{2}\ln(Q_z-q_z)-\frac{\alpha\beta q_z}{2Q_z}$$

With
- $\xi\sim\mathcal{N}_\beta(0,1)$
- $Q_z\equiv\rho\langle\lambda\rangle_\nu/\alpha$
- $\hat{Q}_z\equiv 1/Q_z$
- and the auxiliary functions $\mathcal{Z}_0(b,a)\equiv \mathbb{E}_{z\sim\mathcal{N}_\beta(0,1)}\left[P_0(z)e^{-\frac{\beta}{2}a|z|^2+\beta b\cdot z}\right]$, $\mathcal{Z}_{\rm out}(y;\omega,v)\equiv\mathbb{E}_{z\sim\mathcal{N}_\beta(0,1)}\left[P_{\rm out}\left(y\middle|\sqrt{v}z+\omega\right)\right]$