Fundamental limits of high-dimensional estimation

A stroll between statistical physics, probability, and random matrix theory

Antoine Maillard

Under the supervision of Florent Krzakala & Lenka Zdeborová

Département de Physique
École normale supérieure
ENS | PSL
ETH zürich
FONDATION CFM POUR LA RECHERCHE

Rencontre des Jeunes Physicien.ne.s – November 2nd 2022

What is statistical inference ?

Input data

Φ_μ → **Model** $\mathbf{X}^\star$ → y_μ

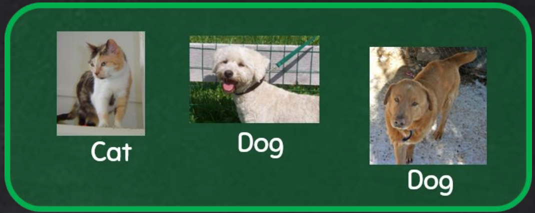
Output data

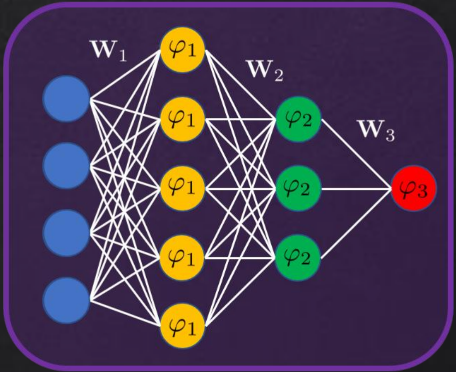
"Learning from data"

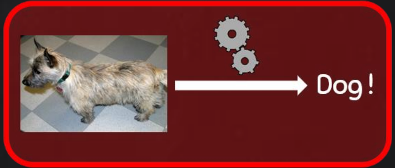
$\Phi = \{\Phi_1, \cdots, \Phi_m\} \quad \mathbf{Y} = \{y_1, \cdots, y_m\}$

?

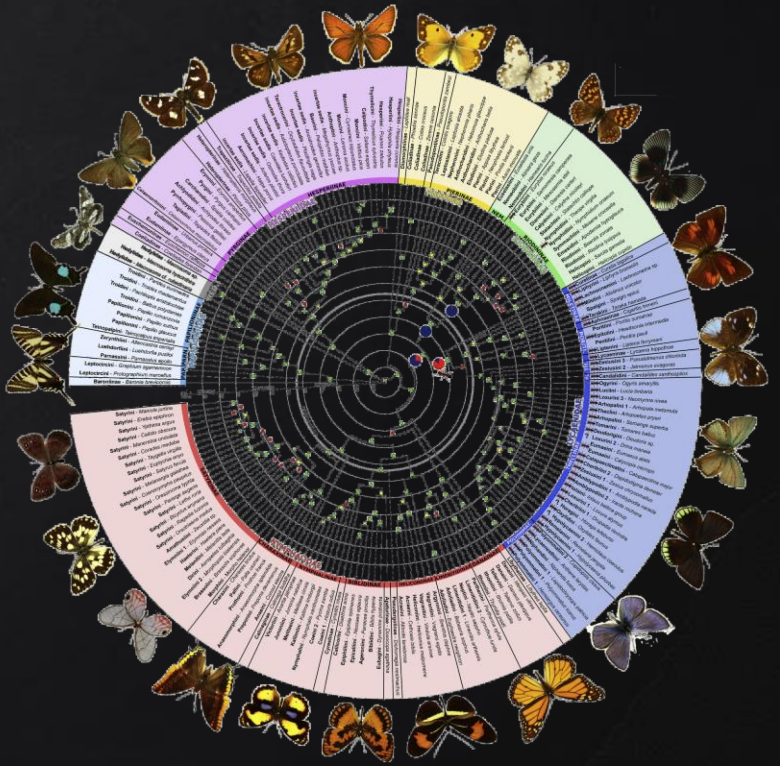
"Signal" $\mathbf{X}^\star$







- Supervised learning in neural networks
- Signal processing
- Phase retrieval
- Quantitative finance, particle physics, evolutionary biology,...



[Espeland&al '18]

High-dimensional statistics

$$\left\{ \mathbf{\Phi}_\mu \in \mathbb{R}^p \rightarrow \boxed{\text{Model } \mathbf{X}^\star \in \mathbb{R}^n} \rightarrow y_\mu \right\}_{\mu=1,\cdots,m}$$

Data deluge

Gigantic databases and explosion of computing power.

↔

Theoretical revolution of the 2000s

High-dimensional statistics

"Modern machine learning": GoogLeNet [Szegedy&al '15]: $n \simeq 5 \times 10^6$ and $m \simeq 10^6$.

Fundamental limits

Can we recover $\mathbf{X}^\star$...

- Perfectly ? Partially ?
- Efficiently ? With which algorithms ?

+

"High-dimensional" limit

Number of parameters $n \to \infty$

+

Number of data $m \to \infty$

Bayesian estimation – Statistical physics

?

$\left\{ \mathbf{\Phi}_\mu \in \mathbb{R}^p \rightarrow \boxed{\text{Model } \mathbf{X}^\star \in \mathbb{R}^n} \rightarrow y_\mu \right\}_{\mu = 1, \cdots, m}$

Bayes' law

Likelihood

$$\boxed{\mathbb{P}(\mathbf{x}|\mathbf{Y},\mathbf{\Phi}) = \frac{\mathbb{P}(\mathbf{x})\mathbb{P}(\mathbf{Y}|\mathbf{x},\mathbf{\Phi})}{\mathbb{P}(\mathbf{Y}|\mathbf{\Phi})}}$$

Prior knowledge on $\mathbf{X}^\star$

"Statistical mechanics 101"

Gibbs-Boltzmann distribution

$$\boxed{\mathbb{P}(x_1, \cdots, x_n) = \frac{e^{-\mathcal{H}(\mathbf{x})/T}}{\mathcal{Z}(T)}}$$

"Disordered" model, with Hamiltonian $\mathcal{H}(\mathbf{x}) = -\log \mathbb{P}(\mathbf{x}) - \log \mathbb{P}(\mathbf{Y}|\mathbf{x}, \mathbf{\Phi}) \quad (T = 1)$

Random variables (noise, ...)

Where do we go from here?

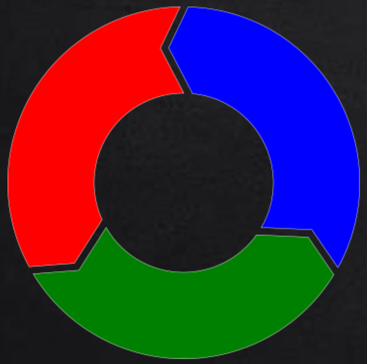
[Hopfield '82; Mézard&Parisi '85; Gardner&Derrida '89; Anderson '89; Mézard&Montanari '09; ...]

Deep and detailed connection

➢ Bayesian estimation problems	⟺	➢ Statistical physics
➢ Posterior distribution	⟺	➢ Gibbs–Boltzmann distribution
➢ High-dimensional limit	⟺	➢ Thermodynamic limit
➢ Randomness of the observations (noise, ...)	⟺	➢ Disordered systems, "spin glasses"

<u>When is learning/inference possible ?</u>

Structure of the data



Architecture of the model

Algorithms

[Zdeborová '20]

Statistical physics allow to study each of these pieces!

Phase retrieval [A.M., Loureiro, Krzakala, Zdeborová 20]

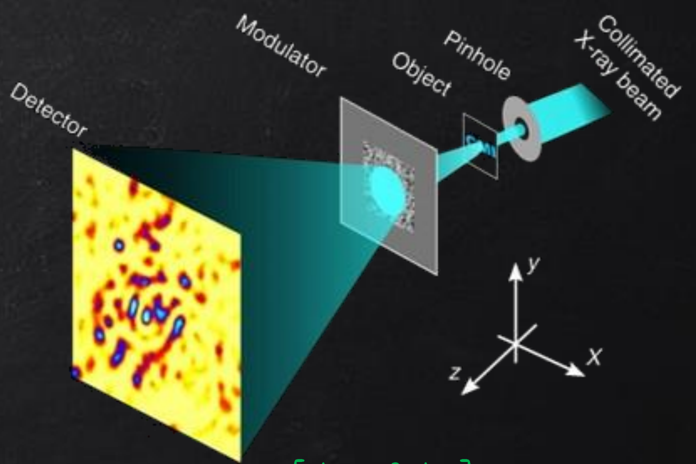
Goal : Recover $\mathbf{X}^\star \in \mathbb{C}^n$ from phaseless measurements

$$Y_\mu = \frac{1}{n}|\mathbf{\Phi}_\mu \cdot \mathbf{X}^\star|^2$$

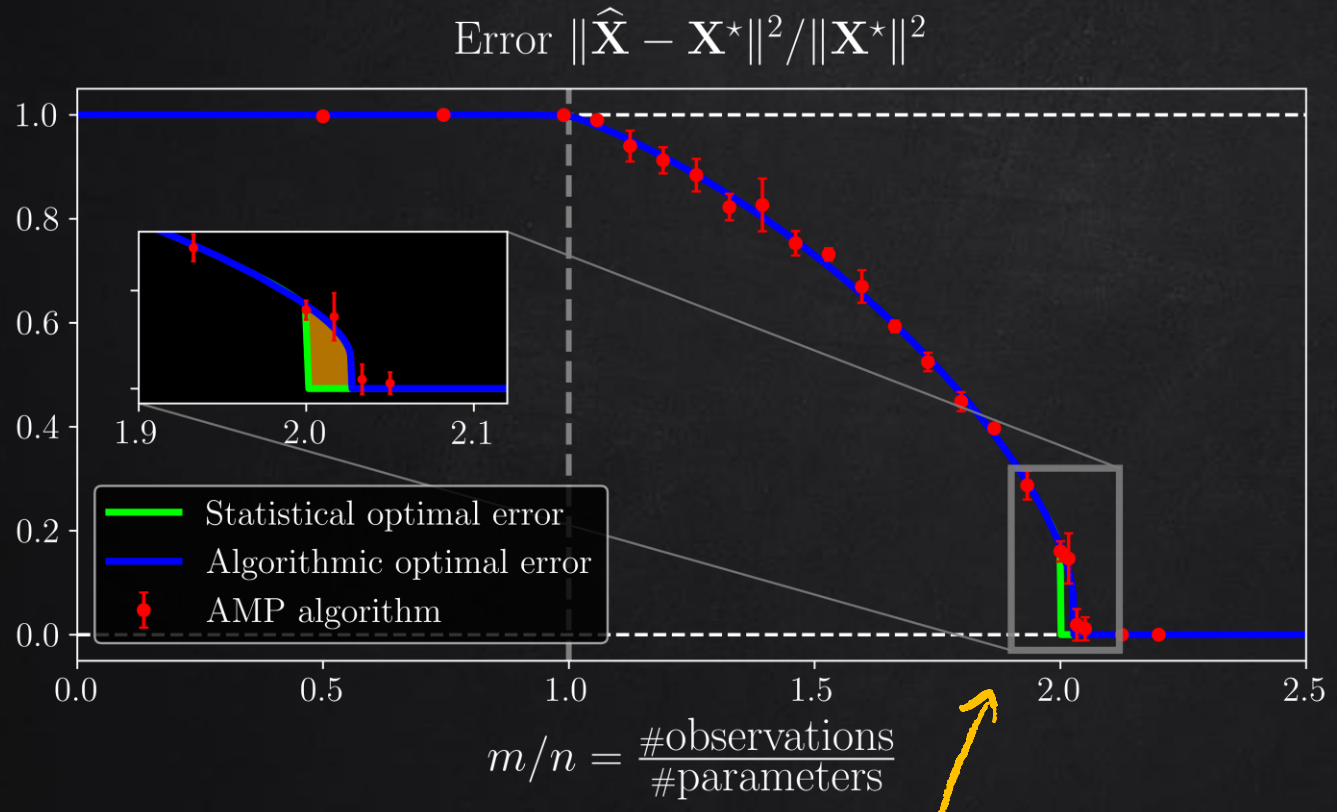
Imaging in complex media

$$\mathbf{\Phi}_\mu \sim \mathcal{CN}(0, \mathrm{I}_n)$$

Model



[Zhang&al 16]



First-order phase transition: no known polynomial-time algorithm succeeds ! "Hard phase"

Use statistical physics !

❖ Analytical predictions using Parisi's replica theory

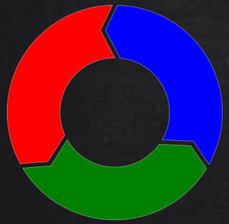
2021

❖ Efficient (message-passing) algorithms

❖ Rigorous proofs using involved probabilistic tools

Concluding remarks

Theory of inference/learning



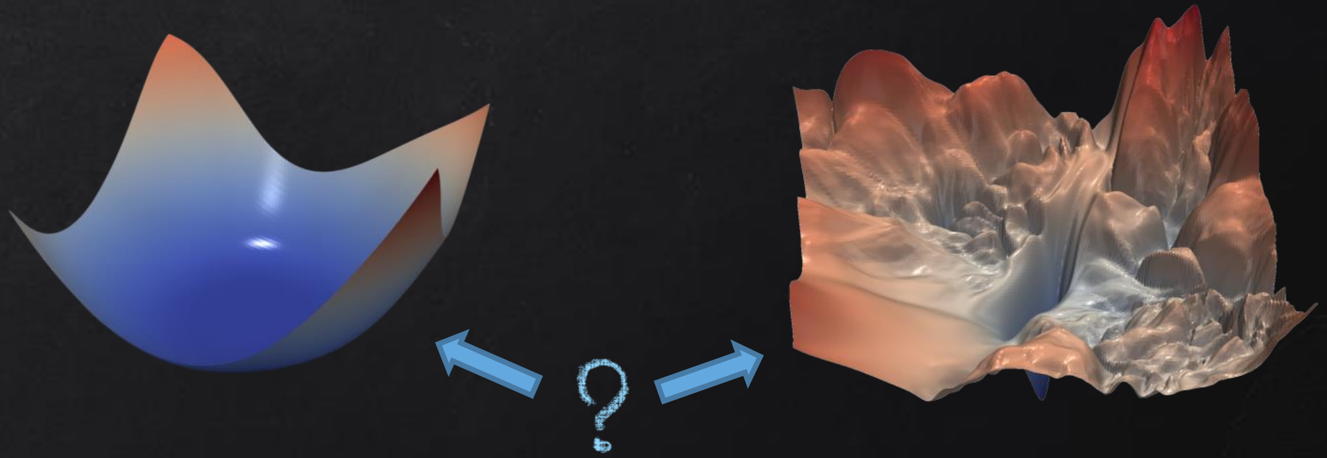
A toolbox from statistical physics and probability theory

Replica, message-passing, Plefka expansions, DMFT...

Large deviations, concentration inequalities, random matrix theory...,

Some applications:

- Statistical-to-computational gaps – "Hard phases"
- Understand optimization landscapes
- ....



Many open problems to crack !

THANK YOU !