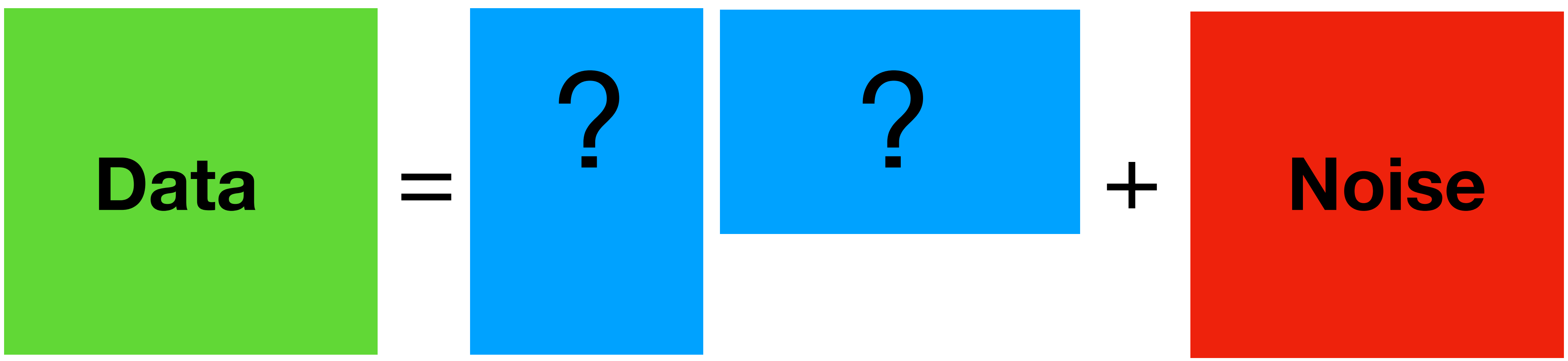
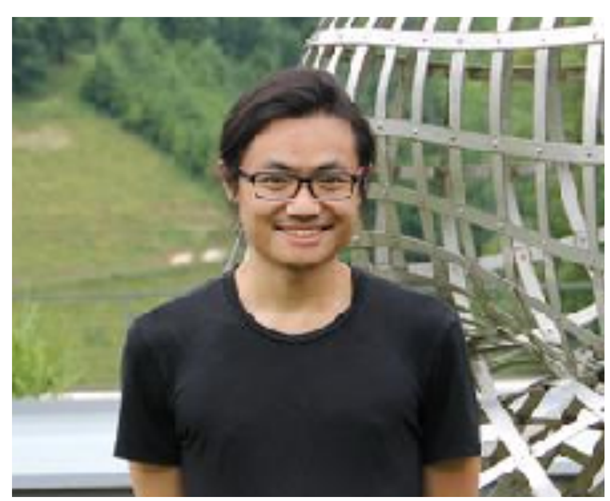
Phase diagram of extensive-rank symmetric matrix denoising beyond rotational invariance

Jean Barbier

International Center for Theoretical Physics
Trieste, Italy

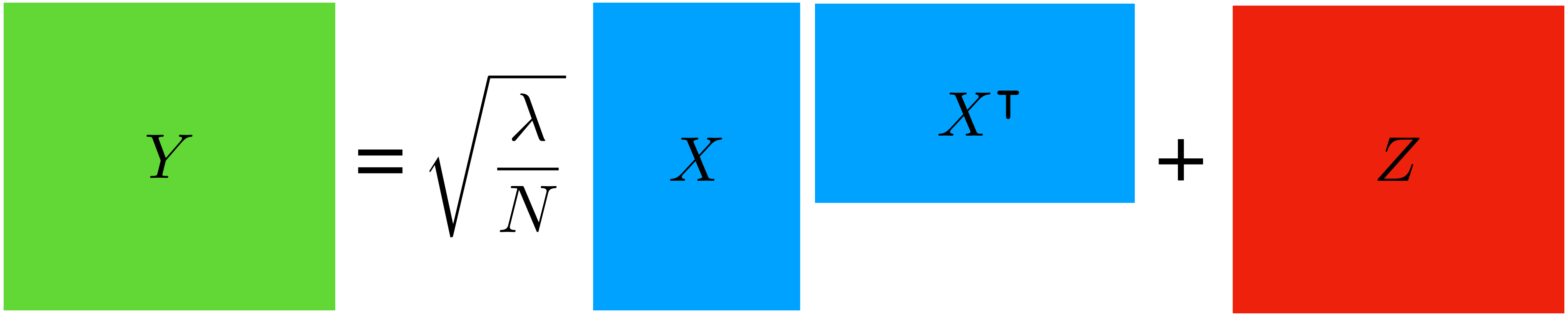


F. Camilli (ICTP)



J. Ko (Waterloo U)

K. Okajima (Tokyo U)



$$Y = \sqrt{\frac{\lambda}{N}} X X^{\intercal} + Z \qquad Y, Z \in \mathbb{R}^{N \times N} \qquad X \in \mathbb{R}^{N \times M}$$

$$Z_{ij} = Z_{ji} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1 + \delta_{ij})$$

$$X_{i\mu} \overset{i.i.d.}{\sim} P_X \quad \mathbb{E}X = 0, \ \mathbb{E}X^2 = 1$$

Known (Bayes optimal)

$$\text{Rot. inv. } P_X = \mathcal{N}(0,1) \ \text{ VS } \ \text{NOT Rot. inv. } P_X = \tfrac{1}{2}\delta_{-1} + \tfrac{1}{2}\delta_{+1}$$

Non-symmetric: dictionary learning $Y = \sqrt{\frac{\lambda}{N}} U V^{\intercal} + Z$

Representation learning, signal processing, recommender systems…

Bayesian optimal setting

Posterior

$$dP(x|Y) \propto \exp\Big(-\frac{1}{4}\|Y - \sqrt{\frac{\lambda}{N}} x x^\intercal \|^2\Big) \prod_{i,\mu}^{N,M} dP_X(x_{i\mu})$$

$$= \frac{1}{\mathcal{Z}} \exp\Big(\frac{1}{2}\sqrt{\frac{\lambda}{N}} \mathrm{Tr} Y x x^\intercal - \frac{\lambda}{4N}\|x x^\intercal\|^2\Big) \prod_{i,\mu}^{N,M} dP_X(x_{i\mu})$$

Minimum mean-square error (MMSE)

$$\lim_{N\to\infty} \frac{1}{MN^2}\mathbb{E}\|XX^\intercal - \mathbb{E}[xx^\intercal|Y]\|^2$$

Bayesian optimal setting

Posterior

$$dP(x|Y) \propto \exp\Big(-\frac{1}{4}\|Y - \sqrt{\frac{\lambda}{N}}xx^\intercal\|^2\Big)\prod_{i,\mu}^{N,M} dP_X(x_{i\mu})$$

$$= \frac{1}{\mathcal{Z}}\exp\Big(\frac{1}{2}\sqrt{\frac{\lambda}{N}}\mathrm{Tr}Yxx^\intercal - \frac{\lambda}{4N}\|xx^\intercal\|^2\Big)\prod_{i,\mu}^{N,M} dP_X(x_{i\mu})$$

Mutual information (MI)

$$\lim_{N\to\infty}\frac{1}{MN}I(XX^\intercal;Y)$$

$$= \lim \frac{1}{MN}\Big(H(Y) - H(Y|XX^\intercal)\Big)$$

$$= \frac{\lambda}{4} - \lim \frac{1}{MN}\mathbb{E}_Y \ln \mathcal{Z}$$

$$= \frac{\lambda}{4} - \lim \frac{1}{MN}\mathbb{E}_Y \ln \int e^{\frac{1}{2}\sqrt{\frac{\lambda}{N}}\mathrm{Tr}Yxx^\intercal - \frac{\lambda}{4N}\|xx^\intercal\|^2}\prod_{i,\mu}^{N,M} dP_X(x_{i\mu})$$

Free entropy $\phi_{N,M}$

$$Y = \sqrt{\frac{\lambda}{N}}\, X X^{\intercal} + Z$$

RMT

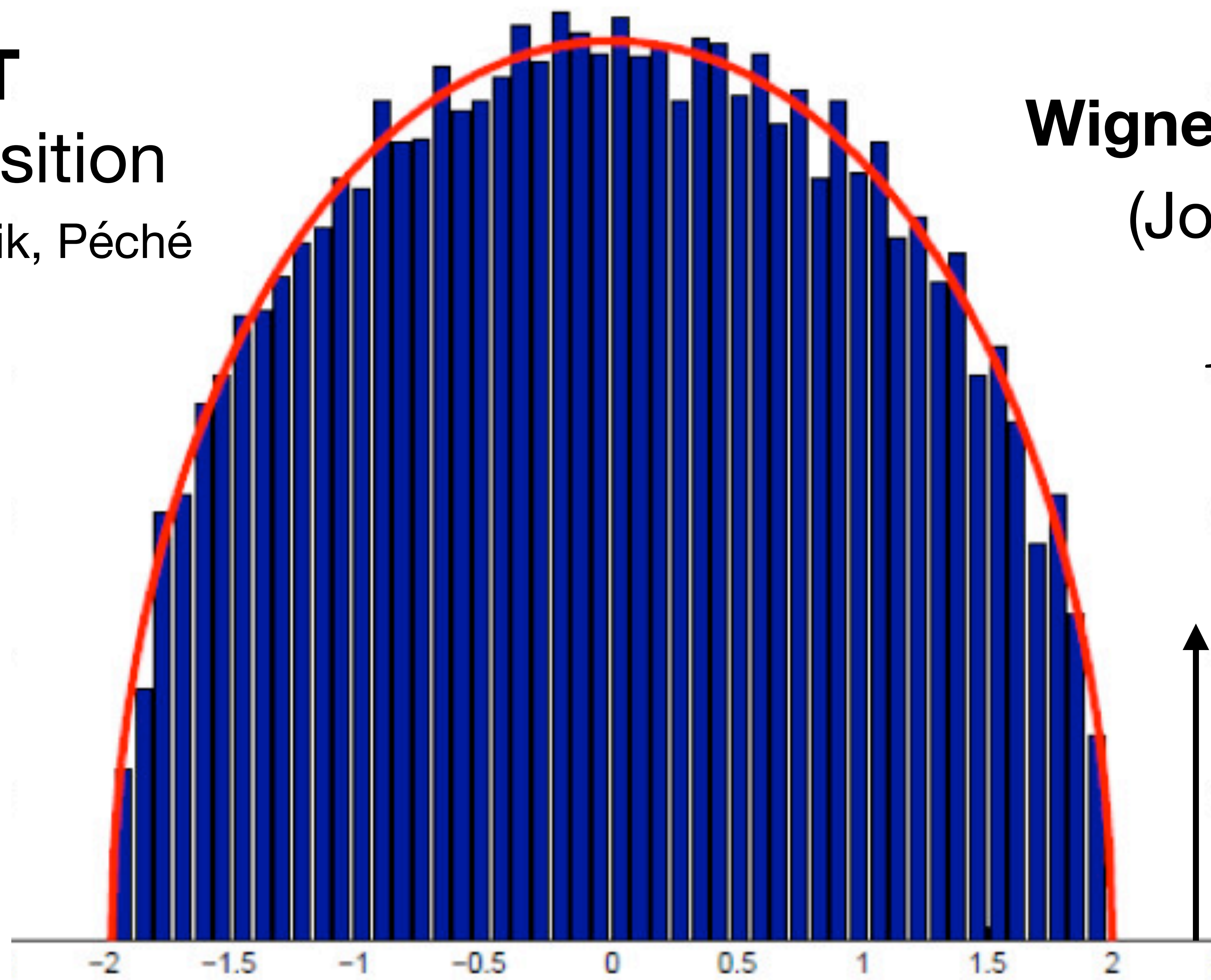
BBP transition

Ben-Arous, Baik, Péché

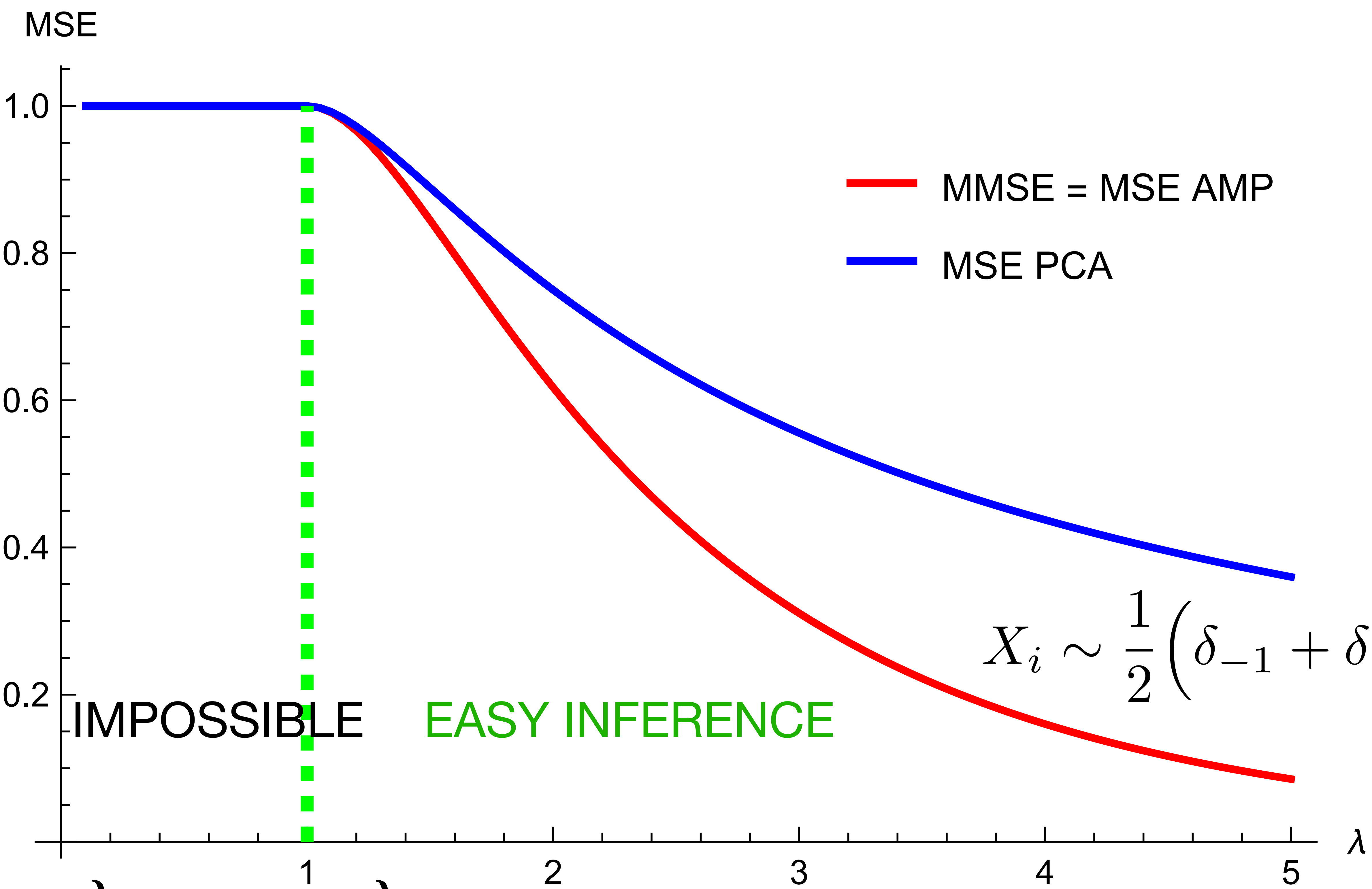
Wigner spiked model

(Johnstone 10')

$$M = 1$$



Bayesian Inference / Information theory

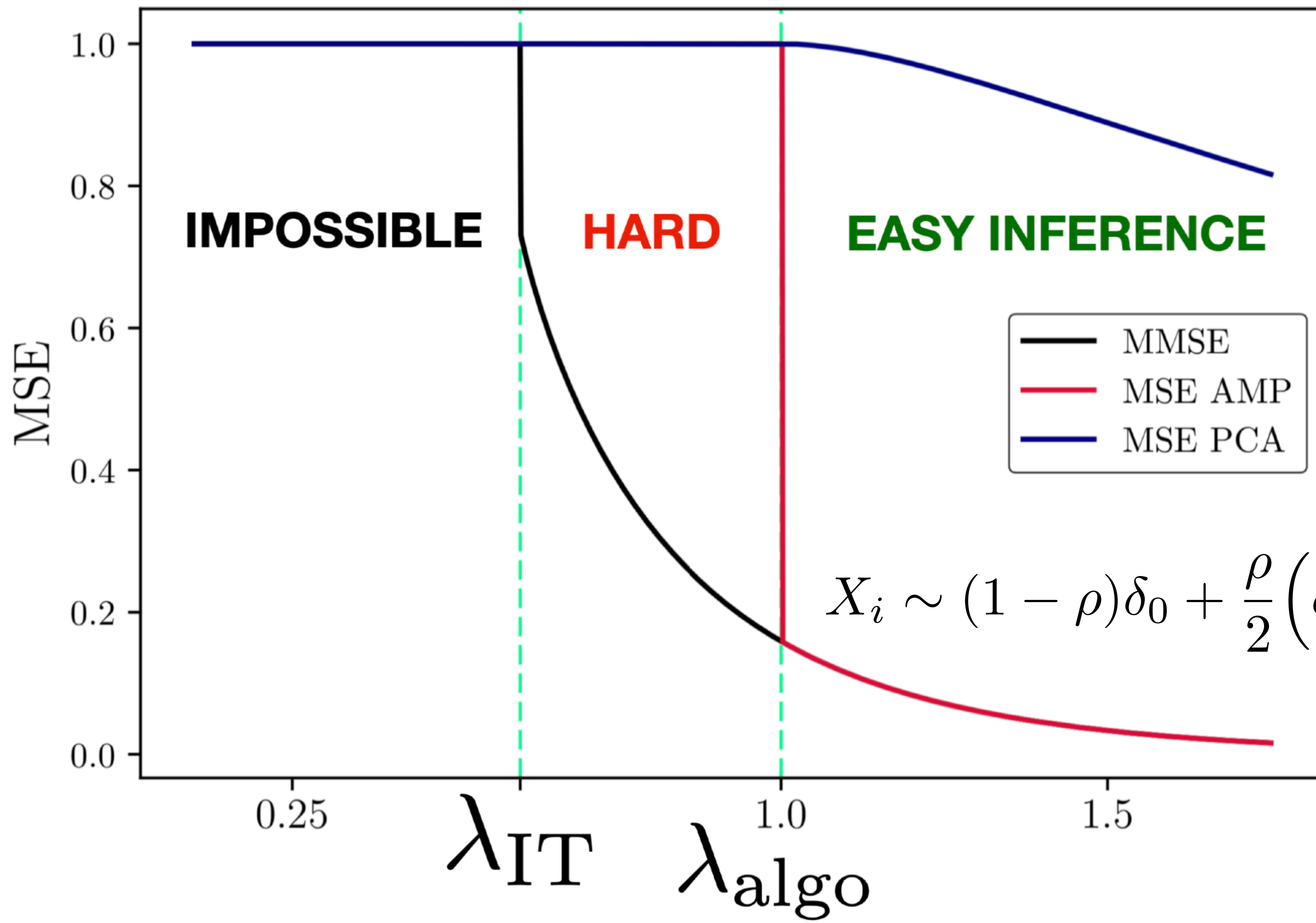


$$X_i \sim \frac{1}{2}\left(\delta_{-1} + \delta_1\right)$$

$$\lambda_{\mathrm{Algo}} = \lambda_{\mathrm{IT}}$$

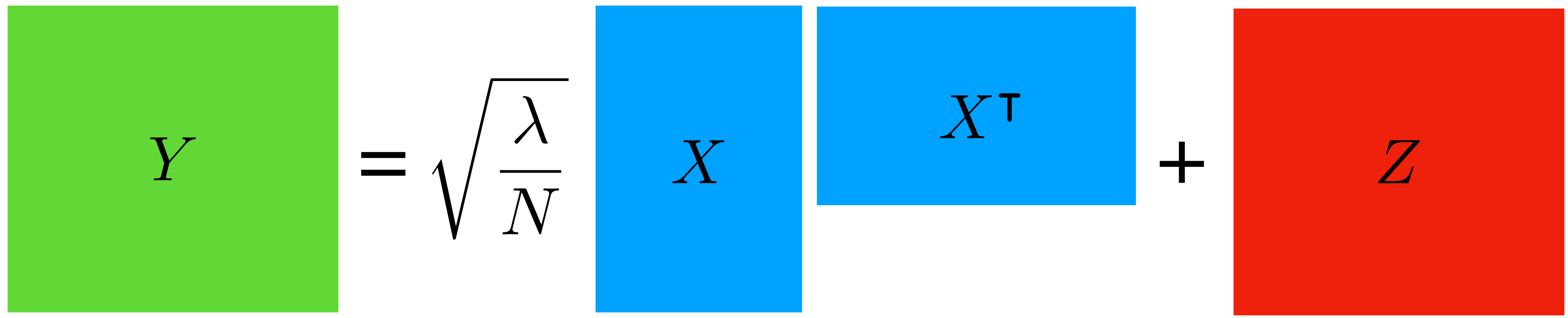
Macris, Korada
Montanari, Abbe, Deshpande
Macris, Barbier, Dia, Krzakala, Zdeborova, Lesieur

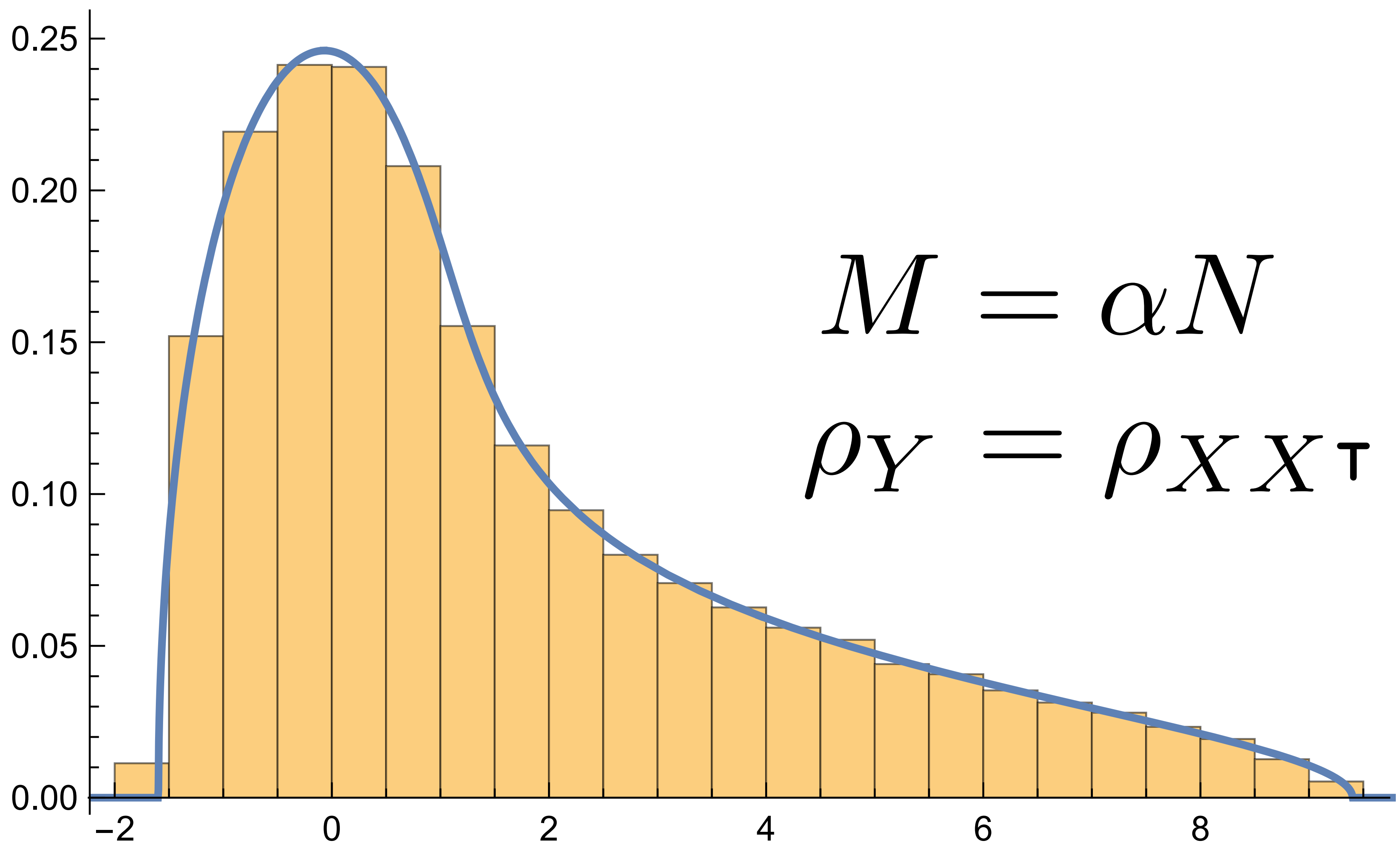
Bayesian Inference / Information theory



$$X_i \sim (1-\rho)\delta_0 + \frac{\rho}{2}\Big(\delta_{-1} + \delta_1\Big)$$

Macris, Korada
Montanari, Abbe, Deshpande
Macris, Barbier, Dia, Krzakala, Zdeborova, Lesieur

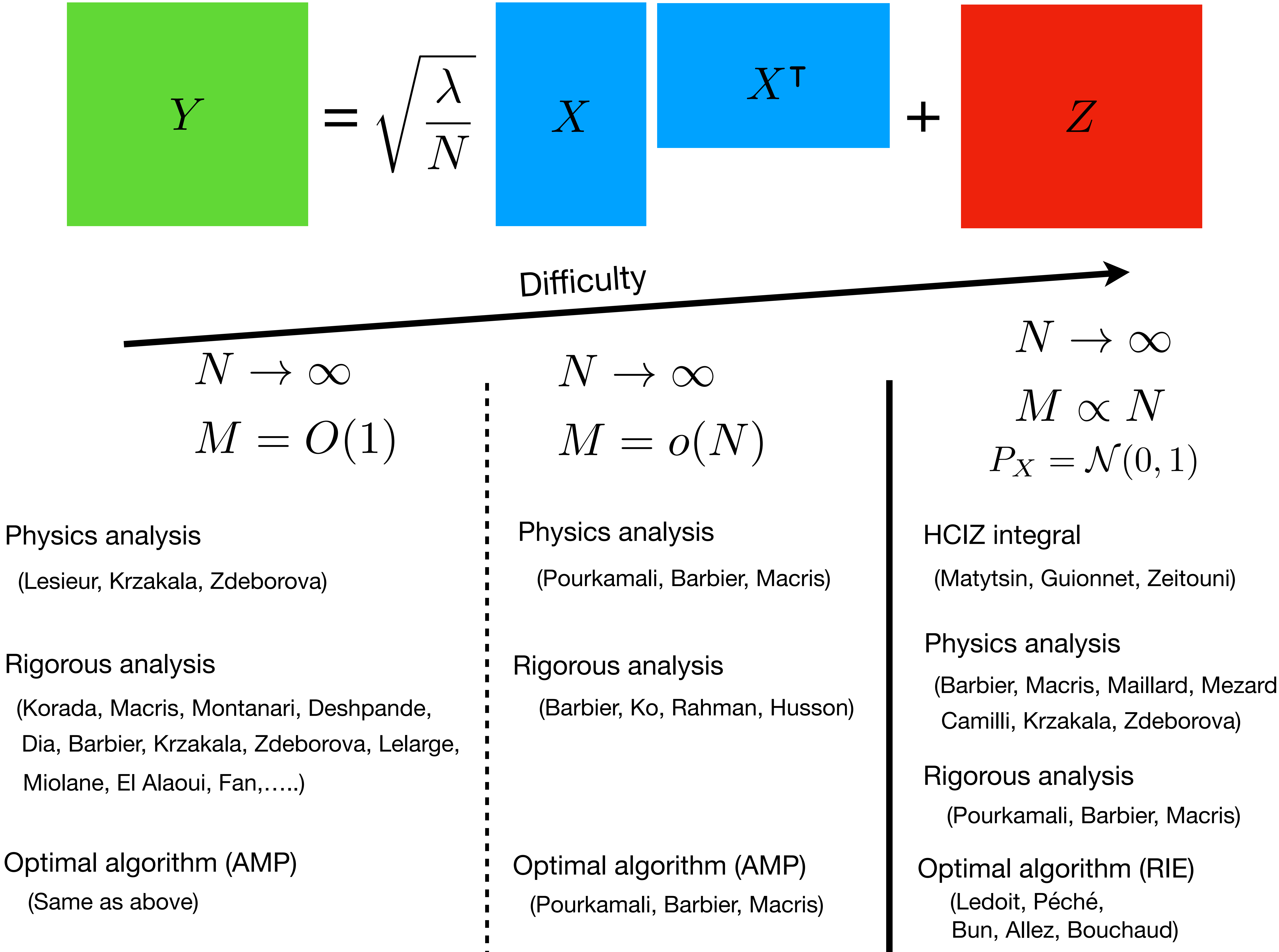




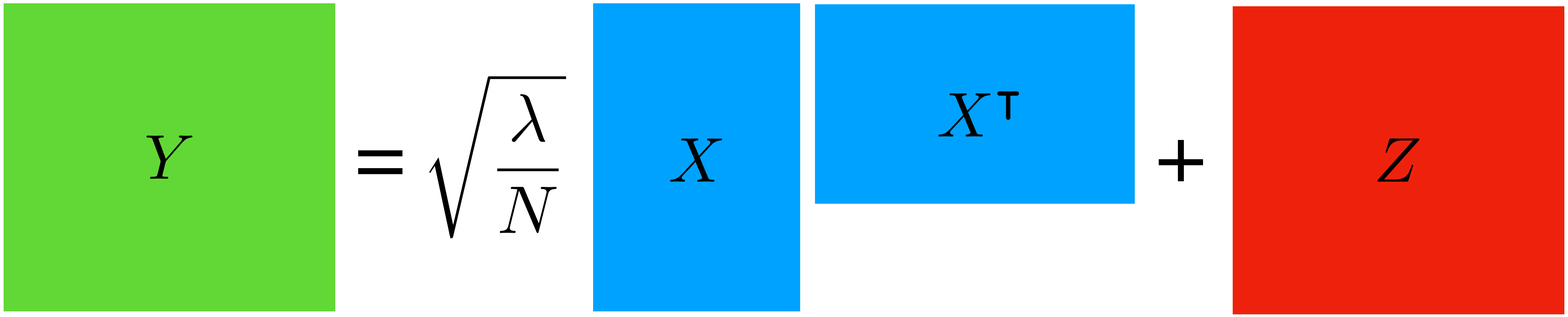
$$M = \alpha N$$

$$\rho_Y = \rho_{XX^\intercal} \boxplus \rho_Z$$

Spectrum (and other statistics) are **universal** in law of X entries, but…



$Y = \sqrt{\frac{\lambda}{N}} X X^\intercal + Z$
Difficulty
$N \to \infty$
$M = O(1)$
$N \to \infty$
$M = o(N)$
$N \to \infty$
$M \propto N$
$P_X = \mathcal{N}(0,1)$
Physics analysis
(Lesieur, Krzakala, Zdeborova)
Rigorous analysis
(Korada, Macris, Montanari, Deshpande, Dia, Barbier, Krzakala, Zdeborova, Lelarge, Miolane, El Alaoui, Fan,…..)
Optimal algorithm (AMP)
(Same as above)
Physics analysis
(Pourkamali, Barbier, Macris)
Rigorous analysis
(Barbier, Ko, Rahman, Husson)
Optimal algorithm (AMP)
(Pourkamali, Barbier, Macris)
HCIZ integral
(Matytsin, Guionnet, Zeitouni)
Physics analysis
(Barbier, Macris, Maillard, Mezard, Camilli, Krzakala, Zdeborova)
Rigorous analysis
(Pourkamali, Barbier, Macris)
Optimal algorithm (RIE)
(Ledoit, Péché, Bun, Allez, Bouchaud)



Difficulty →

$N \to \infty$

$M = O(1)$

Physics analysis

(Lesieur, Krzakala, Zdeborova)

Rigorous analysis

(Korada, Macris, Montanari, Deshpande, Dia, Barbier, Krzakala, Zdeborova, Lelarge, Miolane, El Alaoui, Fan,…..)

Optimal algorithm (AMP)

(Same as above)

$N \to \infty$

$M = o(N)$

Physics analysis

(Pourkamali, Barbier, Macris)

Rigorous analysis

(Barbier, Ko, Rahman, Husson)

Optimal algorithm (AMP)

(Pourkamali, Barbier, Macris)

$N \to \infty$

$M \propto N$

$P_X \neq \mathcal{N}(0,1)$

HCIZ integral

(Matytsin, Guionnet, Zeitouni)

Physics analysis

(Barbier, Macris, Maillard, Mezard, Camilli, Krzakala, Zdeborova)

Rigorous analysis

(Pourkamali, Barbier, Macris)

Optimal algorithm (RIE)

(Ledoit, Péché, Bun, Allez, Bouchaud)

Cross-breeding a matrix model with an Hopfield model

(Not really) a matrix model

$$\mathcal{Z} = \int_{\mathbb{R}^{N\times M}} e^{\frac{1}{2}\sqrt{\frac{\lambda}{N}}\mathrm{Tr} Y xx^\intercal - \frac{\lambda}{4N}\|xx^\intercal\|^2} \prod_{i,\mu}^{N,M} dP_X(x_{i\mu})$$

Matrix model WITHOUT rotational invariance (so not really one…)

$$\mathcal{Z}_{\mathrm{MM}} = \int e^{\frac{1}{2}\sqrt{\frac{\lambda}{N}}\mathrm{Tr} Y xx^\intercal - \frac{\lambda}{4N}\|xx^\intercal\|^2} \prod_{i,\mu}^{N,M} dx_{i\mu}$$

$$= \int e^{-\frac{\lambda}{4N}\|\Sigma\|^2}\Big(\int e^{\frac{1}{2}\sqrt{\frac{\lambda}{N}}\mathrm{Tr} Y O\Sigma O^\intercal} d\mu(O)\Big) d\Sigma$$

$xx^\intercal \to O\Sigma O^\intercal$

HCIZ / spherical integral +
saddle point over spectrum

A more complicated Hopfield model

$$\mathcal{Z} = \int_{\mathbb{R}^{N\times M}} e^{\frac{1}{2}\sqrt{\frac{\lambda}{N}}\sum_{\mu\leq M} x_\mu^\intercal Y x_\mu - \frac{\lambda}{4N}\sum_{\mu,\nu}(x_\mu^\intercal x_\nu)^2} \prod_{i,\mu}^{N,M} dP_X(x_{i\mu})$$

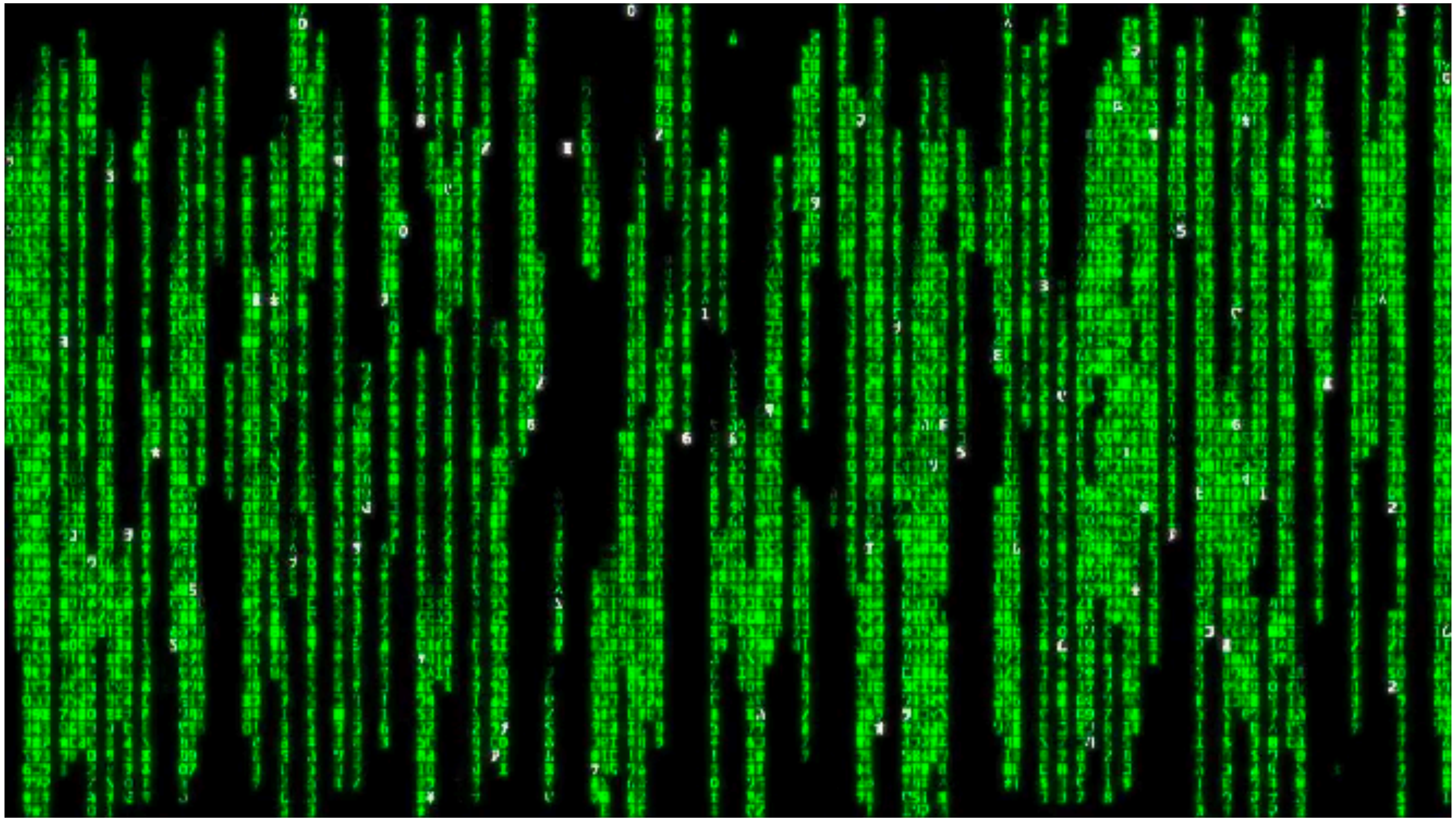
$$Y = \sqrt{\frac{\lambda}{N}} \sum_{\mu\leq M} X_\mu X_\mu^\intercal + Z$$

An Hopfield model where ALL patterns must be recovered JOINTLY

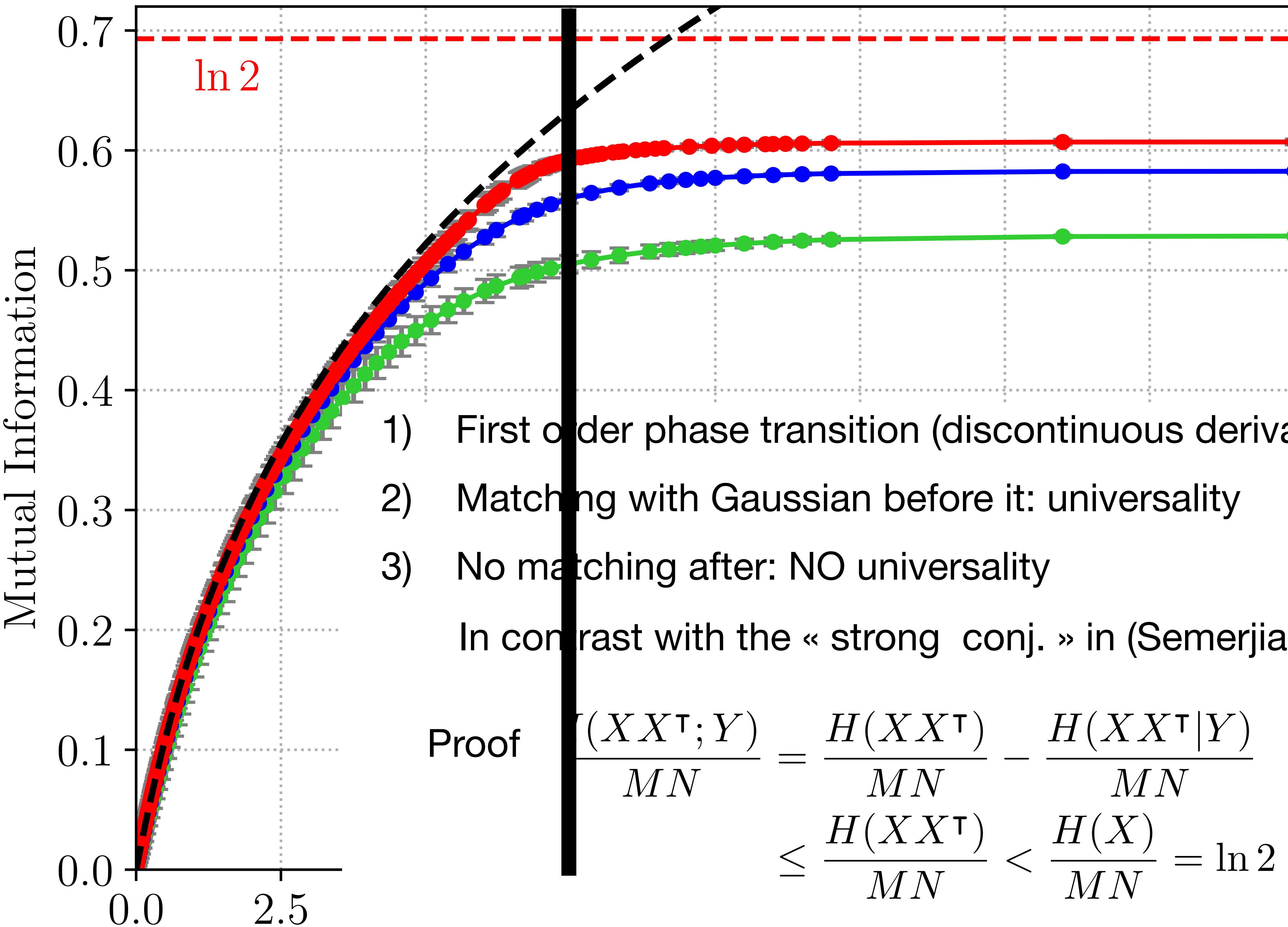
$$\mathcal{Z}_{\text{Hopfield}} = \int_{\mathbb{R}^N} e^{\frac{1}{2}\sqrt{\frac{\lambda}{N}} x^\intercal J x} \prod_i^N dP_X(x_i)$$

$$J = \frac{1}{\sqrt{N}} \sum_{\mu\leq M} X_\mu X_\mu^\intercal$$

Numerical insights



$$P_X = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1} \qquad M = \alpha N = N/2$$



1) First order phase transition (discontinuous derivative)

2) Matching with Gaussian before it: universality

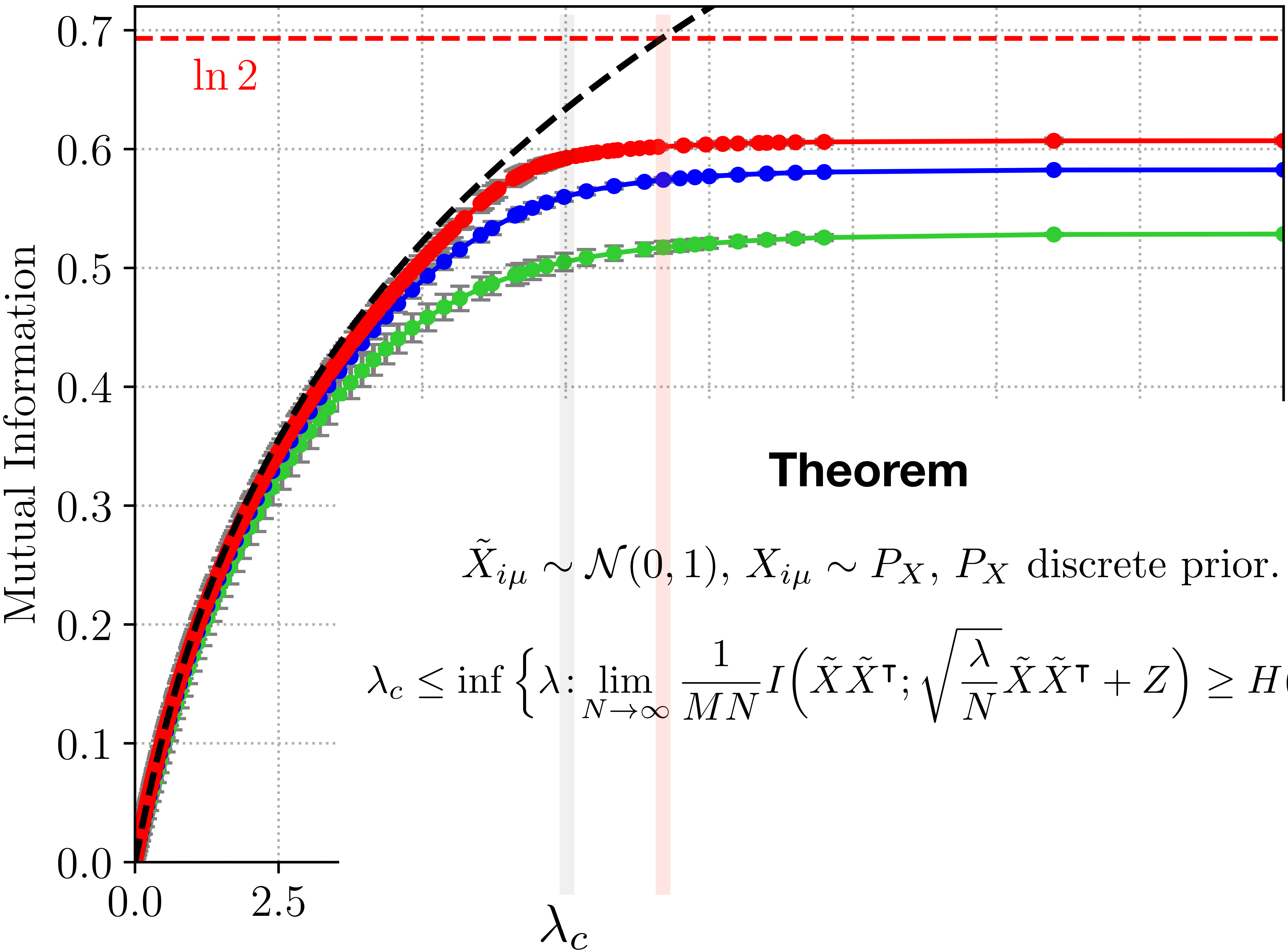
3) No matching after: NO universality

In contrast with the « strong conj. » in (Semerjian 24)

Proof
$$\frac{I(XX^\intercal;Y)}{MN} = \frac{H(XX^\intercal)}{MN} - \frac{H(XX^\intercal|Y)}{MN}$$
$$\leq \frac{H(XX^\intercal)}{MN} < \frac{H(X)}{MN} = \ln 2$$

+ MI Gaussian prior unbounded in SNR

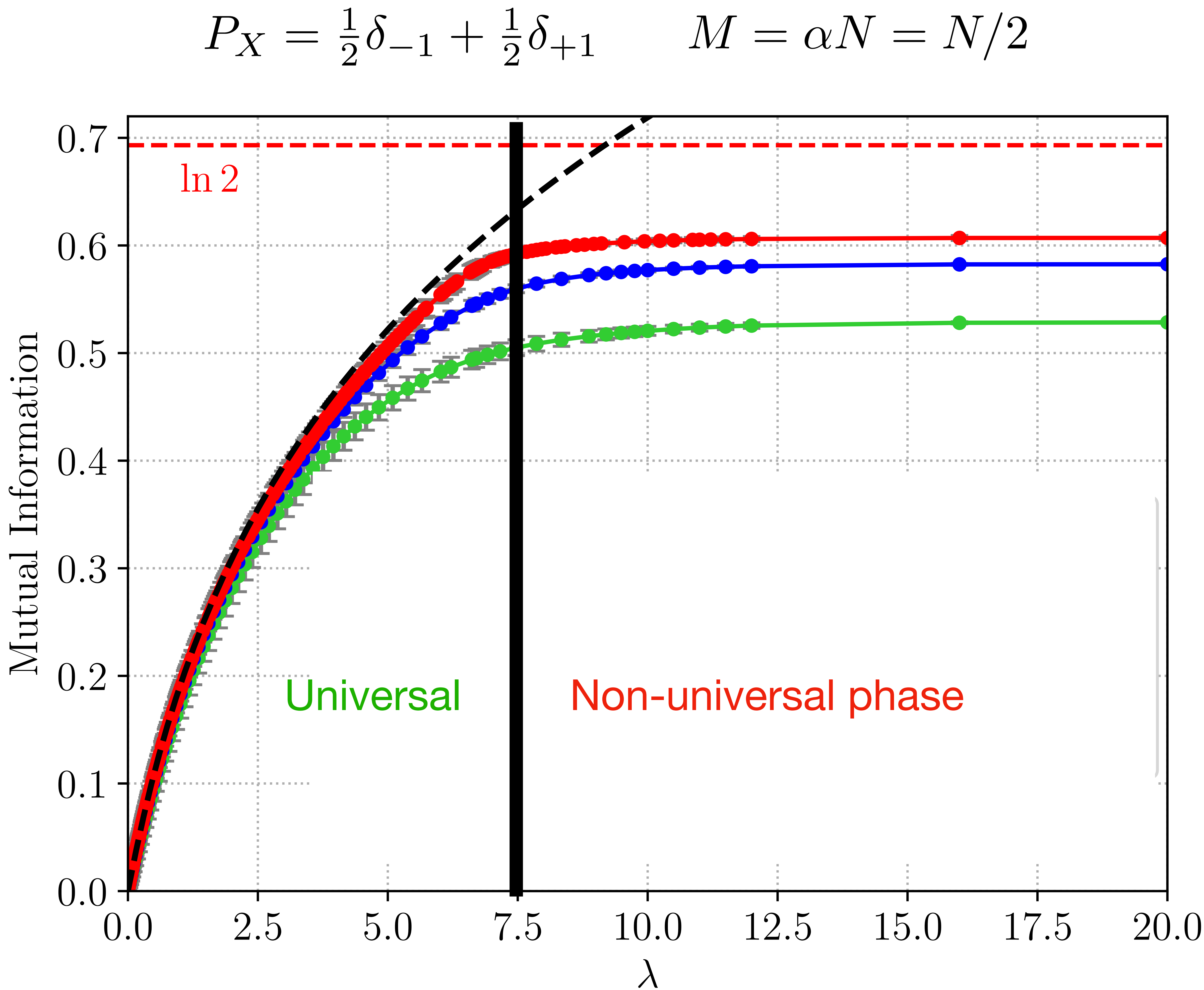
$$P_X = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1} \qquad M = \alpha N = N/2$$



Theorem

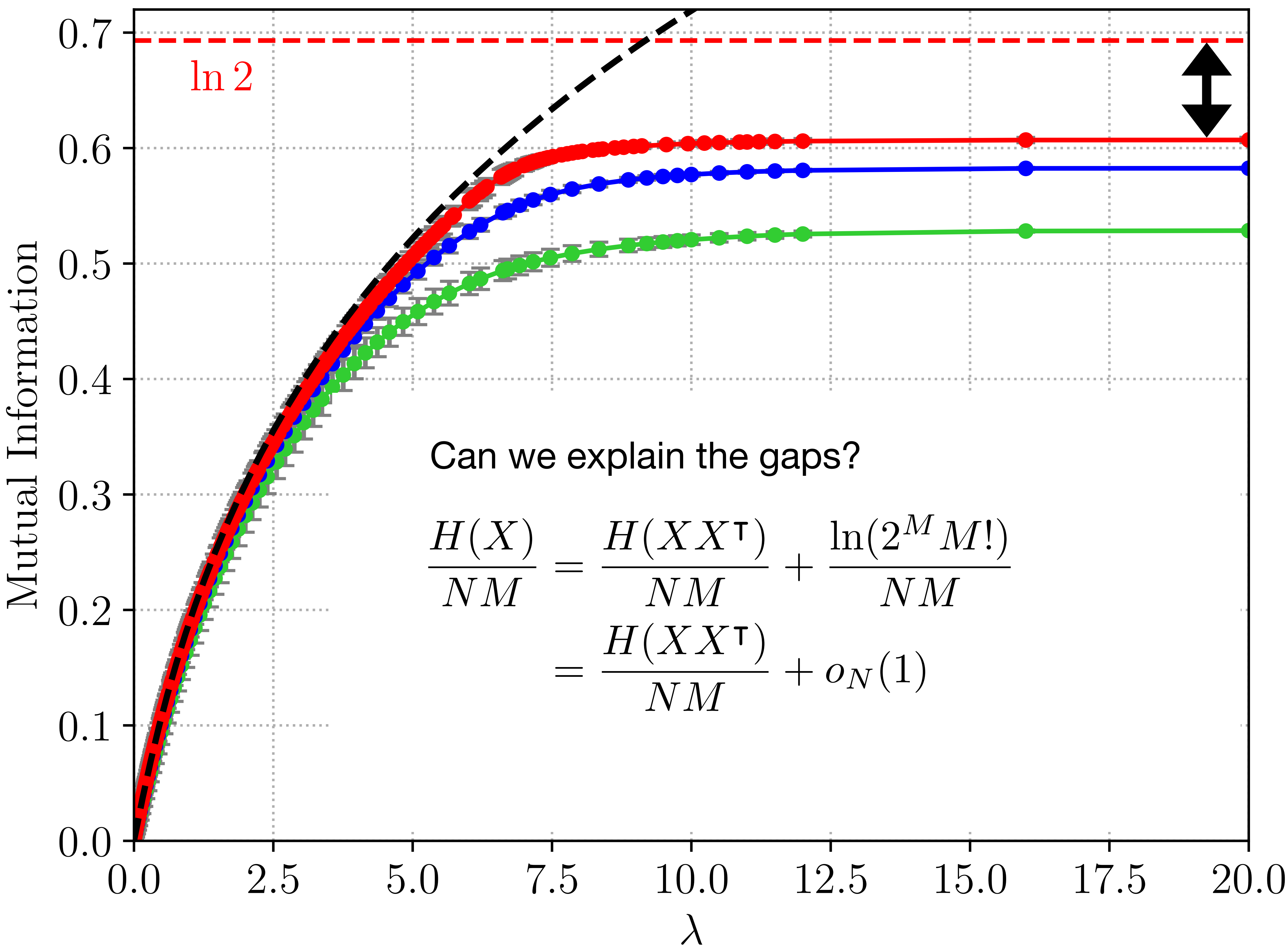
$\tilde{X}_{i\mu} \sim \mathcal{N}(0,1)$, $X_{i\mu} \sim P_X$, P_X discrete prior.

$$\lambda_c \leq \inf\left\{\lambda \colon \lim_{N\to\infty} \frac{1}{MN} I\left(\tilde{X}\tilde{X}^\intercal; \sqrt{\frac{\lambda}{N}}\tilde{X}\tilde{X}^\intercal + Z\right) \geq H(X)\right\}$$

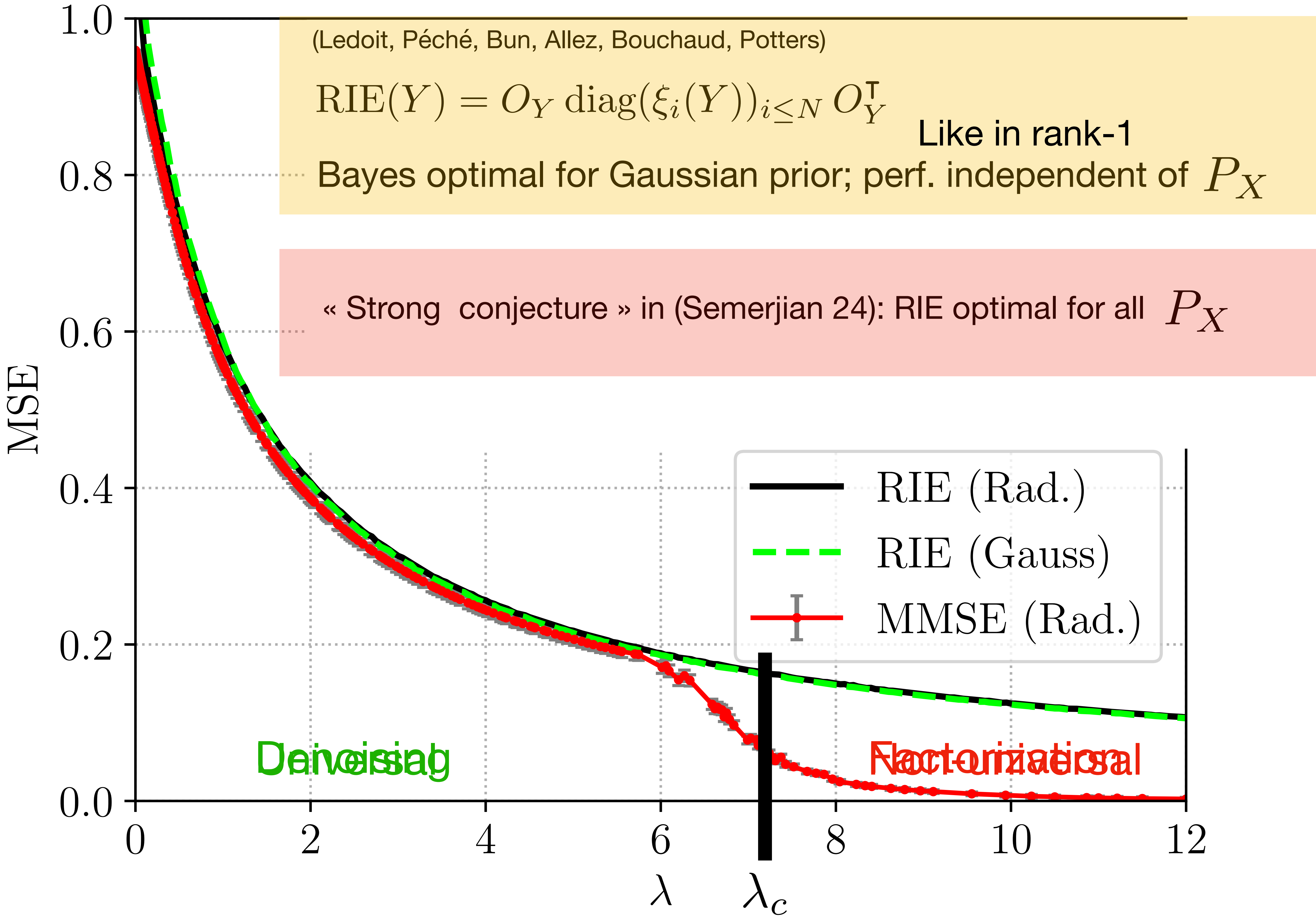


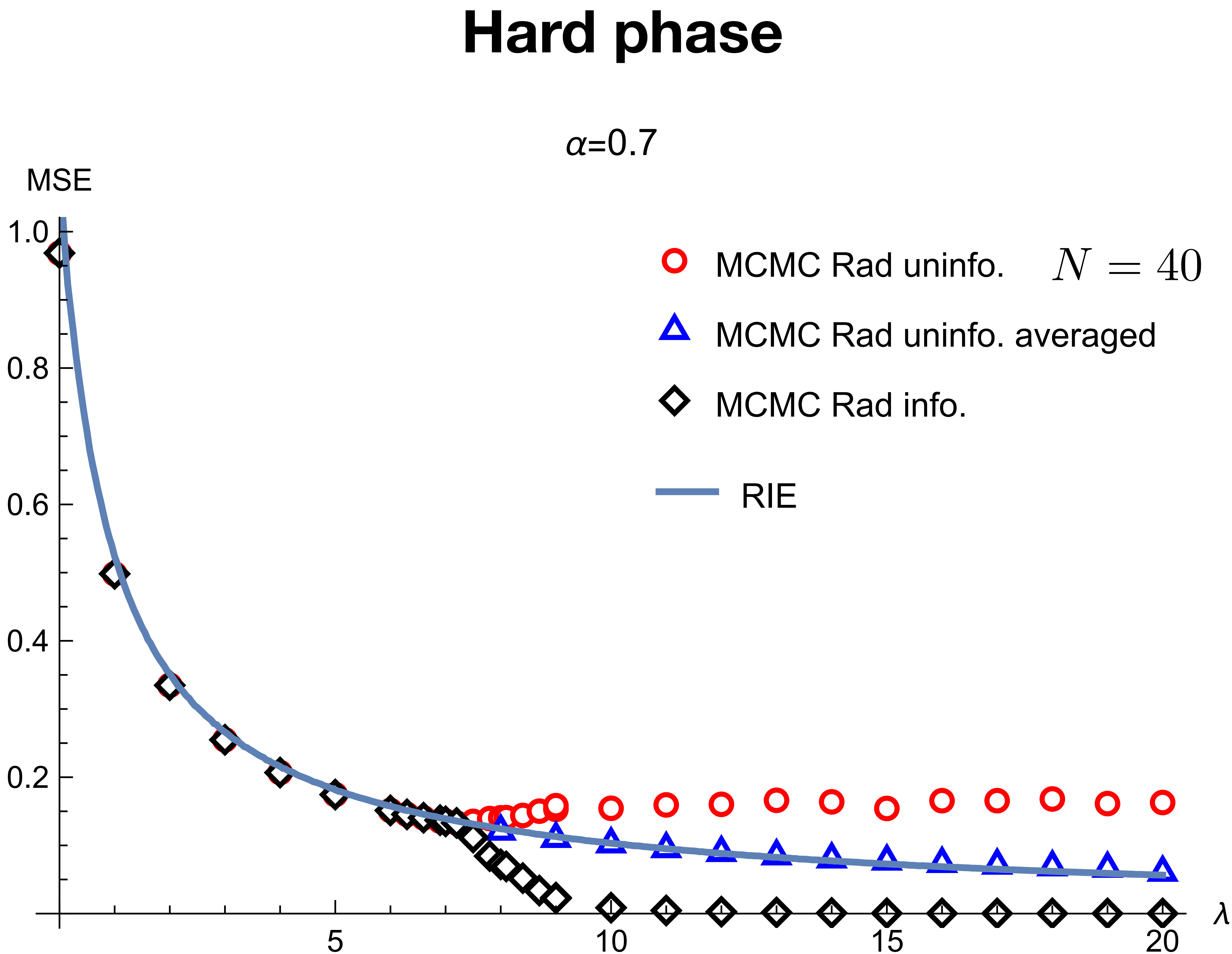
$P_X = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$ $M = \alpha N = N/2$
Mutual Information
$\ln 2$
Universal
Non-universal phase
λ
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

$$P_X = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1} \qquad M = \alpha N = N/2$$



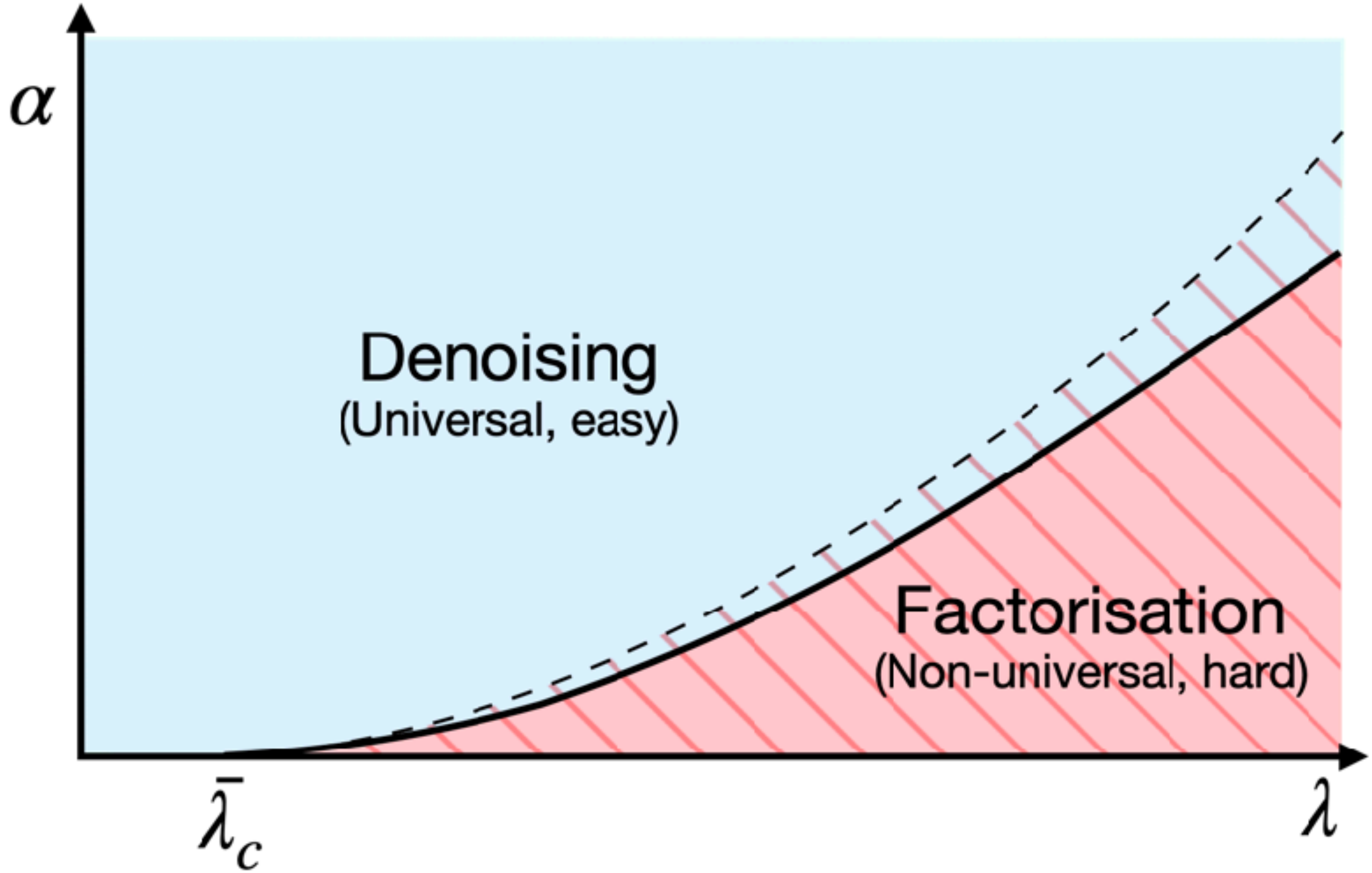
$$\frac{1}{MN^2}\mathbb{E}\|XX^\intercal - \mathbb{E}[xx^\intercal|Y]\|^2 = 4\frac{1}{MN}\frac{d}{d\lambda}I(XX^\intercal;Y)$$





Hard phase
α=0.7
MSE
1.0
0.8
0.6
0.4
0.2
5
10
15
20
λ
MCMC Rad uninfo.
$N = 40$
MCMC Rad uninfo. averaged
MCMC Rad info.
RIE

Phase diagram



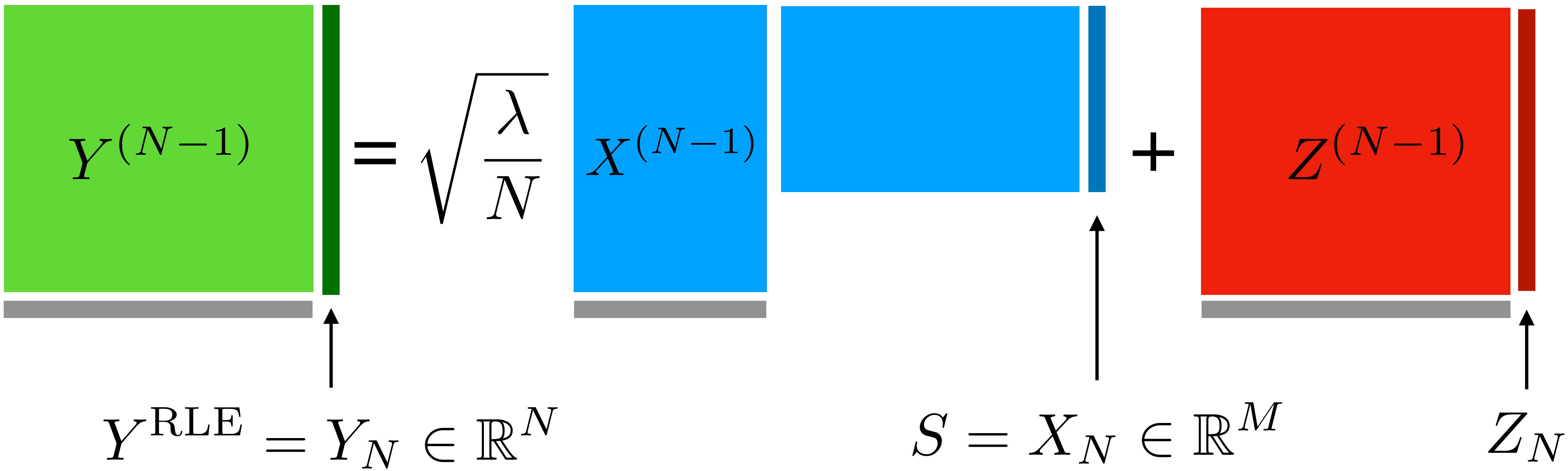
Information-theoretically, recovering ALL ``patterns” at once is roughly equally hard as recovering just one (rank-one problem).

Algorithmically it is a very different story…

Multiscale mean-field theory

$$O(MN) \longrightarrow O(1)$$

$O(MN) \longrightarrow O(M)$ **cavity method**



Effective M-dim pb: random linear estimation with uncertain design

$$Y_N = \sqrt{\frac{\lambda}{N}} X^{(N-1)} S + Z_N \qquad Y^{(N-1)}(X^{(N-1)})$$

$$(s, x^{(N-1)}) \sim P(s|Y_N, x^{(N-1)}) \times P(x^{(N-1)}|Y^{(N-1)})$$

$$Y_N = \sqrt{\frac{\lambda'}{N}} X^{(N-1)} S + Z_N \qquad Y^{(N-1)}(X^{(N-1)})$$

$$\frac{1}{MN^2}\mathbb{E}\|XX^\intercal - \mathbb{E}[xx^\intercal|Y]\|^2 = \frac{1}{MN^2}\sum_{i,j=1}^{N}\left(\mathbb{E}(X_i^\intercal X_j)^2 - \mathbb{E}\langle x_i^\intercal x_j\rangle^2\right)$$

Measurement MMSE

(Barbier, Macris, Dia et al. 17')

$$Y_N = \sqrt{\frac{\lambda'}{N}} X^{(N-1)} S + Z_N \qquad Y^{(N-1)}(X^{(N-1)})$$

$$(s, x^{(N-1)}) \sim P(s|Y_N, x^{(N-1)}) \times P(x^{(N-1)}|Y^{(N-1)})$$

Bulk measure simplification: mean-field ansatz

$$P(x^{(N-1)}|Y^{(N-1)})$$

$$\Rightarrow P(x^{(N-1)}|Y_\sigma^{\mathrm{eff,bulk}}) = \prod_{i,\mu} P(x_{i\mu}|Y_{i\mu}^{\mathrm{eff,bulk}} = \sqrt{\sigma} X_{i\mu} + Z_{i\mu}^{\mathrm{eff,bulk}})$$

A bit of work + Barbier, Macris, Miolane et al. PNAS 19'

$$\Rightarrow I\big((X^{(N-1)}, S); Y_N | Y_\sigma^{\mathrm{eff,bulk}}\big)$$ Low-dim. ``Replica symmetric formula”

$$O(M) \longrightarrow O(1)$$

$\sigma?$

Consistency between cavity and bulk (random) marginals

-> Exchangeability among rows

$$s_\mu \sim P(s_\mu | Y_\mu^{\mathrm{eff,cavity}} = \sqrt{r(\sigma,\lambda)} S_\mu + Z_\mu^{\mathrm{eff,cavity}})$$

$$x_{i\mu} \sim P(x_{i\mu} | Y_{i\mu}^{\mathrm{eff,bulk}} = \sqrt{\sigma} X_{i\mu} + Z_{i\mu}^{\mathrm{eff,bulk}})$$

$$\Rightarrow \sigma(\lambda) = r(\sigma,\lambda)$$

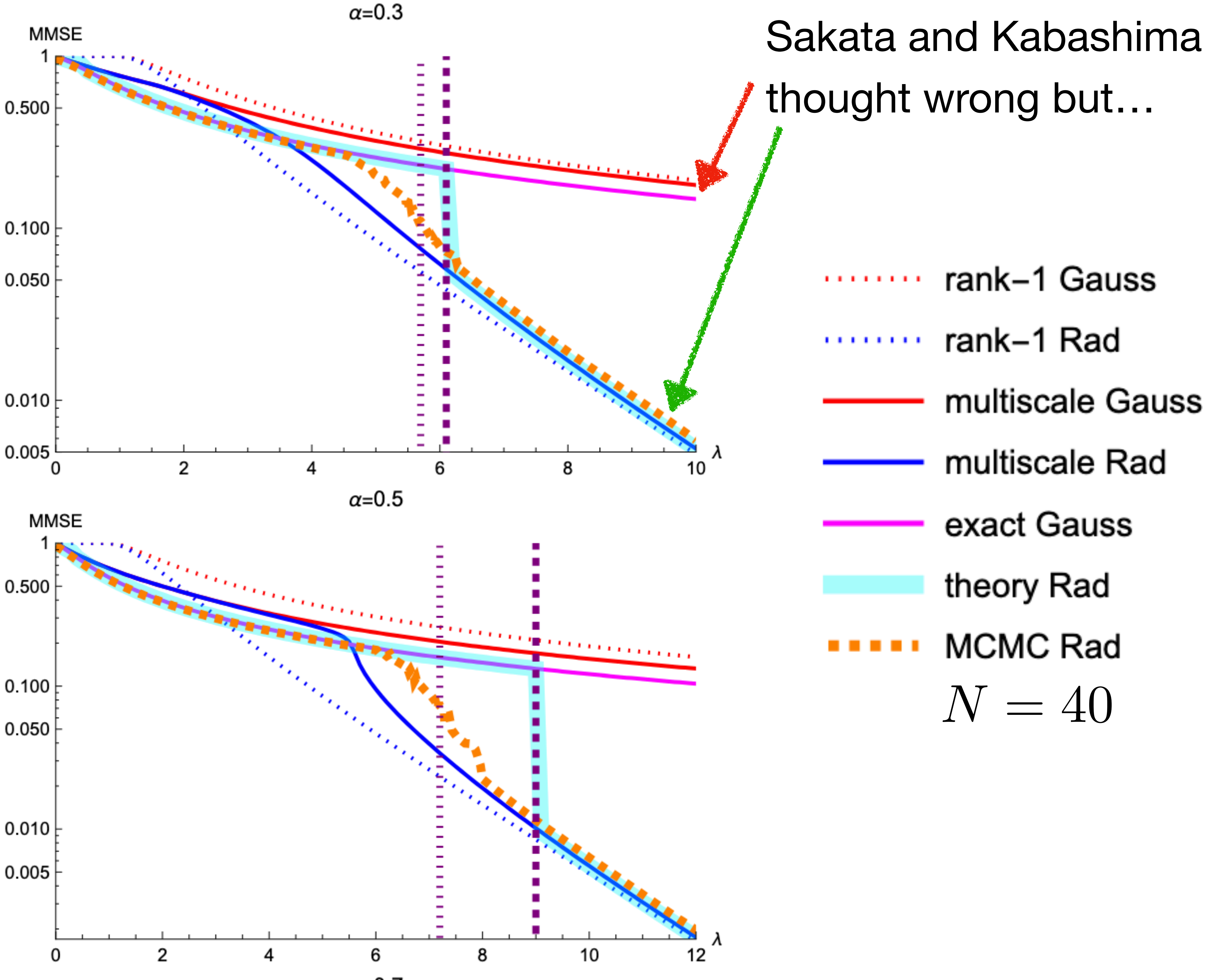
$$\Rightarrow I\Big((X^{(N-1)}, S); Y_N | Y_\sigma^{\mathrm{eff}}\Big)$$

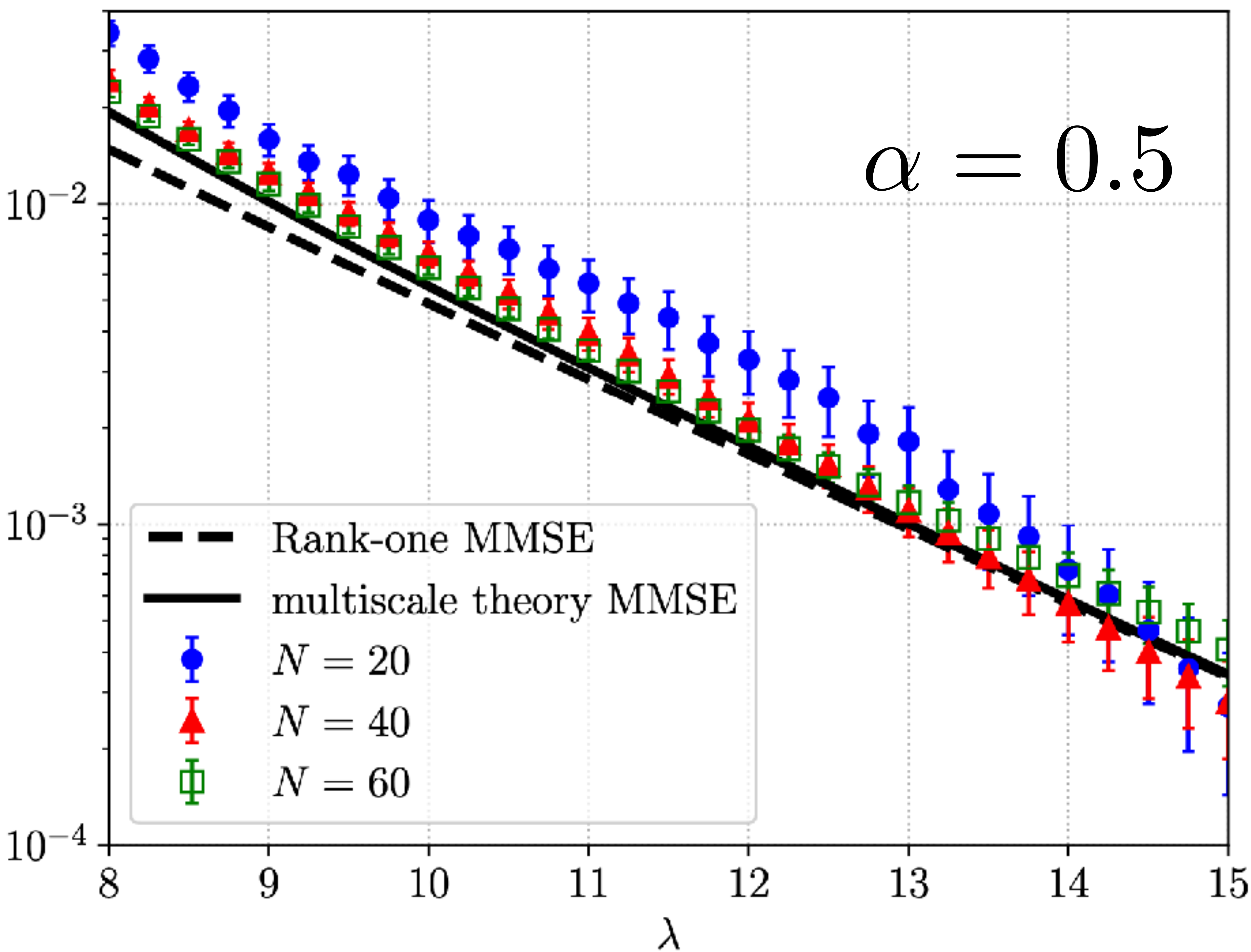
$$\Rightarrow \frac{1}{MN^2} \mathbb{E} \| XX^\intercal - \mathbb{E}[xx^\intercal | Y] \|^2$$

Matches Sakata and Kabashima replica prediction 13’

Sakata and Kabashima 13'

thought wrong but…





$\alpha = 0.5$
Rank-one MMSE
multiscale theory MMSE
$N = 20$
$N = 40$
$N = 60$
10^{-2}
10^{-3}
10^{-4}
8
9
10
11
12
13
14
15
λ

Nature of the phase transition

λ_c

Matrix overlap: is factorisation possible?

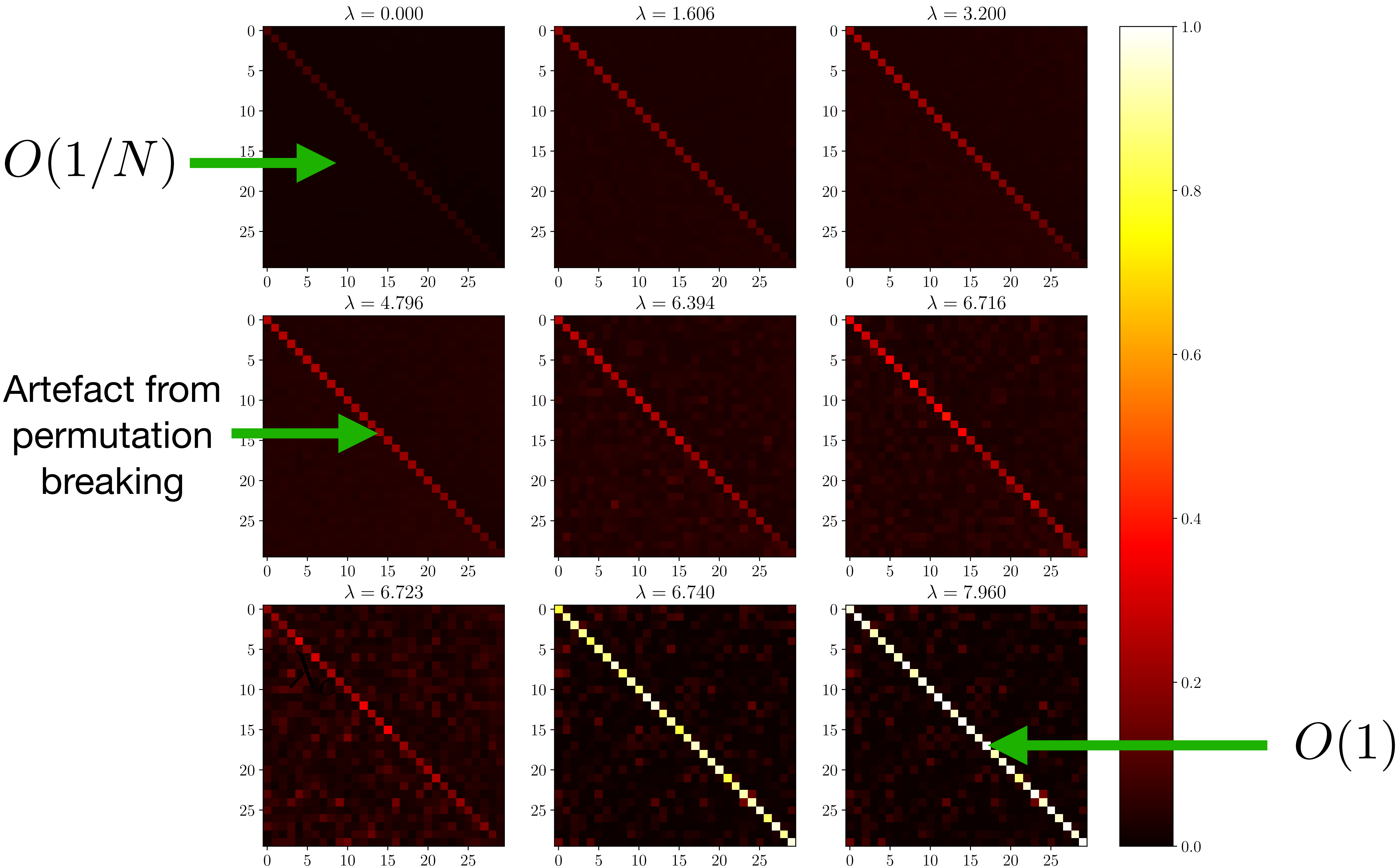
MMSE not enough to probe everything

Optimally column-permuted

Overlap $$Q = \frac{X^\intercal x}{N} = \Big(\frac{X_\mu^\intercal x_\nu}{N}\Big)_{\mu,\nu\leq M} \in \mathbb{R}^{M\times M} \qquad x \sim P(\,\cdot\,|Y)$$

$$\frac{1}{MN^2}\mathbb{E}\|XX^\intercal - \langle xx^\intercal\rangle\|^2 = \frac{1}{M}\Big(\mathbb{E}\Big\|\frac{X^\intercal X}{N}\Big\|^2 - \mathbb{E}\langle\|Q\|^2\rangle\Big)$$

$$N^{-2}\langle (X^\intercal x)^{\odot 2} \rangle$$



Nature of the phase transition

Universal

Information-theoretic quantities (MI and MMSE) asymptotically <u>independent</u> of $P_X, \mathbb{E}X = 0, \mathbb{E}X^2 = 1$

Denoising

``Effective rotational invariance" -> <u>RIE is optimal.</u>

$XX^\intercal$ can be estimated by RIE, X cannot

Mixed

Posterior patterns $(x_\mu)_{\mu \leq M}$ have $O(1/\sqrt{N})$ projections on planted patterns (X_μ) but <u>larger than in random case</u> -> Information mixed.

Delocalised

Posterior patterns $(x_\mu)_{\mu \leq M}$ have $O(1/\sqrt{N})$ projections in the ``quasi basis" (X_μ)

No feature learning

The internal features (factorised, discrete) of $S = XX^\intercal$ <u>are not seen</u>.

λ_c

Non-universal

… <u>dependent</u> of prior.

Factorisation

Breaking of invariance -> <u>RIE is NOT optimal.</u>

Better strategy exits, which allows to reconstruct X

Ferromagnetic Retrieval

Each posterior pattern has $O(1)$ overlap with one planted one.

Localised

… $O(1)$ …

Feature learning

The internal features are exploited.

Conclusion

- Numerical insights on the phase diagram of matrix denoising with extensive rank: a model that ``interpolates” between an RMT-like matrix model at low SNR and a mean-field spin model at high SNR
- Denoising/Universality -> Factorisation/Non-universality: 1st order transition. Extensive-rank generalisation of the BBP transition and its Bayesian counterpart
- Simple and versatile multiscale mean-field theory
- Predicts that factorisation is hard but possible: RIE best poly algorithm?
- Paves the way for the analysis of more complex inference, spin and matrix models

Work in progress

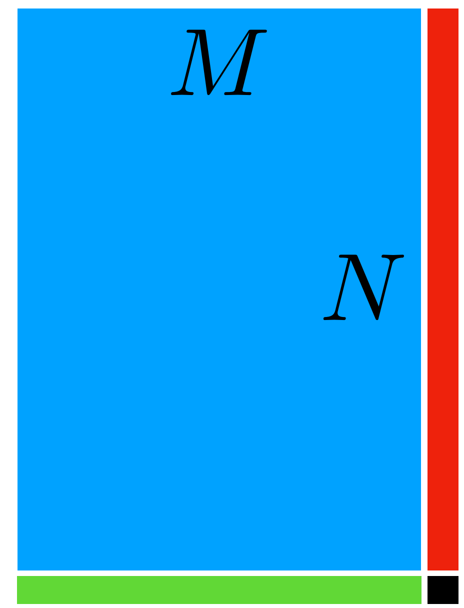
- More rigorous control
- Other extensive-rank models (Hopfield, Bayesian neural networks,…)

In addition

$O(MN) \longrightarrow O(M)$ **cavity method**

Usually $$\frac{1}{N}\mathbb{E}\ln \mathcal{Z}_N = \frac{1}{N}\sum_{i\le N-1}\big(\mathbb{E}\ln\mathcal{Z}_{i+1} - \mathbb{E}\ln\mathcal{Z}_i\big) \approx \lim_{N\to\infty}(\mathbb{E}\ln\mathcal{Z}_{N+1} - \mathbb{E}\ln\mathcal{Z}_N)$$

$X =$



Theorem 1 (Multiscale Aizenman–Sims–Starr identity). *Let $M = M_N = \alpha N$. We have*

$$\lim_{N\to\infty}\phi_{N,M} = \frac{1}{2}\lim_{N\to\infty}\frac{\Delta_{\mathrm{row}}(N)}{M} + \frac{1}{2}\lim_{N\to\infty}\frac{\Delta_{\mathrm{col}}(N)}{N}$$

where

$$\Delta_{\mathrm{row}}(N) := \mathbb{E}\ln\mathcal{Z}_{N+1,M_{N+1}} - \mathbb{E}\ln\mathcal{Z}_{N,M_{N+1}},$$
$$\Delta_{\mathrm{col}}(N) := \mathbb{E}\ln\mathcal{Z}_{N,M_N+1} - \mathbb{E}\ln\mathcal{Z}_{N,M_N}.$$

Proposition 2 (Equivalence of cavity representations).

Suppose that $\lim_{N\to\infty}\frac{\Delta_{\mathrm{row}}(N)}{M_N}$ *and* $\lim_{N\to\infty}\frac{\Delta_{\mathrm{col}}(N)}{N}$ *exist and they are respectively equal to L_{row} and L_{col}. Then $L_{\mathrm{row}} = L_{\mathrm{col}}$.*