SIMONS INSTITUTE for the Theory of Computing

EPFL

Mirror event!

Rigorous Evidence for Information-Computation Trade-offs, Sep. 2021

IdEPHIX

INFORMATION, LEARNING & PHYSICS LAB.

TO BE (HARD) OR NOT BE: THE SPIKED MATRIX-TENSOR MODEL

Florent Krzakala

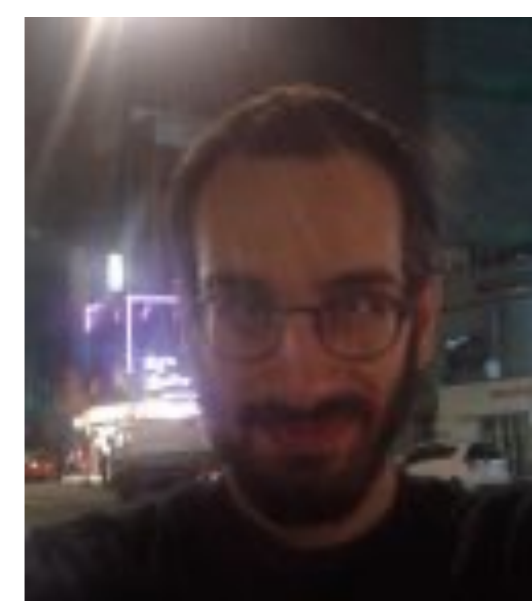
S. Sarao
(Saclay -> Oxford)

G. Biroli
(ENS Paris)

L.Zdeborova
(EPFL)

C. Chamarotta
(King's_> Roma)

P. Urbani
(Ipht Saclay)



GENERIC QUESTIONS:

For high-dimensional non-convex learning problems:

- How hard is it to sampling from the posterior ?
- How hard is it to compute the ML estimate?

- How does the energy/loss landscape affect the behaviour of gradient descent & sampling algorithms?
- Does the presence of spurious minima matter?

- How these approaches compare with message passing?

IN THIS TALK

An attempt to answer these questions in a simple yet generic problem

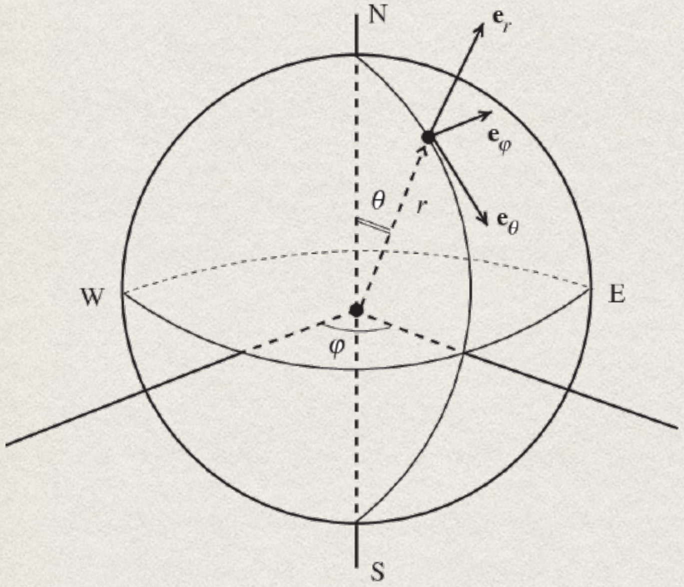
A synthetic problems where the *optimal performances* can be determined, the *energy landscape* characterised, and the behaviour of many algorithms (*Message passing, Sampling, Gradient descent*) analysed …

👉The Matrix-Tensor spiked model on the sphere

SPIKED MATRIX-TENSOR PROBLEM

$\mathbf{x}^* \in \mathbb{R}^N$

Choose a normed vector $\|\mathbf{x}^*\|_2^2 = \mathbf{N}$, randomly on the sphere in N-dimension



Create a rank-noise noisy (symmetric) matrix

$$Y_{ij} = \frac{1}{\sqrt{N}} x_i^* x_j^* + \xi_{ij} \qquad \xi_{ij} \sim \mathcal{N}(0, \Delta_2)$$

Create a rank-noise noisy (symmetric) tensor

$$T_{i_1 \dots i_p} = \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} x_{i_1}^* \dots x_{i_p}^* + \xi_{i_1 \dots i_p}$$

$$\xi_{i_1, \dots, i_p} \sim \mathcal{N}(0, \Delta_p)$$

Given the matrix T and the tensor Y, can one recover **x***?

SPIKED MATRIX-TENSOR PROBLEM

- For the same signal $\mathbf{x}^*$ in $\mathbb{R}^N$ & observe a matrix Y <u>and</u> a tensor T:

$$Y = \frac{1}{\sqrt{N}} \mathbf{x}^* \mathbf{x}^{*T} + \sqrt{\Delta_2} W \qquad W_{ij} \sim \mathcal{N}(0,1)$$

$$T = \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} \mathbf{x}^{*\otimes p} + \sqrt{\Delta_p} Z \qquad Z_{ijk} \sim \mathcal{N}(0,1)$$

- Can one recover $\mathbf{x}^*$ from T and Y?

SPIKED MATRIX-TENSOR PROBLEM

PLANTED VERSION OF THE '2+P' SPIN GLASS IN STATISTICAL PHYSICS

- Define the Hamiltonian (or cost function):

$$\mathscr{H}(x) = -\frac{1}{\Delta_2\sqrt{N}}\sum_{i<j} Y_{ij}x_i x_j - \frac{\sqrt{(p-1)!}}{\Delta_p N^{(p-1)/2}}\sum_{i_1<\ldots<i_p} T_{i_1\ldots i_p} x_{i_1}\ldots x_{i_p}$$

spherical constraint: $\sum_{i=1}^{N} x_i^2 = N$

- Bayes-optimal estimation = marginals of Gibbs measure

$$\hat{\mathbf{x}} = \mathbb{E}_{P(\mathbf{X}|Y,T)}[\mathbf{x}] \qquad P_{\mathbf{Gibbs}}(\mathbf{x}\,|\,Y,T) = \frac{1}{Z(Y,T)}e^{-\mathscr{H}_{Y,T}(x)}$$

- MMSE = $\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2$

[Derrida 81, Mezard-Gross '84, many others]

SPIKED MATRIX-TENSOR PROBLEM

PLANTED VERSION OF THE '2+P' SPIN GLASS IN STATISTICAL PHYSICS

- Define the Hamiltonian (or cost function):

$$\mathcal{H}(x) = -\frac{1}{\Delta_2\sqrt{N}}\sum_{i<j} Y_{ij}x_i x_j - \frac{\sqrt{(p-1)!}}{\Delta_p N^{(p-1)/2}}\sum_{i_1<\ldots<i_p} T_{i_1\ldots i_p} x_{i_1}\ldots x_{i_p}$$

spherical constraint: $\sum_{i=1}^{N} x_i^2 = N$

- Maximum likelihood (MLE)

$$\mathcal{L}(\mathbf{x}) = \frac{1}{2\Delta_2}\|Y - \frac{\mathbf{x}\mathbf{x}^T}{\sqrt{N}}\|_2^2 + \frac{1}{p!\Delta_p}\|T - \sqrt{(p-1)!}\frac{\mathbf{x}^{\otimes p}}{N^{(p-1)/2}}\|_2^2$$

Minimize $\mathcal{L}(\mathbf{x})$ subject to $\|\mathbf{x}\|_2^2 = N$

①

Computational-Statistical Gaps: What do we know?

OPTIMAL LEARNING

INFORMATION THEORETIC PERFORMANCES

$$P_{\mathbf{Gibbs}}(\mathbf{x} \,|\, Y, T) = \frac{1}{Z(Y,T)} e^{-\mathcal{H}_{Y,T}(x)}$$

From the free energy, $\frac{1}{N} \log Z(Y)$, we can compute *anything*

Mutual Information $\quad \frac{I(X;Y)}{N} = \frac{(\mathbb{E}[X^2])^2}{4\Delta} - \mathbb{E}_Y \left[\frac{1}{N} \log Z(Y) \right]$

Likelihood ratio $\quad \frac{1}{N} \log \left(\frac{P_{\text{spiked}}(Y)}{P_{\text{null}}(Y)} \right) = \frac{1}{N} \log Z(Y)$

Kullback-Leibler $\quad D_{\text{KL}}(P_{\text{Spiked}} || P_{\text{Null}}) = \mathbb{E}_Y \log Z(Y)$

OPTIMAL LEARNING
INFORMATION THEORETIC PERFORMANCES

Theorem 1 (informally): “replica symmetric potential”

For large N, $\frac{1}{N}\log Z(Y,\Delta)$ concentrates around the max of $\Phi_{\rm RS}(m)$

$$\Phi_{\rm RS} = \frac{1}{2}\log(1-m) + \frac{m}{2} + \frac{m^2}{4\Delta_2} + \frac{m^p}{2p\Delta_p} \qquad m \in [0,1]$$

[Mannelli, Biroli, Cammarota, **FK**, Urbani, & Zdeborová ’18]

OPTIMAL LEARNING

INFORMATION THEORETIC PERFORMANCES

Theorem 1 (informally): "replica symmetric potential"

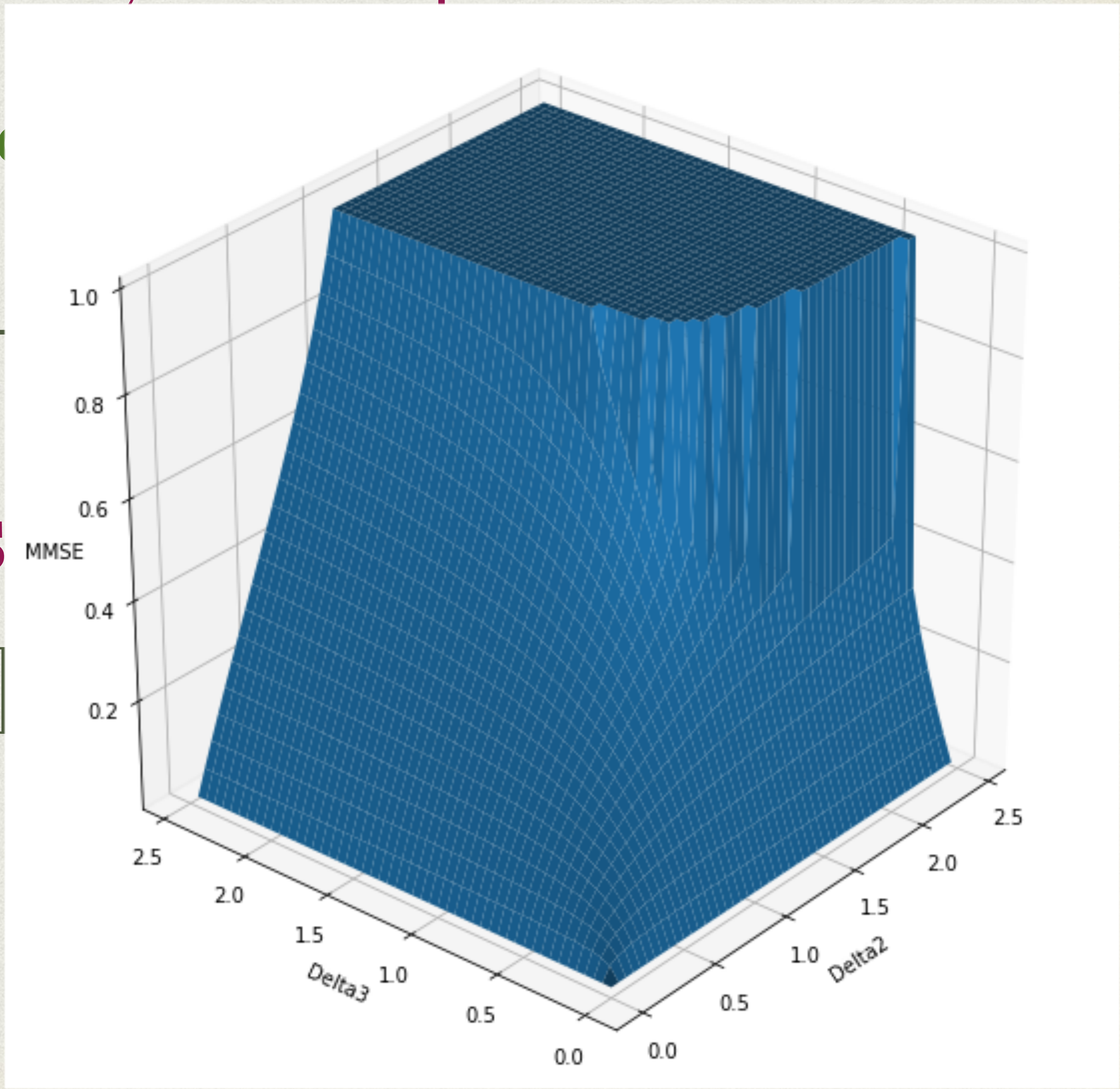
For large N, $\frac{1}{N} \log Z(Y, \Delta)$ conc

$$\Phi_{\mathrm{RS}} = \frac{1}{2}\log(1-m) + \frac{m}{2} +$$

Theorem 2 (informally): MMS

$$MMSE = 1 - \mathrm{argmax}\left[\Phi_{\mathrm{rs}}(m)\right]$$

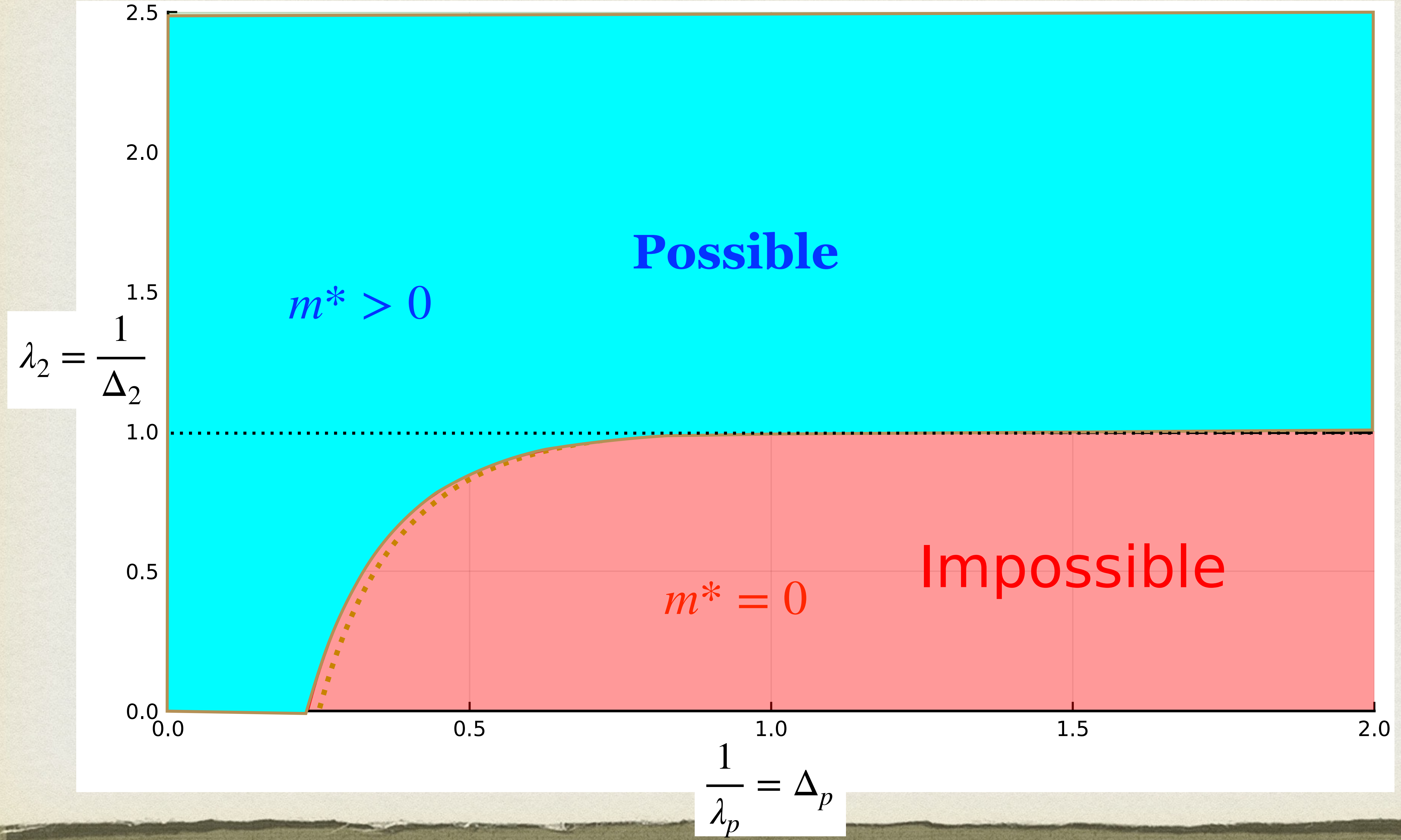
$$m* = 1 - \frac{1}{1 + \frac{m*}{\Delta_2} + \frac{(m*)^{p-1}}{\Delta_p}}$$



[Mannelli, Biroli, Cammarota, **FK**, Urbani, & Zdeborová '18]

OPTIMAL LEARNING

INFORMATION THEORETIC PERFORMANCES



WHAT ARE THE BEST ALGORITHMS SO FAR?

APPROXIMATE MESSAGE PASSING
AN ITERATIVE THRESHOLDING ALGORITHM

- **Approximate Message Passing**

$$\mathbf{B}^{(2,t)} = \frac{1}{\Delta_2 \sqrt{N}} Y \hat{\mathbf{x}}^t - \frac{1}{\Delta_2} \hat{\sigma}^t \hat{\mathbf{x}}^{t-1}$$

$$\mathbf{B}^{(p,t)} = \frac{\sqrt{(p-1)!}}{\Delta_p N^{(p-1)/2}} T(\hat{\mathbf{x}}^t)^{\otimes p-1} - \frac{p-1}{\Delta_p} \hat{\sigma}^t \left[\frac{(\hat{\mathbf{x}}^t) \cdot (\hat{\mathbf{x}}^{t-1})}{N} \right]^{p-2} \hat{\mathbf{x}}^{t-1}$$

$$A^{(2,t)} = \frac{\|\hat{\mathbf{x}}^t\|_2^2}{\Delta_2 N} \qquad A^{(p,t)} = \frac{(\|\hat{\mathbf{x}}^t\|_2^2)^{p-1}}{\Delta_p N}$$

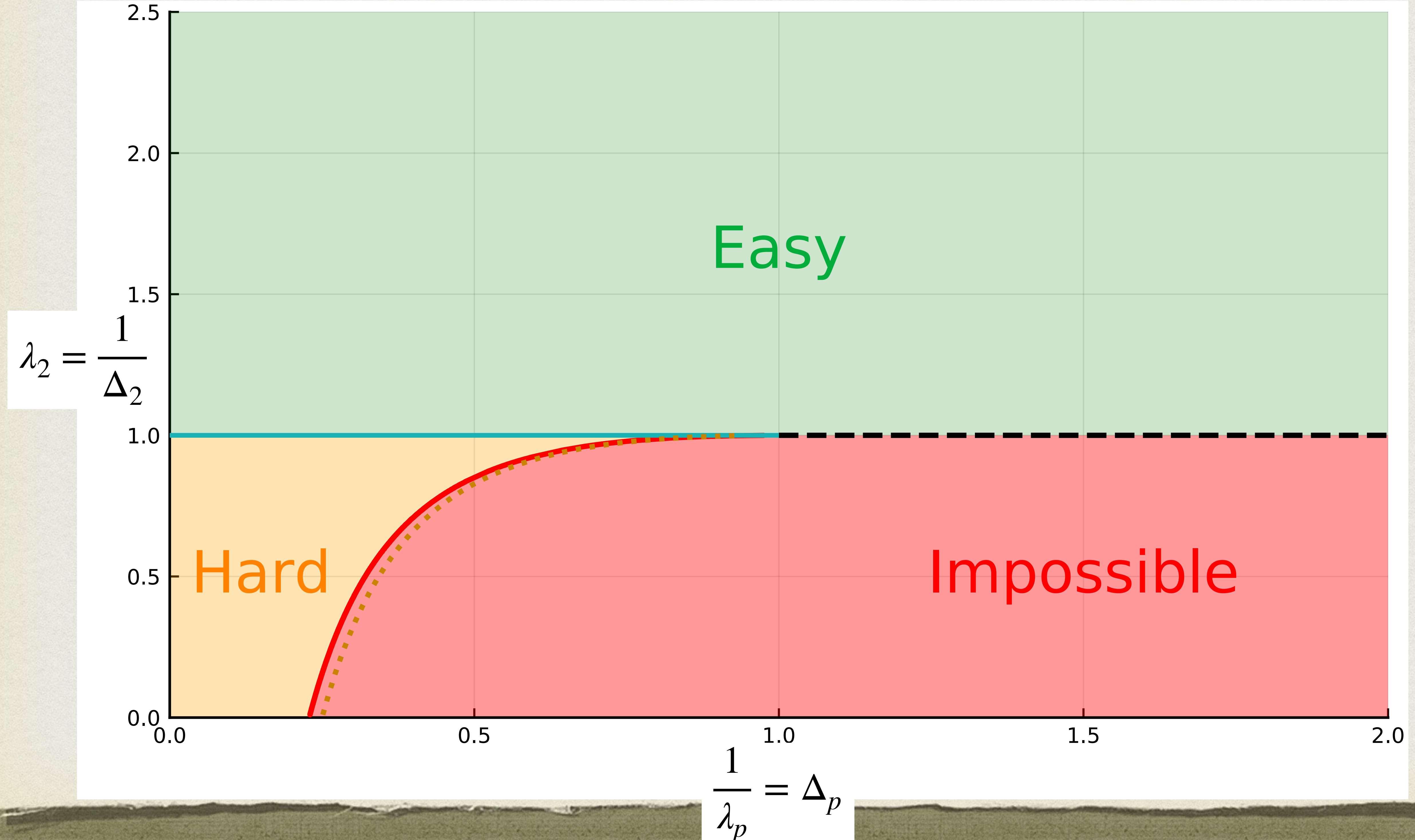
$$\hat{x}^{t+1} = \eta(A^{(2,t)} + A^{(p,t)}, \mathbf{B}^{(2,t)} + \mathbf{B}^{(p,t)}) \qquad \eta(A, B) = \frac{B}{1 + A}$$

$$\hat{\sigma}^{t+1} = \frac{1}{1 + A^{(2,t)} + A^{(p,t)}}$$

[Montanari, Richard ’15; Mannelli, Biroli, Cammarota, **FK**, Urbani, & Zdeborová ’18]

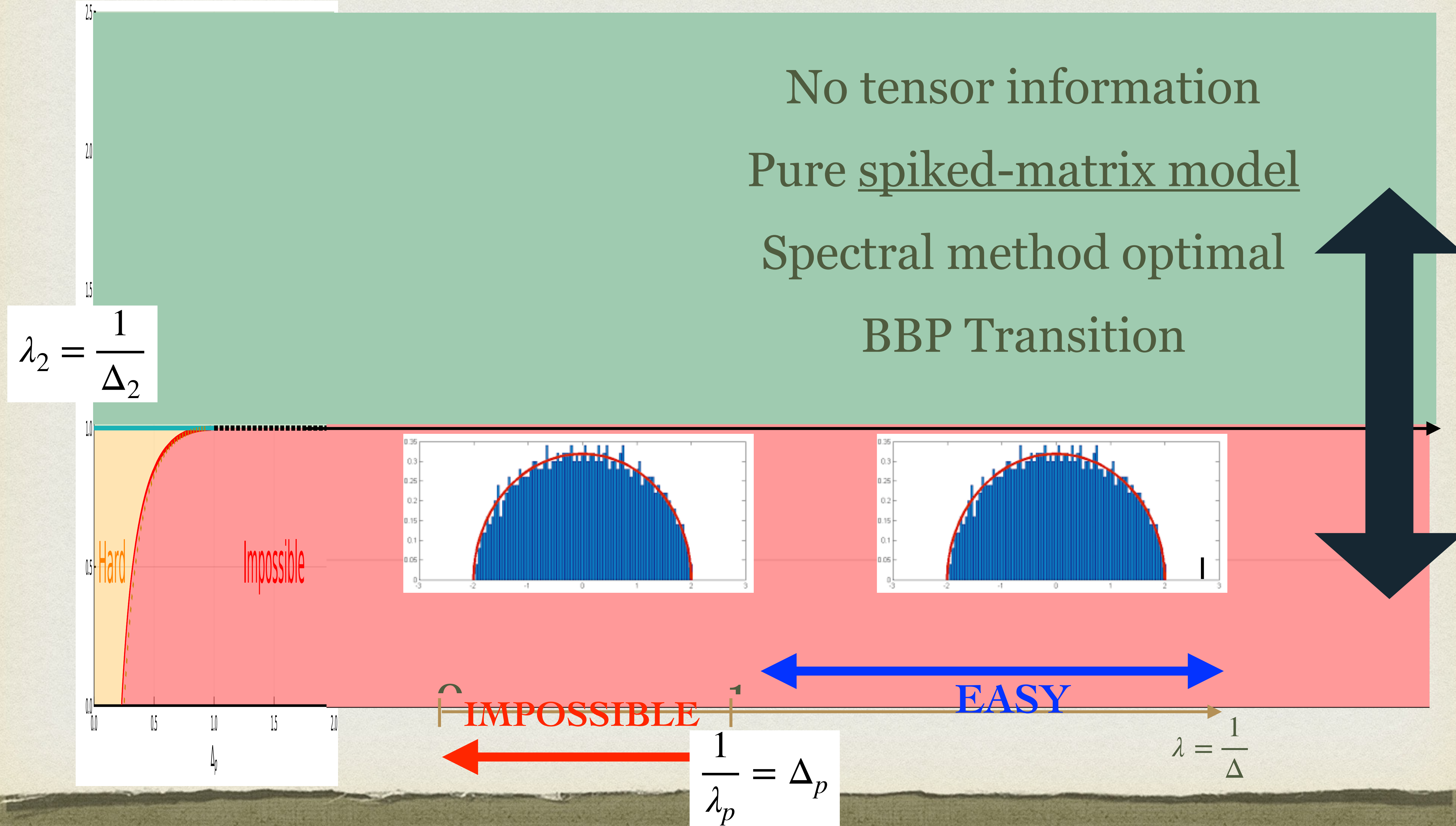
THE HARD PHASE OF AMP

Bayes-optimal performance & AMP



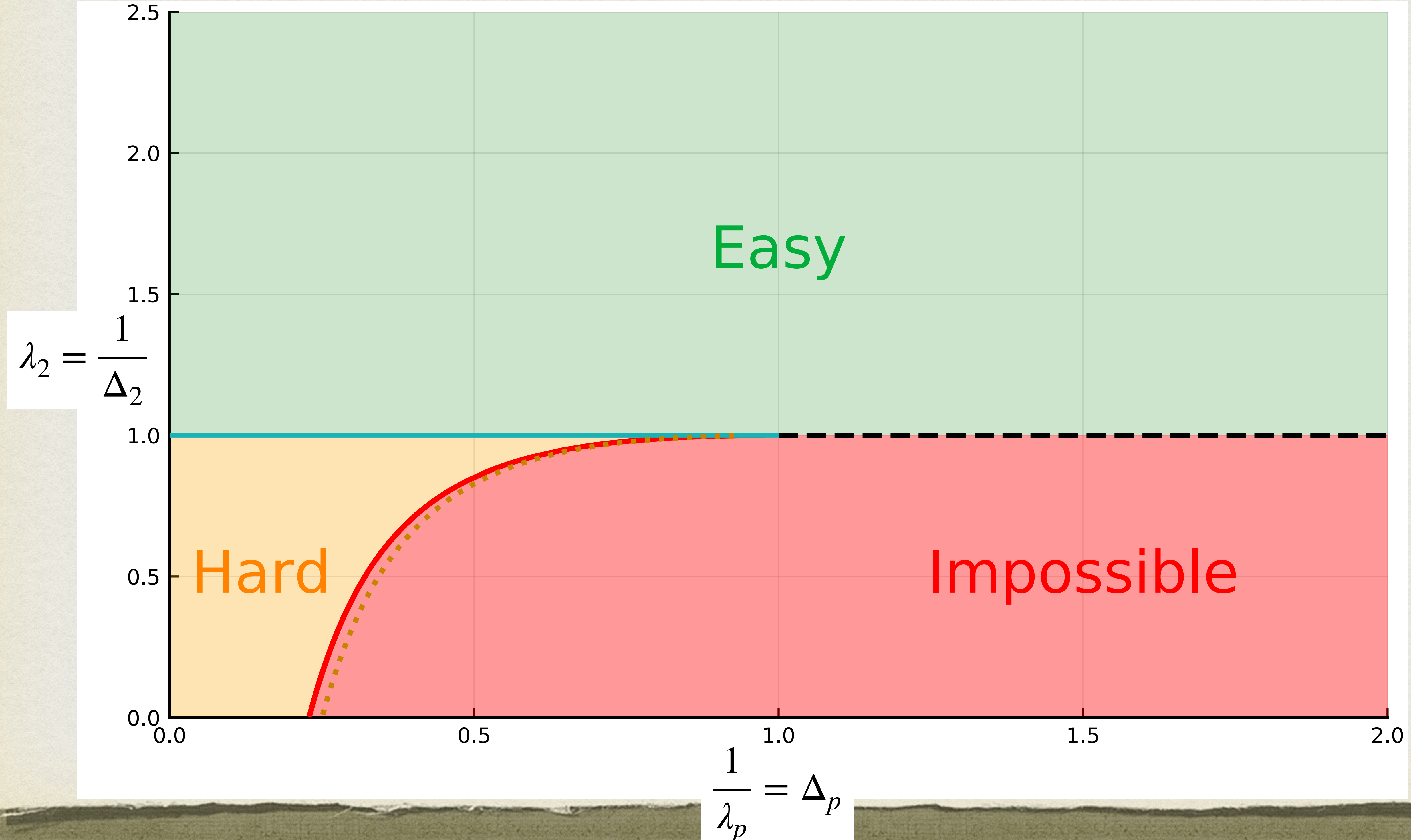
TWO LIMITS

Bayes-optimal performance & AMP



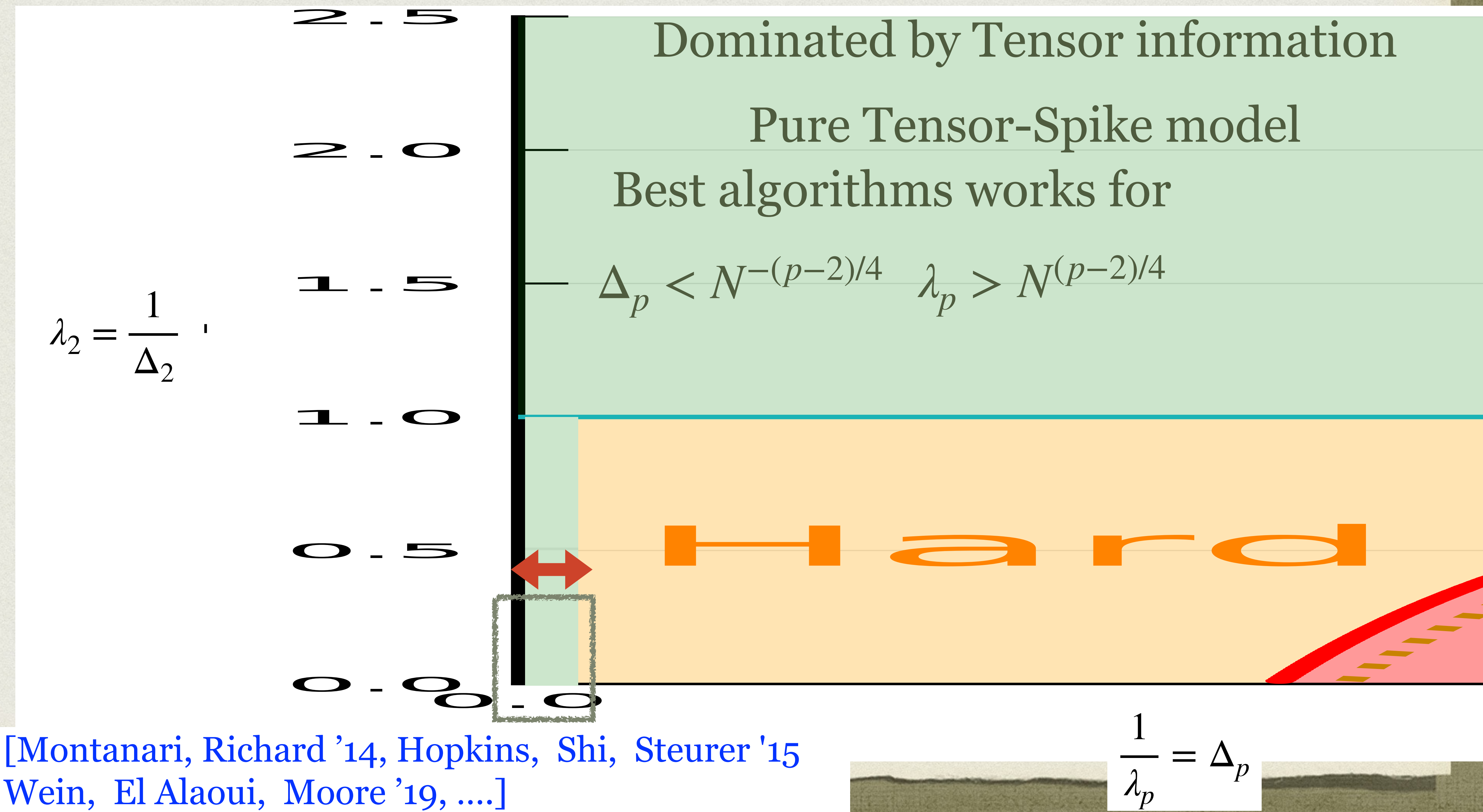
TWO LIMITS

Bayes-optimal performance & AMP



TWO LIMITS

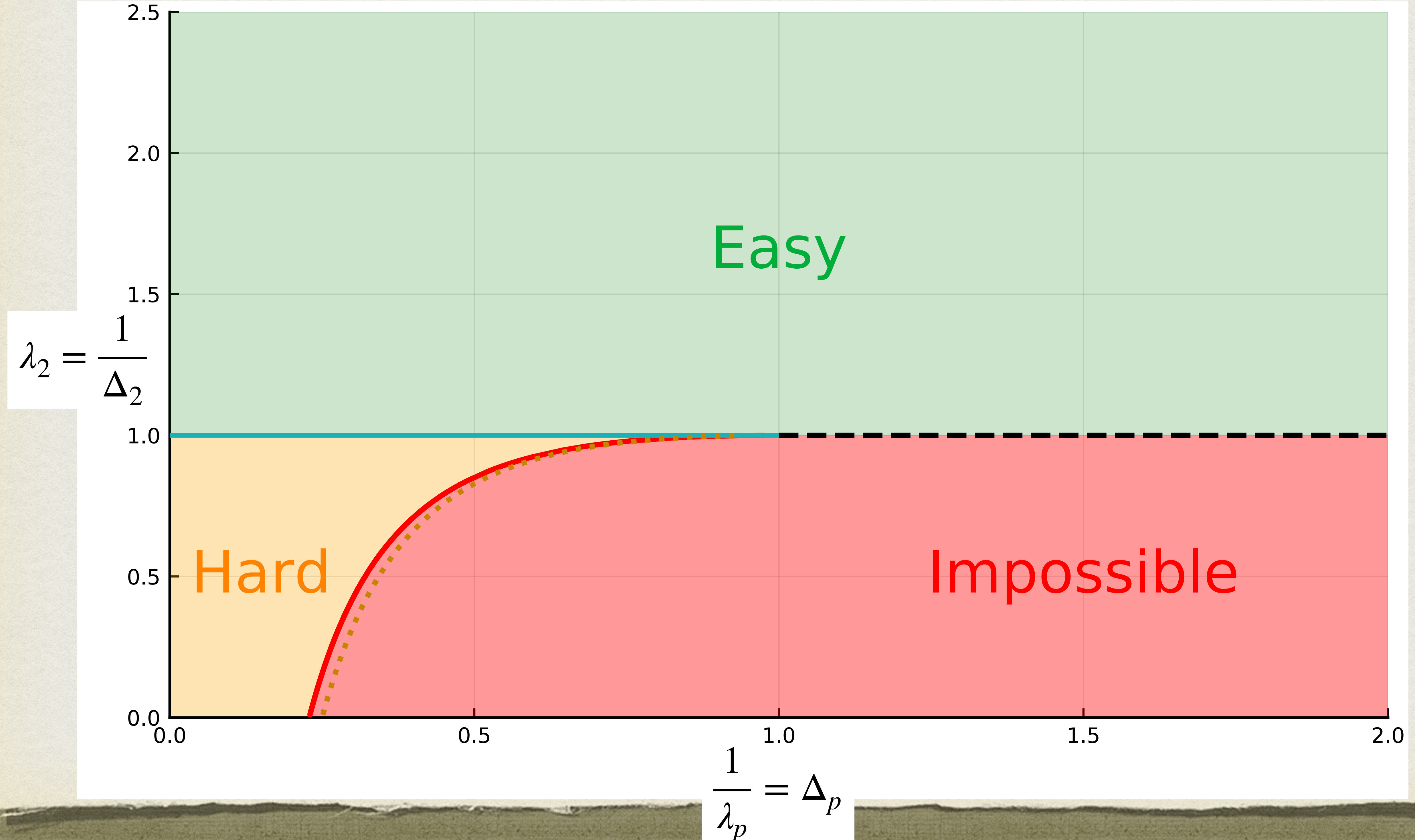
Bayes-optimal performance & AMP



[Montanari, Richard '14, Hopkins, Shi, Steurer '15
Wein, El Alaoui, Moore '19, ...]

THE HARD PHASE OF AMP

Bayes-optimal performance & AMP



WHAT ABOUT PRACTICAL OFF-THE-SHELVES ALGORITHMS?

Sampling the posterior with MCMC or Langevin ?

MLE with gradient descent?

Connection gradient descent with property of the landscapes?

Presence/Absence of spurious minima?

All these can be studied analytically & quantitively in the spherical spiked matrix-tensor model

②

Sampling the posterior with Langevin dynamics

LANGEVIN ALGORITHM

$$\langle \eta_i(t)\eta_j(t')\rangle = 2\delta_{ij}\delta(t-t')$$

spherical constraint

T=1 noise

$$\dot{x}_i(t) = -\mu(t)x_i(t) - \frac{\partial \mathcal{H}}{\partial x_i} + \eta_i(t)$$

gradient

At large time (exponential in N) samples the posterior measure.

Where does it go in large but constant time?

Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model

L. F. Cugliandolo and J. Kurchan
Phys. Rev. Lett. **71**, 173 – Published 5 July 1993

ITHM

Article

Pa

From

Zeitschrift für Physik B Condensed Matter
June 1993, Volume 92, Issue 2, pp 257–271 | Cite as

The spherical*p*-spin interaction spin-glass model

The dynamics

Authors | Authors and affiliations

A. Crisanti, H. Horner, H. -J. Sommers

CUGLIANDOLO-KURCHAN EQUATIONS FOR DYNAMICS OF SPIN-GLASSES.

GERARD BEN AROUS, AMIR DEMBO, AND ALICE GUIONNET

ABSTRACT. We study the Langevin dynamics for the family of spherical p-spin disordered mean-field models of statistical physics. We prove that in the limit of system size N approaching infinity, the empirical state correlation and integrated response functions for these N-dimensional coupled diffusions converge almost surely and uniformly in time, to the non-random unique strong solution of a pair of explicit non-linear integro-differential equations, first introduced by Cugliandolo and Kurchan.

Probability Theory and Related Fields **136**, 619–660 (2006)

Martin, Siggia, Rose '73, De Dominicis '76, Jensen '76

LANGEVIN STATE EVOLUTION

$$C_N(t,t') \equiv \frac{1}{N}\sum_{i=1}^{N} x_i(t)x_i(t')\,,$$

$$\overline{C}_N(t) \equiv \frac{1}{N}\sum_{i=1}^{N} x_i(t)x_i^*\,, \quad = m_{\text{Langevin}}(t)$$

$$R_N(t,t') \equiv \frac{1}{N}\sum_{i=1}^{N} \partial x_i(t)/\partial h_i(t')|_{h_i=0}\,,$$

$$\frac{\partial}{\partial t}C(t,t') = 2R(t',t) - \mu(t)C(t,t') + Q'(\overline{C}(t))\overline{C}(t') + \int_0^t dt'' R(t,t'')Q''(C(t,t''))C(t',t'') + \int_0^{t'} dt'' R(t',t'')Q'(C(t,t''))$$

$$\frac{\partial}{\partial t}R(t,t') = \delta(t-t') - \mu(t)R(t,t') + \int_{t'}^t dt'' R(t,t'')Q''(C(t,t''))R(t'',t')\,,$$

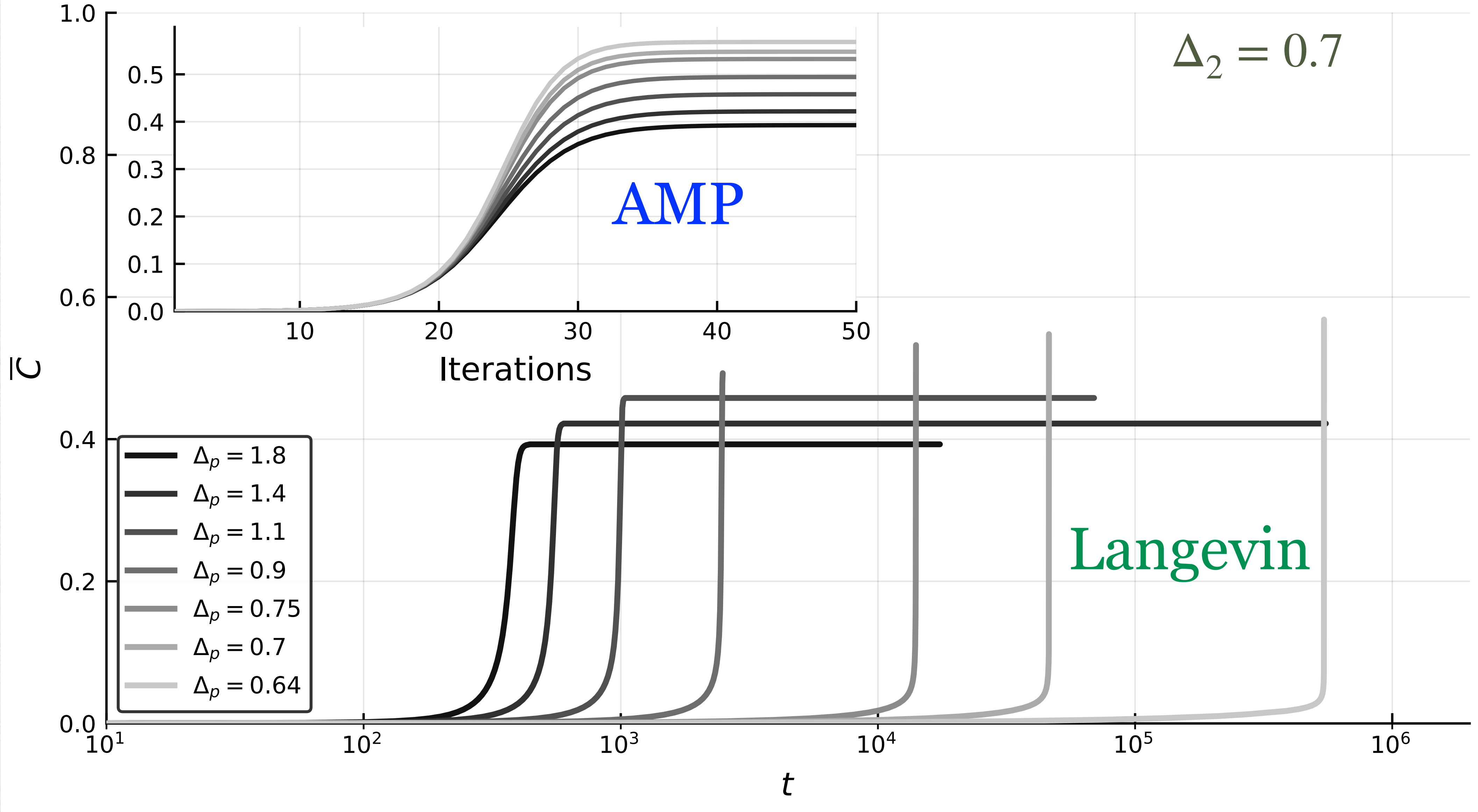
$$\frac{\partial}{\partial t}\overline{C}(t) = -\mu(t)\overline{C}(t) + Q'(\overline{C}(t)) + \int_0^t dt'' R(t,t'')\overline{C}(t'')Q(C(t,t''))\,, \qquad Q(x) = x^2/(2\Delta_2) + x^p/(p\Delta_p).$$

Generalization of the **CHSCK** equations that includes the spike **x***.

[Mannelli, Biroli, Cammarota, **FK**, Urbani, & Zdeborová '18]

LANGEVIN STATE EVOLUTION (NUMERICAL SOLUTION)

correlation with ground truth

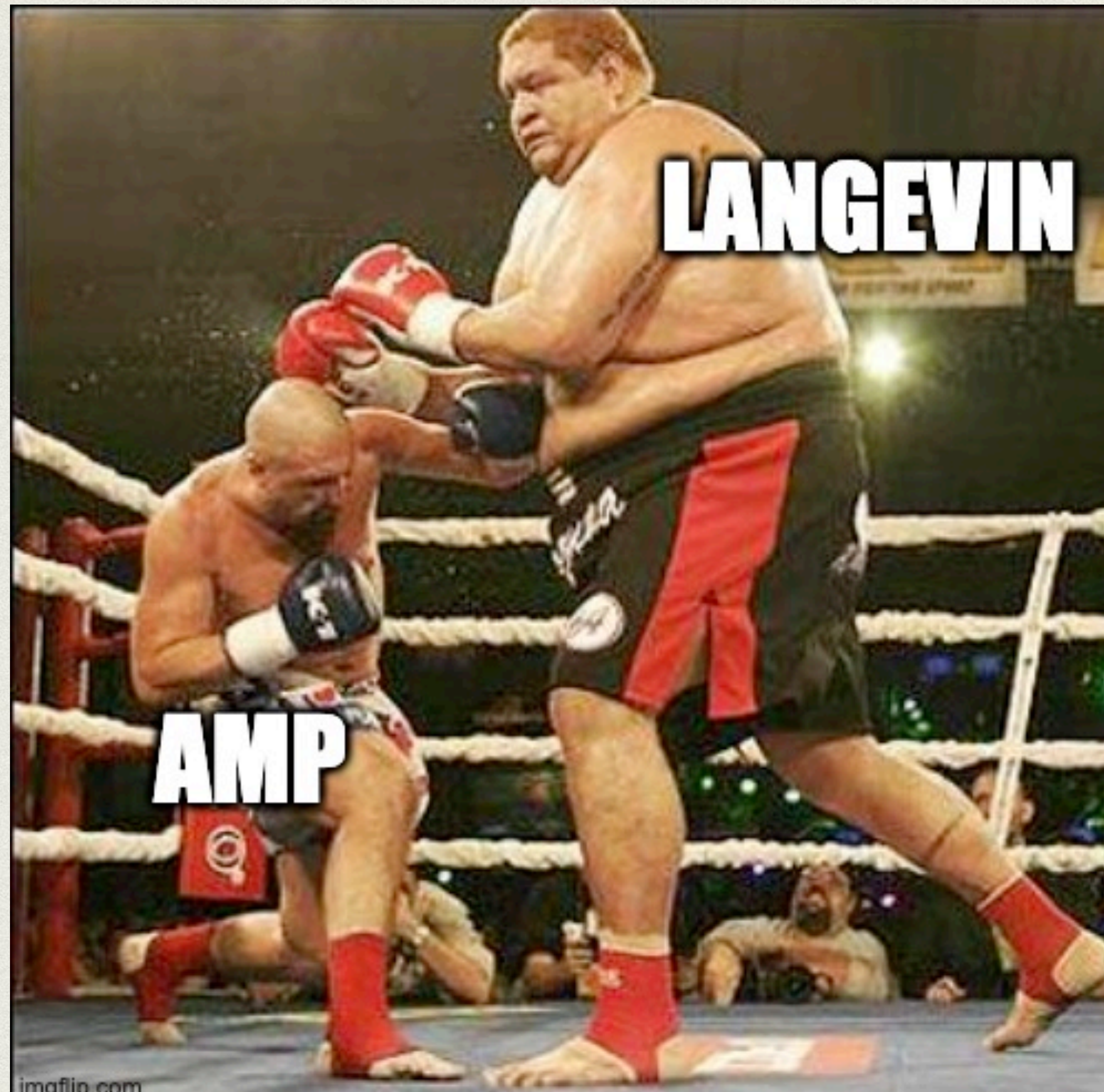


http://github.com/sphinxteam/spiked_matrix-tensor

AMP SAMPLNG
VS LANGEVIN SAMPLING

OK, Langevin is *slower*….

… but does it work as well as AMP in the long run (i.e linear but large time)?



LANGEVIN STATE EVOLUTION (NUMERICAL SOLUTION)

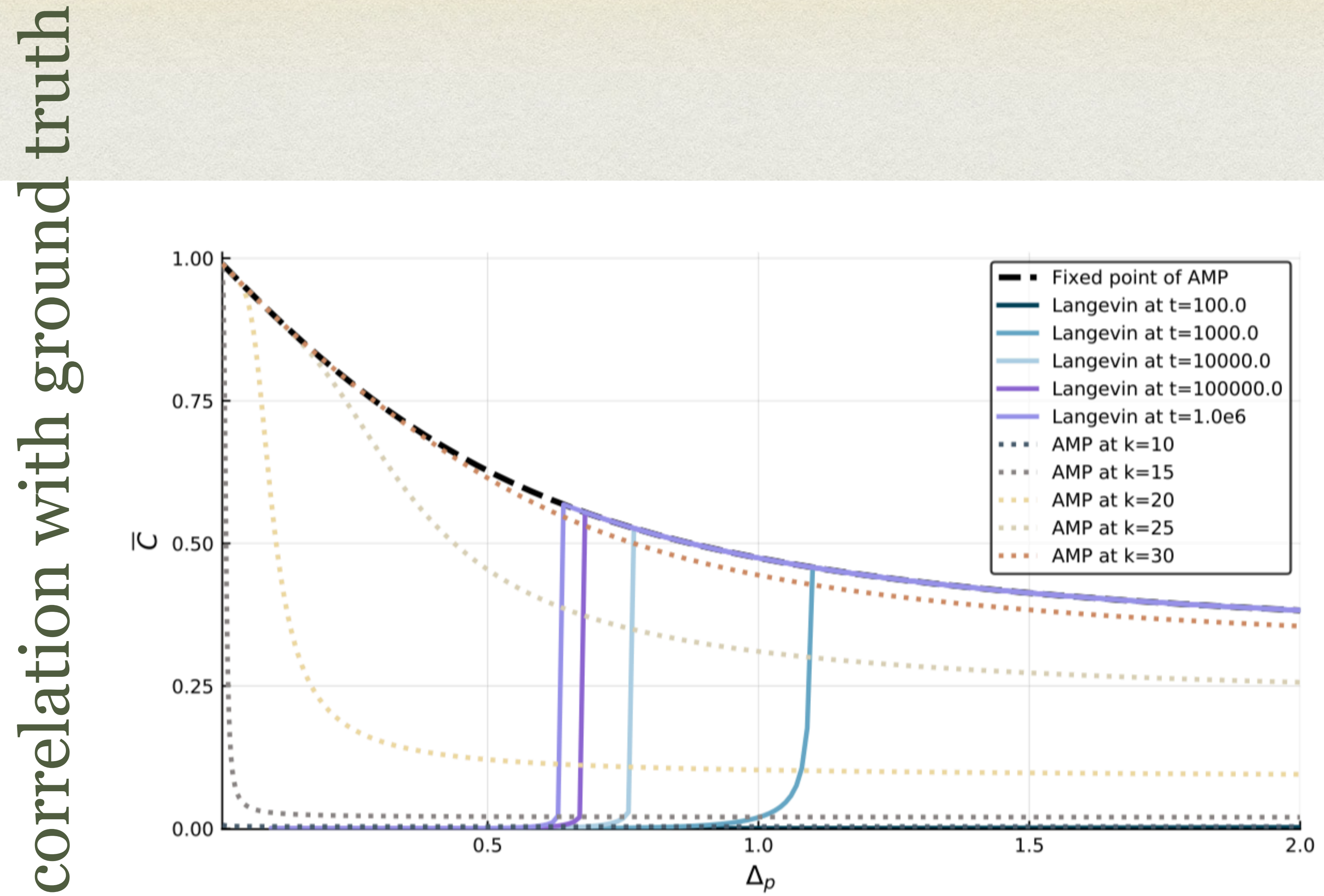
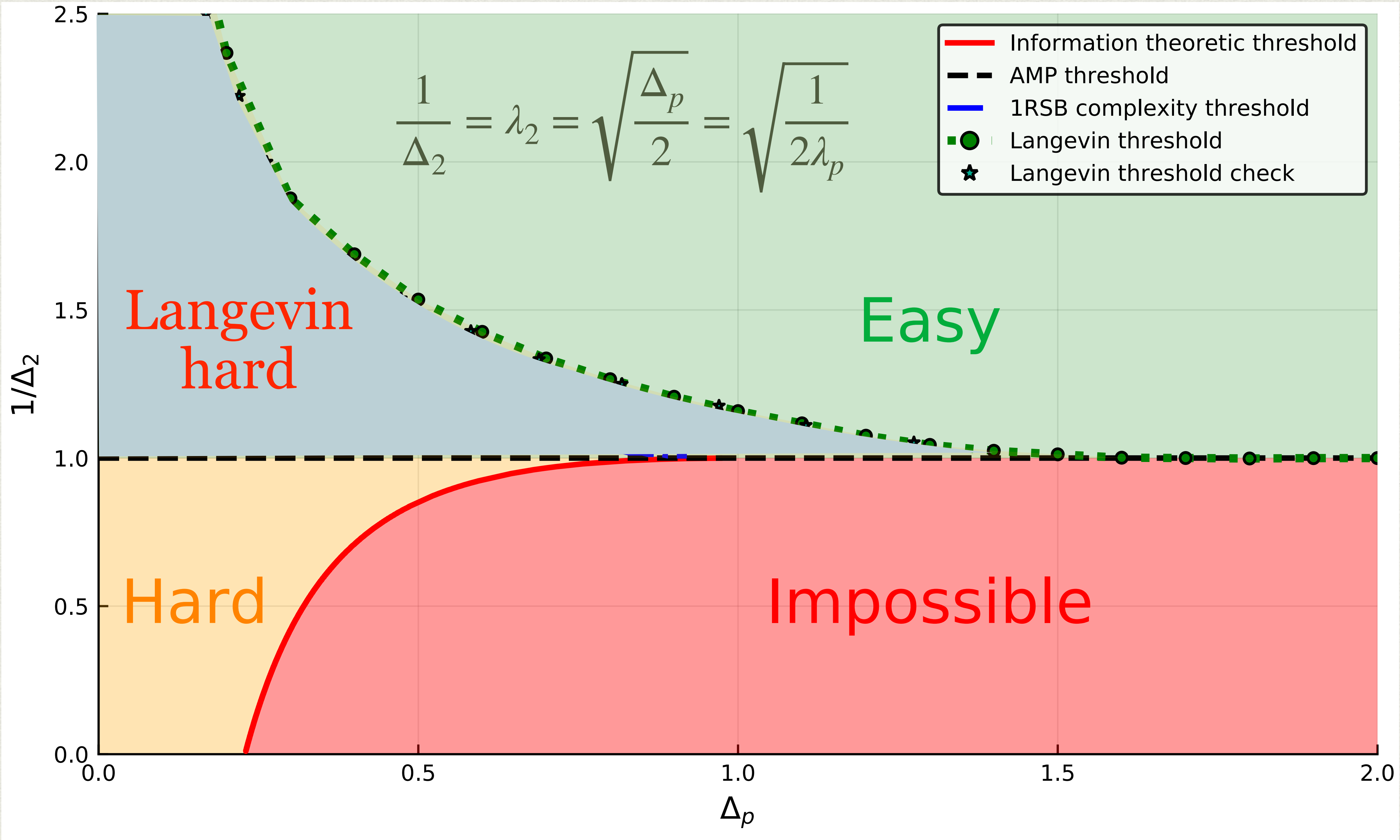


Figure 4: Correlation with the signal of AMP and Langevin at the kth iteration (at time t) for fixed $\Delta_2 = 0.7$.

http://github.com/sphinxteam/spiked_matrix-tensor

A LANGEVIN PHASE TRANSITION



AMP BEATS LANGEVIN

Langevin dynamics display **worst** performances w.r.t. Bayes-AMP

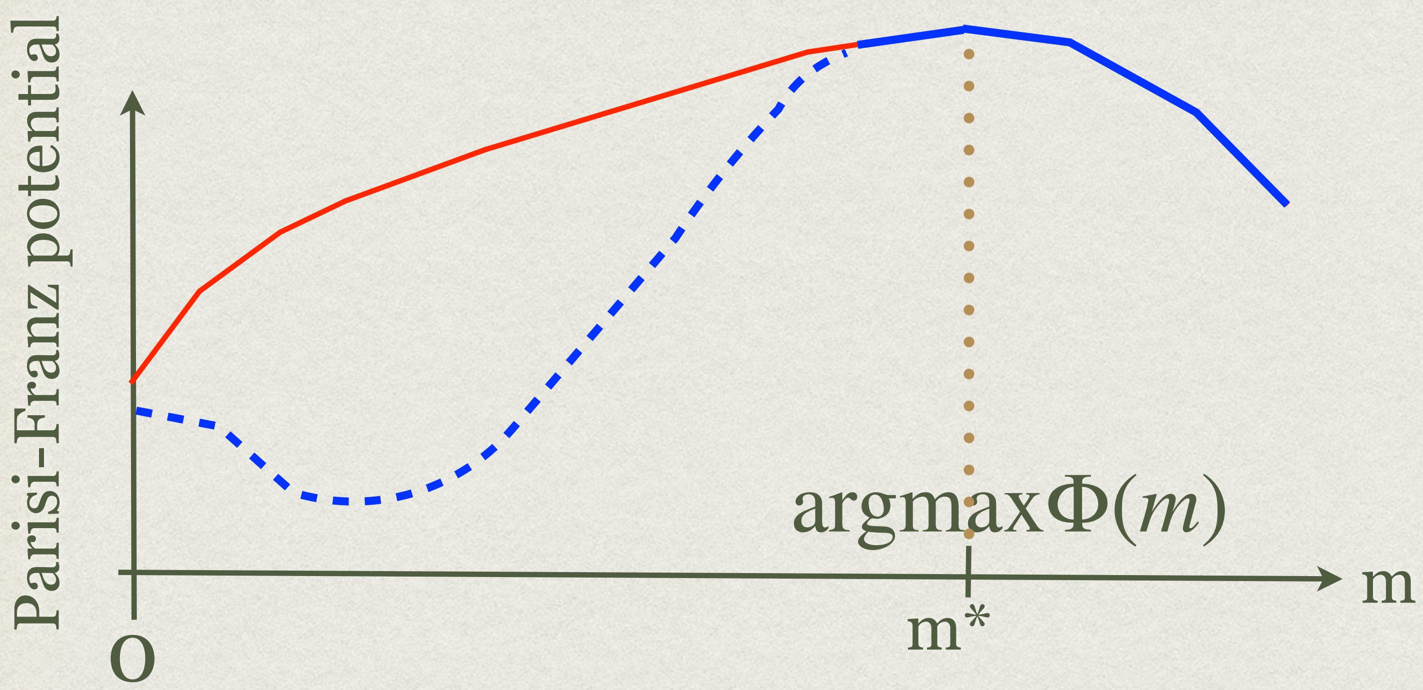
Physicists: "Residual glassiness prevents a correct sampling"

We expect the same picture to hold in all problems having hard phase associated to the first order phase transition.
(e.g. GLM, Teacher-Student Neural networks, ...)

A PARISI-FRANZ VISION

BEWARE: THIS SLIDE IS FOR REPLICA GEEKS

Consider the free energy of a system conditioned at a given overlap m from truth vector **x***



$$\Phi_{FP}(m) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_Y \log Z(Y, m)$$

$$\log Z(Y, m) = \int d\mathbf{x} e^{-\mathscr{H}} \mathbf{1}\left(\mathbf{m} - \frac{\mathbf{x}\cdot\mathbf{x}^*}{\mathbf{N}}\right)$$

AMP try to optimize the Replica-Symetric potentiel

Langevin try to optimize the Actual (RSB) potential

[Antenucci, Franz, Urbani, Zdeborova '19]

③

MLE, Gradient & Landscapes

MLE AND MINIMIZATION

- For the same signal **x*** observe a matrix Y <u>and</u> a tensor T as:

$$Y_{ij} = \frac{1}{\sqrt{N}} x_i^* x_j^* + \xi_{ij} \qquad \xi_{ij} \sim \mathcal{N}(0, \Delta_2)$$

$$T_{i_1 \dots i_p} = \frac{\sqrt{(p-1)!}}{N^{(p-1)/2}} x_{i_1}^* \dots x_{i_p}^* + \xi_{i_1 \dots i_p} \qquad \xi_{i_1, \dots, i_p} \sim \mathcal{N}(0, \Delta_p)$$

- Maximum likelihood (MLE)

$$\mathcal{L}(\mathbf{x}) = \frac{1}{2\Delta_2} \| Y - \frac{\mathbf{x}\mathbf{x}^T}{\sqrt{N}} \|_2^2 + \frac{1}{2\Delta_p} \| T - \sqrt{(p-1)!} \frac{\mathbf{x}^{\otimes p}}{N^{(p-1)/2}} \|_2^2$$

Minimize $\mathcal{L}(\mathbf{x})$ subject to $\|\mathbf{x}\|_2^2 = N$

GRADIENT FLOW

ZERO TEMPERATURE LIMIT OF LANGEVIN

$$\dot{x}_i(t) = -\mu(t)x_i(t) - \frac{\partial \mathcal{H}}{\partial x_i} + \eta_i(t)$$

$$\dot{x}_i(t) = -\mu(t)x_i(t) - \frac{\partial \mathcal{H}}{\partial x_i}$$

Can be analysed again with the Langevin State evolution

Simply the T → 0 limit of the CHSCK equations

LANGEVIN STATE EVOLUTION

ZERO TEMPERATURE LIMIT: GRADIENT FLOW

$$C_N(t,t') \equiv \frac{1}{N}\sum_{i=1}^{N} x_i(t)x_i(t')\,,$$

$$\overline{C}_N(t) \equiv \frac{1}{N}\sum_{i=1}^{N} x_i(t)x_i^*\,, = m_{\mathrm{GD}}(t)$$

$$R_N(t,t') \equiv \frac{1}{N}\sum_{i=1}^{N} \partial x_i(t)/\partial h_i(t')|_{h_i=0}\,,$$

$$\frac{\partial}{\partial t}C(t,t') = -\tilde{\mu}(t)C(t,t') + Q'(m(t))m(t') + \int_0^t dt'' R(t,t'')Q''(C(t,t''))C(t',t'') + \int_0^{t'} dt'' R(t',t'')Q'(C(t,t''))\,,$$

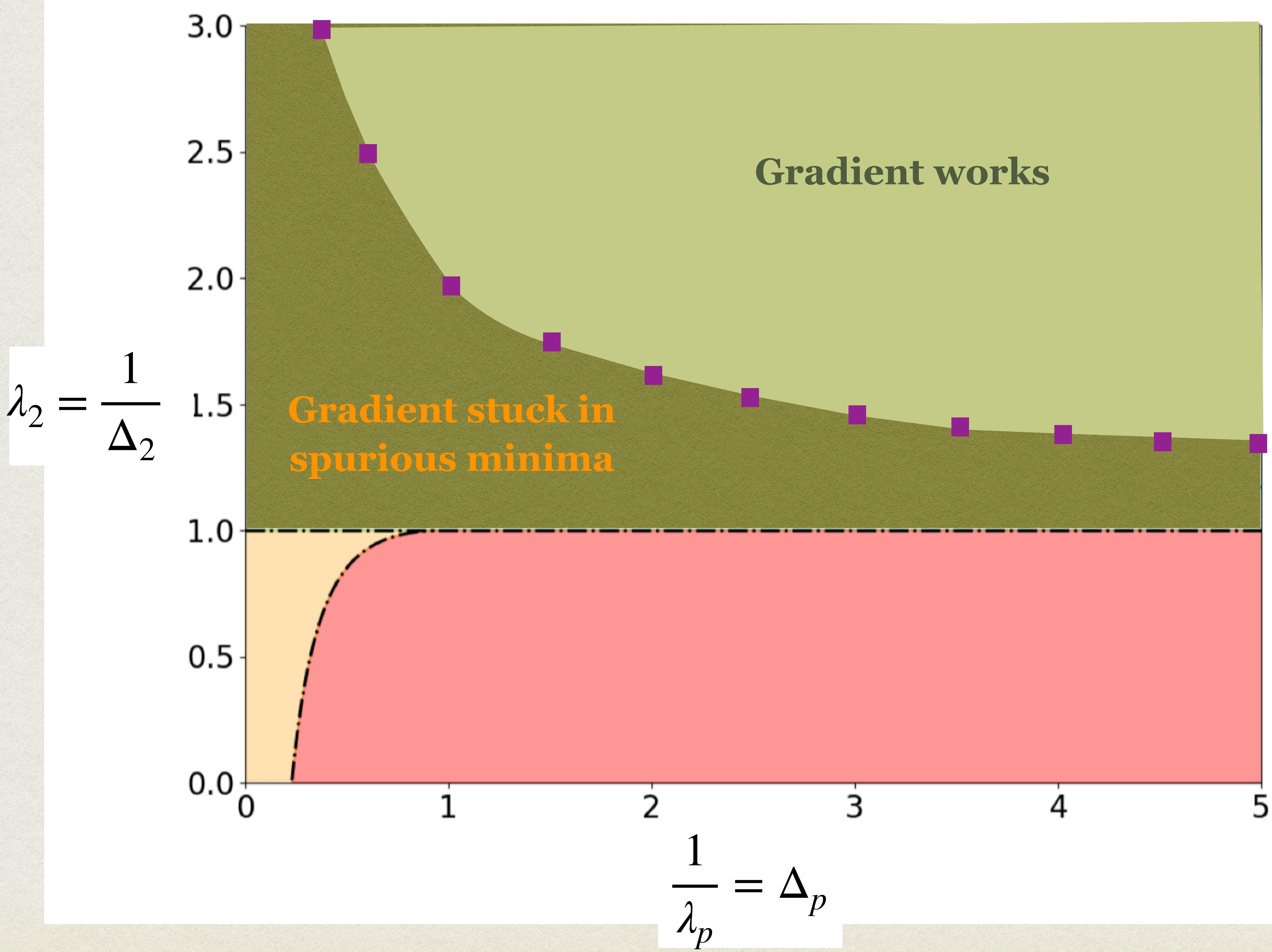
$$\frac{\partial}{\partial t}R(t,t') = -\tilde{\mu}(t)R(t,t') + \int_{t'}^t dt'' R(t,t'')Q''(C(t,t''))R(t'',t')\,,$$

$$\frac{\partial}{\partial t}m(t) = -\tilde{\mu}(t)m(t) + Q'(m(t)) + \int_0^t dt'' R(t,t'')m(t'')Q(C(t,t''))\,,$$

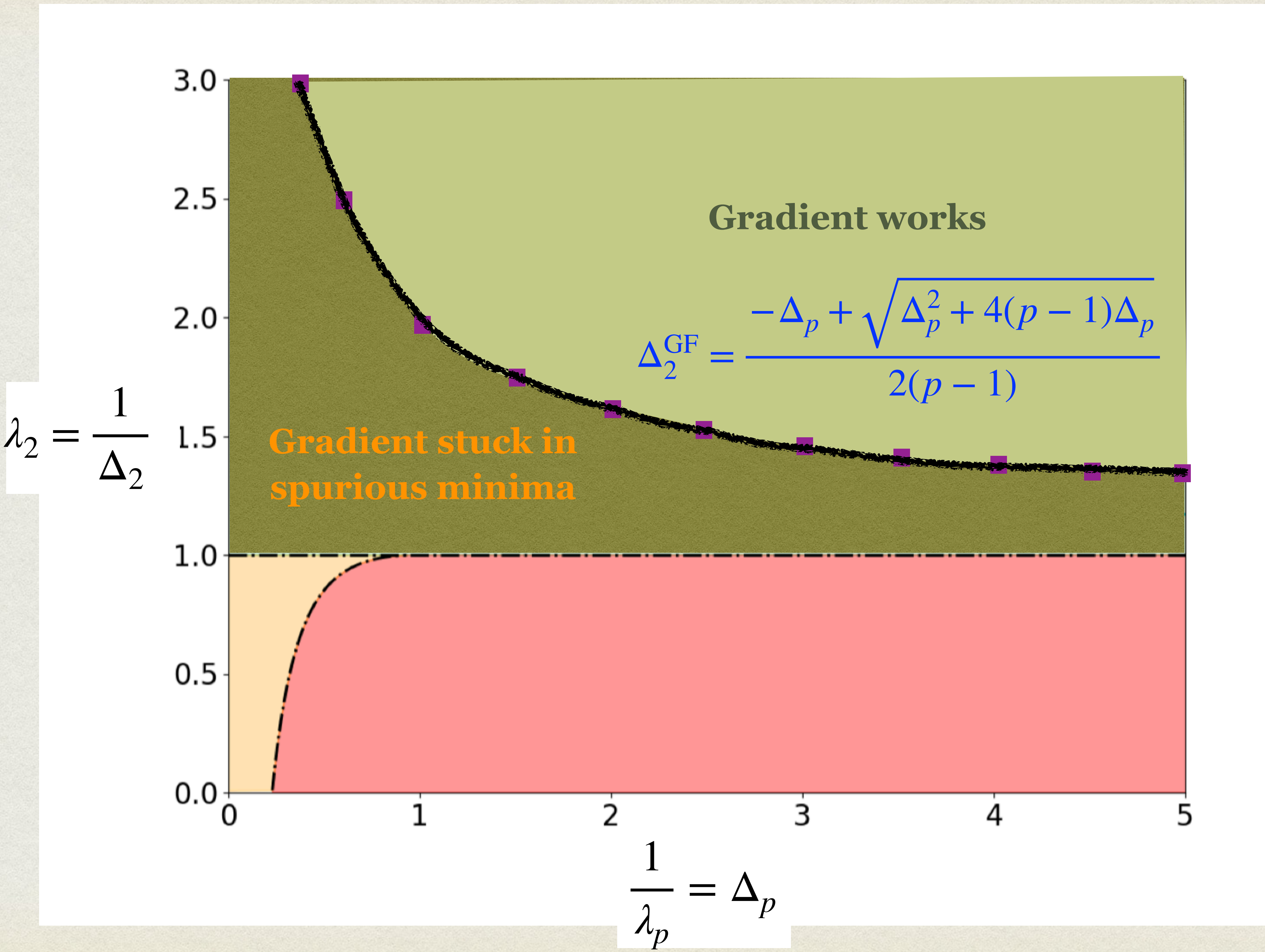
$$Q(x) = x^2/(2\Delta_2) + x^p/(p\Delta_p).$$

[Mannelli, Biroli, Cammarota, FK, Urbani, & Zdeborová ’18]

NUMERICAL INTEGRATION



ANALYTICAL SOLUTION
(LONG TIME EXTRAPOLATION)



ENERGY LANDSCAPE

- Maximum likelihood (MLE)

$$\mathscr{L}(\mathbf{x}) = \frac{1}{2\Delta_2}\|Y - \frac{\mathbf{x}\mathbf{x}^T}{\sqrt{N}}\|_F^2 + \frac{1}{p!\Delta_p}\|T - \sqrt{(p-1)!}\frac{\mathbf{x}^{\otimes p}}{N^{(p-1)/2}}\|_F^2$$

Minimize $\mathscr{L}(\mathbf{x})$ subject to $\|\mathbf{x}\|_2^2 = N$

- Can we compute the property of the energy landscape ?
- Number of minimas/saddles at each energy level?
- Are the minima spurious or good ones?

LANDSCAPE & MINIMAS
THE KAC-RICE FORMULA

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[\mathcal{N}(m, \epsilon_2, \epsilon_p)\right] = \tilde{\Sigma}_{\Delta_2,\Delta_p}(m, \epsilon_2, \epsilon_p)$$

$$\mathcal{N}(m, \epsilon_2, \epsilon_p; \Delta_2, \Delta_p) = e^{\tilde{\Sigma}_{\Delta_2,\Delta_p}(m,\epsilon_2,\epsilon_p)} =$$

$$= \int_{\mathbb{S}^{N-1}} \mathbb{E}[\det H | G = 0, F_2 = N\epsilon_2, F_p = N\epsilon_p, H \succ 0]\, \phi_{G,F_2,F_p}(\sigma, 0, \epsilon_2, \epsilon_p)\, \delta(m - \sigma\cdot\sigma^*)\, \mathrm{d}\sigma$$

[Fyodorov Y. V. ’03; Auffinger A., Ben Arous G., & Cerny J ’13]

LANDSCAPE & MINIMAS

KAC-RICE FOR THE SPIKE MODEL

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[\mathcal{N}(m, \epsilon_2, \epsilon_p)\right] = \tilde{\Sigma}_{\Delta_2,\Delta_p}(m, \epsilon_2, \epsilon_p)$$

$$\tilde{\Sigma}_{\Delta_2,\Delta_p}(m,\epsilon_2,\epsilon_p) = \frac{1}{2}\log\frac{\frac{p-1}{\Delta_p}+\frac{1}{\Delta_2}}{\frac{1}{\Delta_p}+\frac{1}{\Delta_2}} + \frac{1}{2}\log(1-m^2) - \frac{1}{2}\frac{\left(\frac{m^{p-1}}{\Delta_p}+\frac{m}{\Delta_2}\right)^2}{\frac{1}{\Delta_p}+\frac{1}{\Delta_2}}(1-m^2) - 2p\Delta_p\left(\epsilon_p - \frac{m^p}{2p\Delta_p}\right)^2 - 4\Delta_2\left(\epsilon_2 - \frac{m^2}{4\Delta_2}\right)^2 + \Phi(t) - L(\theta,t),$$

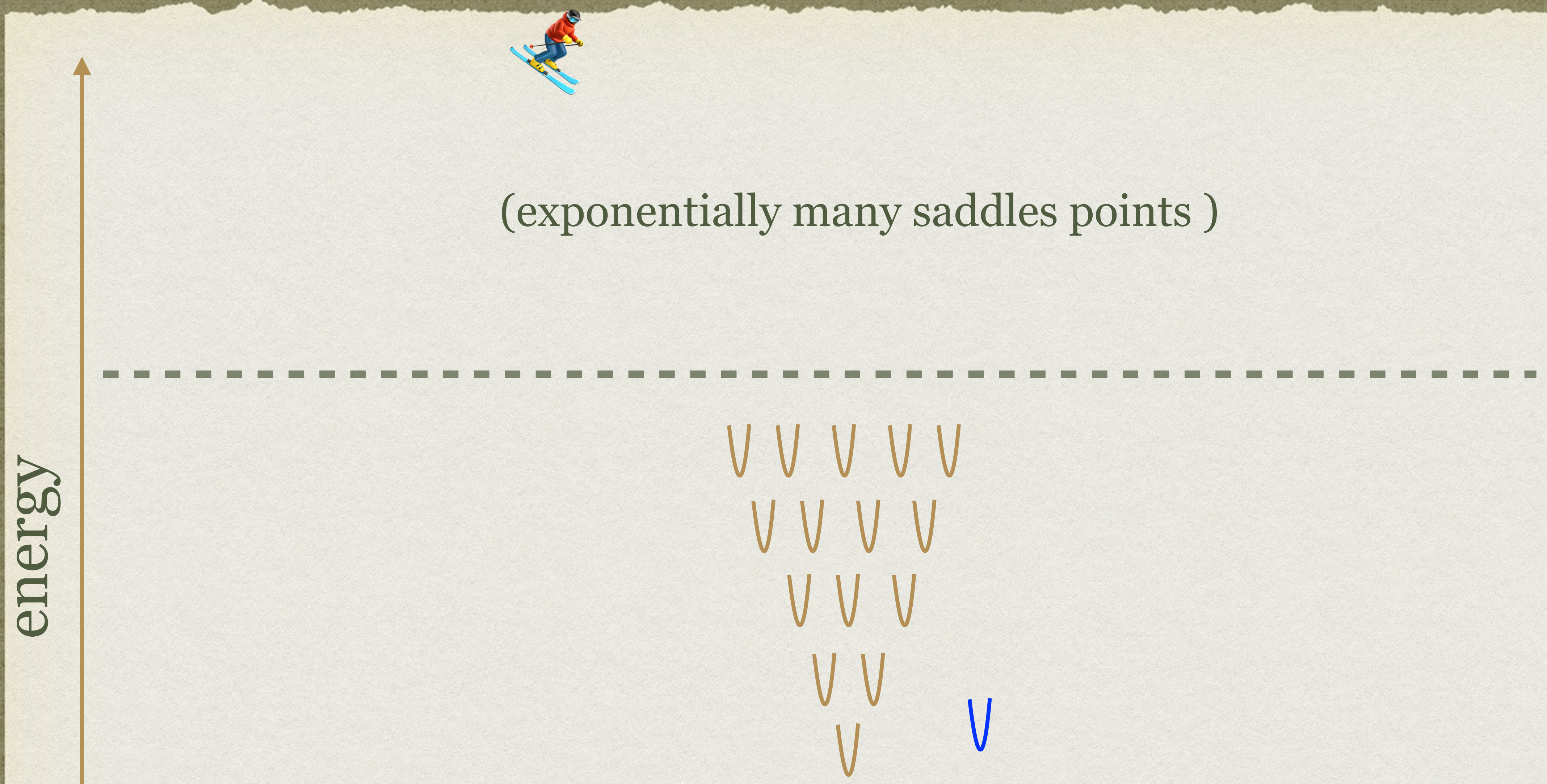
$$\Phi(t) = \frac{t^2}{4} + \mathbb{1}_{|t|>2}\left[\log\left(\sqrt{\frac{t^2}{4}-1}+\frac{|t|}{2}\right) - \frac{|t|}{4}\sqrt{t^2-4}\right]$$

$$L(\theta,t) = \begin{cases} \frac{1}{4}\int_{\theta+\frac{1}{\theta}}^{t}\sqrt{y^2-4}dy - \frac{\theta}{2}\left(t-\left(\theta+\frac{1}{\theta}\right)\right) + \frac{t^2-\left(\theta+\frac{1}{\theta}\right)^2}{8} & \theta>1,\ 2\le t<\frac{\theta^2+1}{\theta} \\ \infty & t<2 \\ 0 & \text{otherwise.} \end{cases}$$

(Note: we also checked that annealed is equal to quenched, thanks to the replica method)

[Ben Arous, Mei, Montanari, & Nica '17, Sarao, **FK**, Urbani & Zdeborova '19]

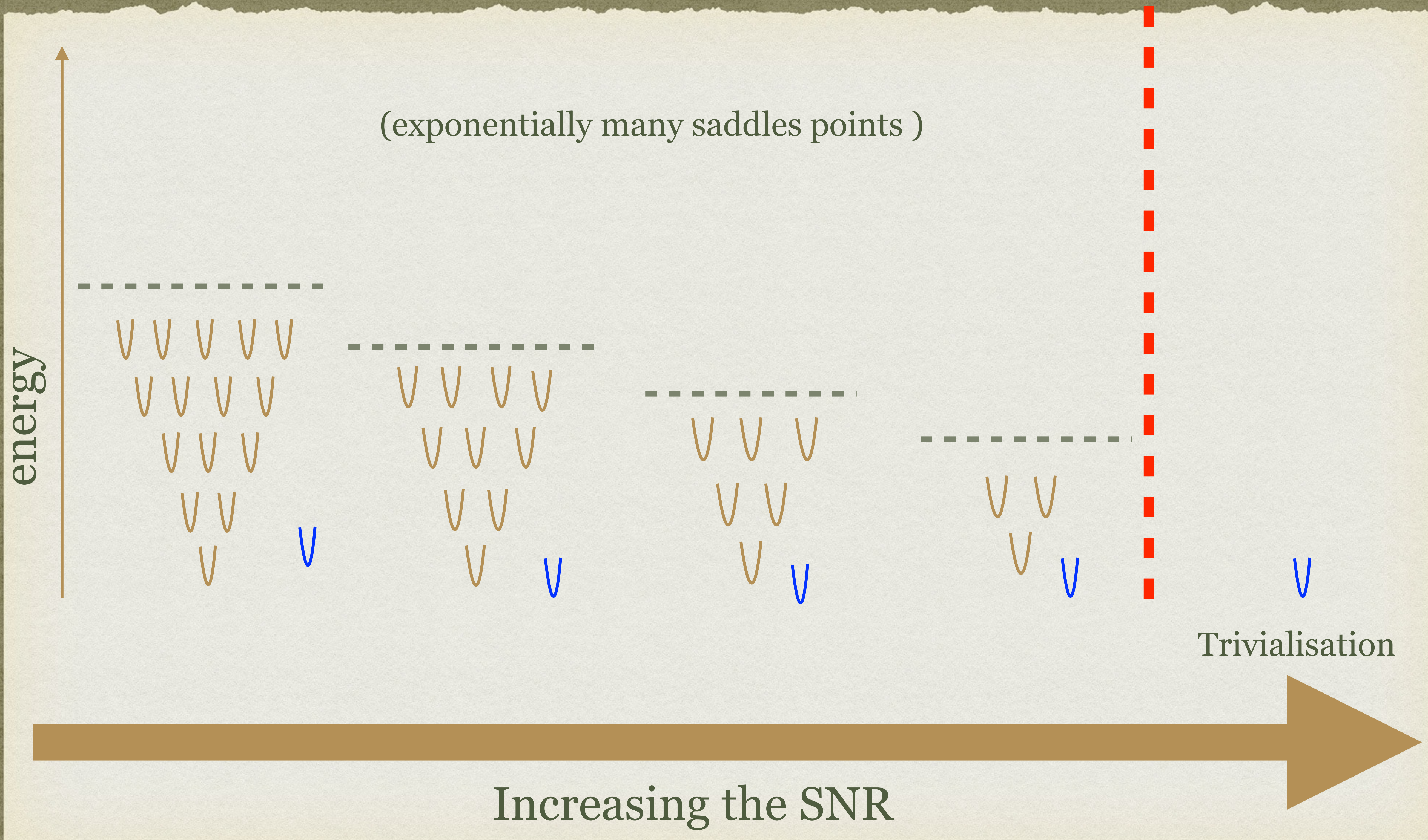
LANDSCAPE ANALYSIS



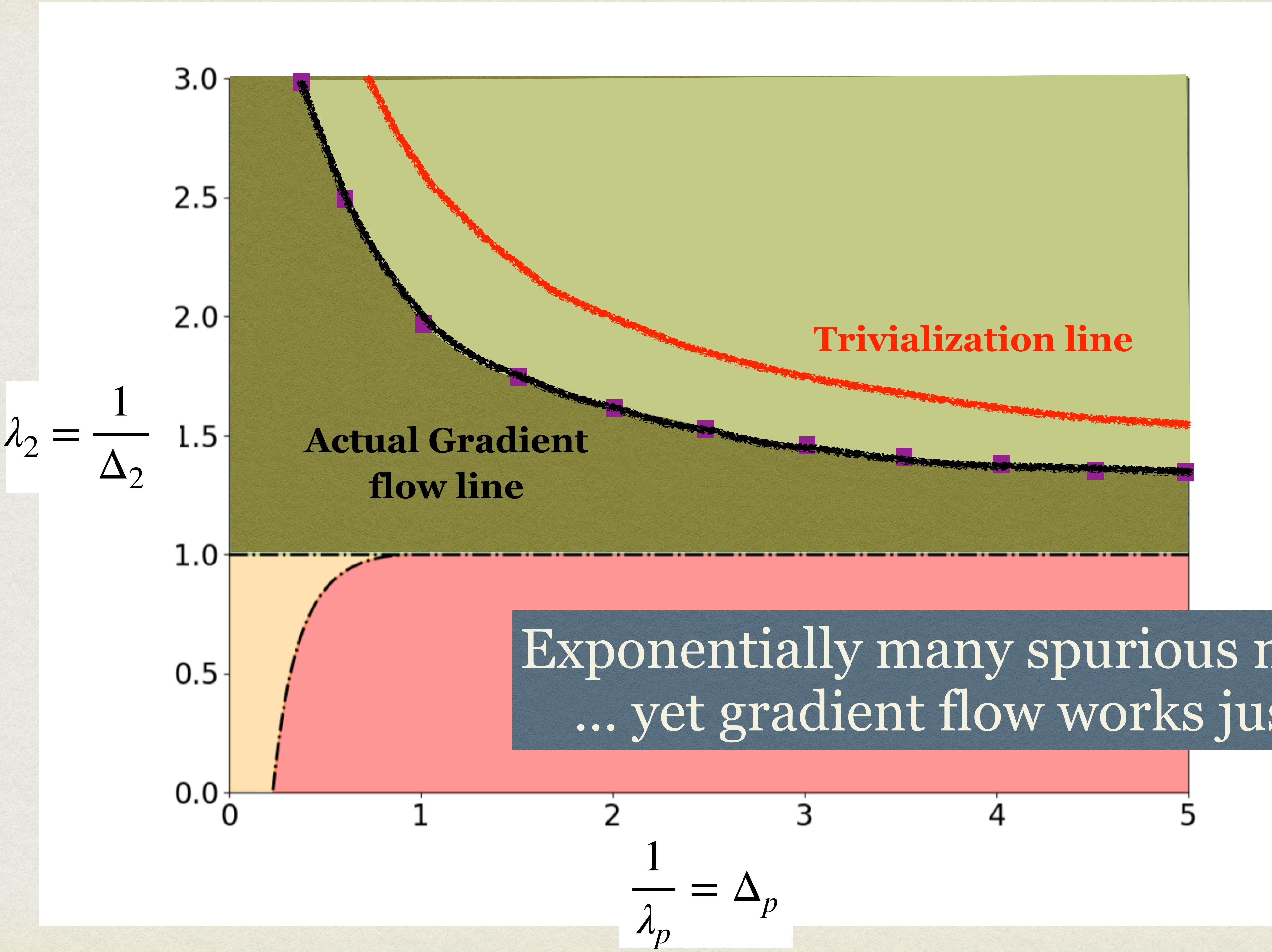
Low SNR / large noise situation

Similar as in Levent, Guney, Ben Arous & LeCun '14

LANDSCAPE ANALYSIS

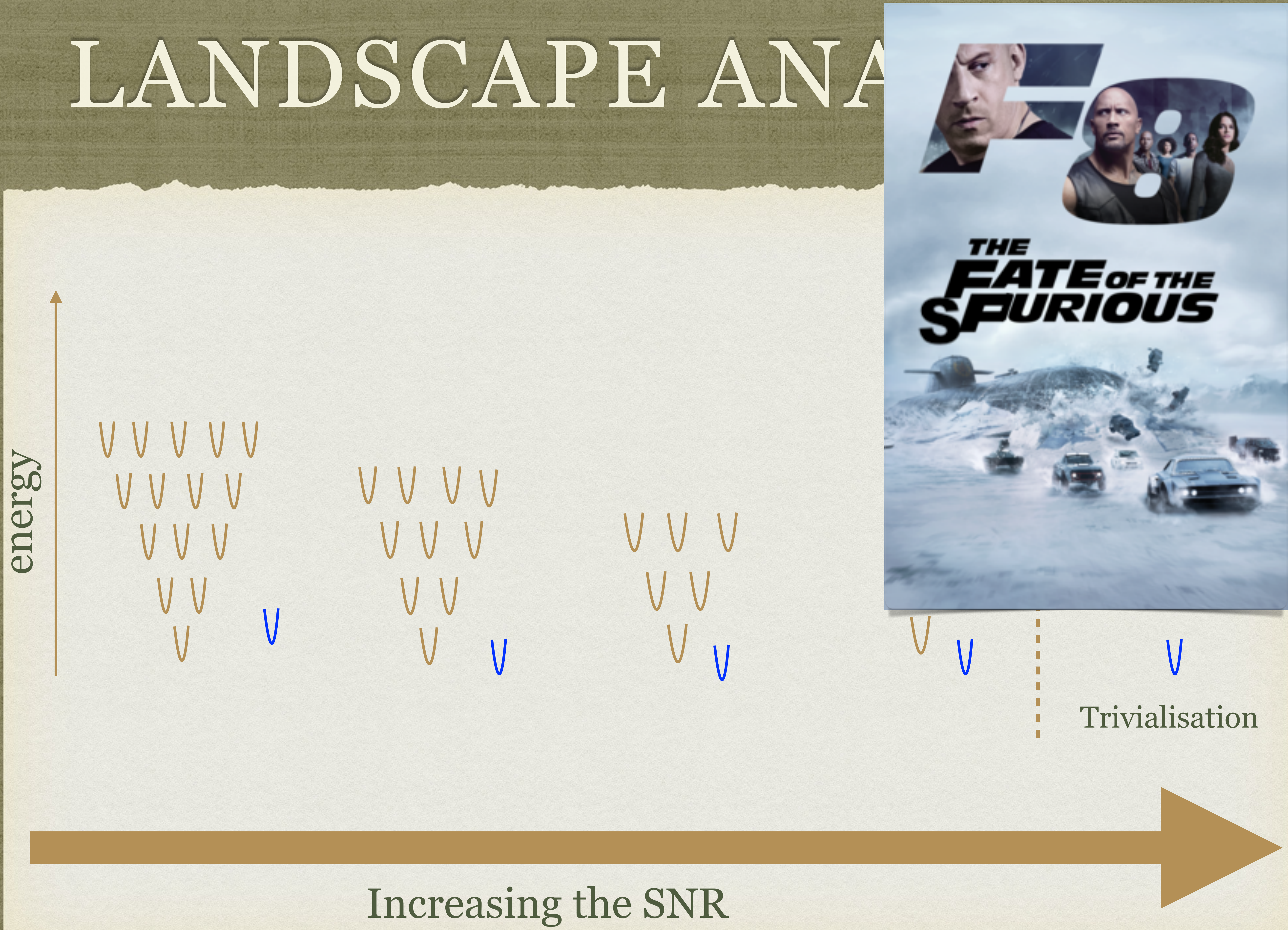


WHAT IS ACTUALLY GOING ON

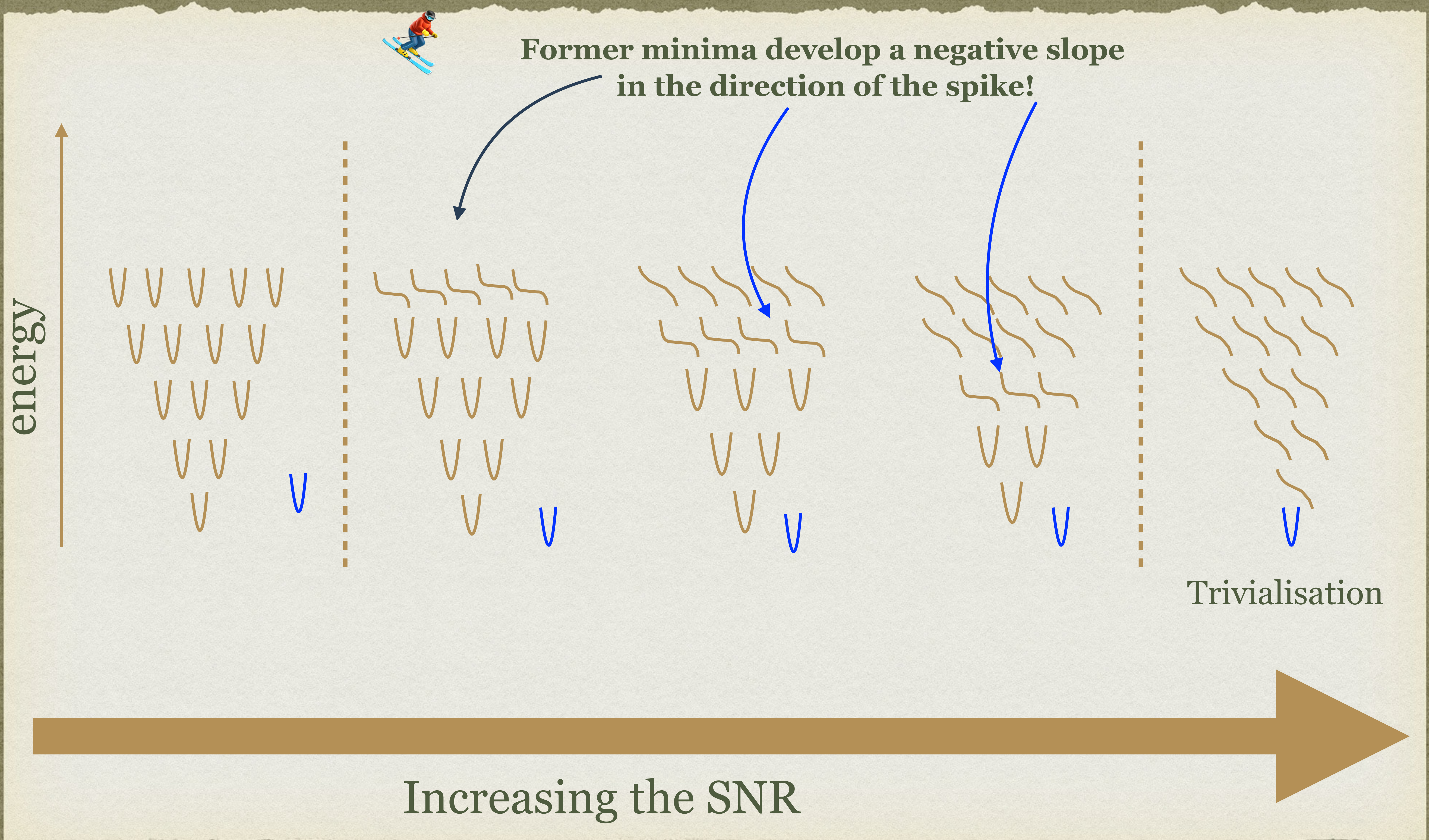


Exponentially many spurious minima…
… yet gradient flow works just fine!

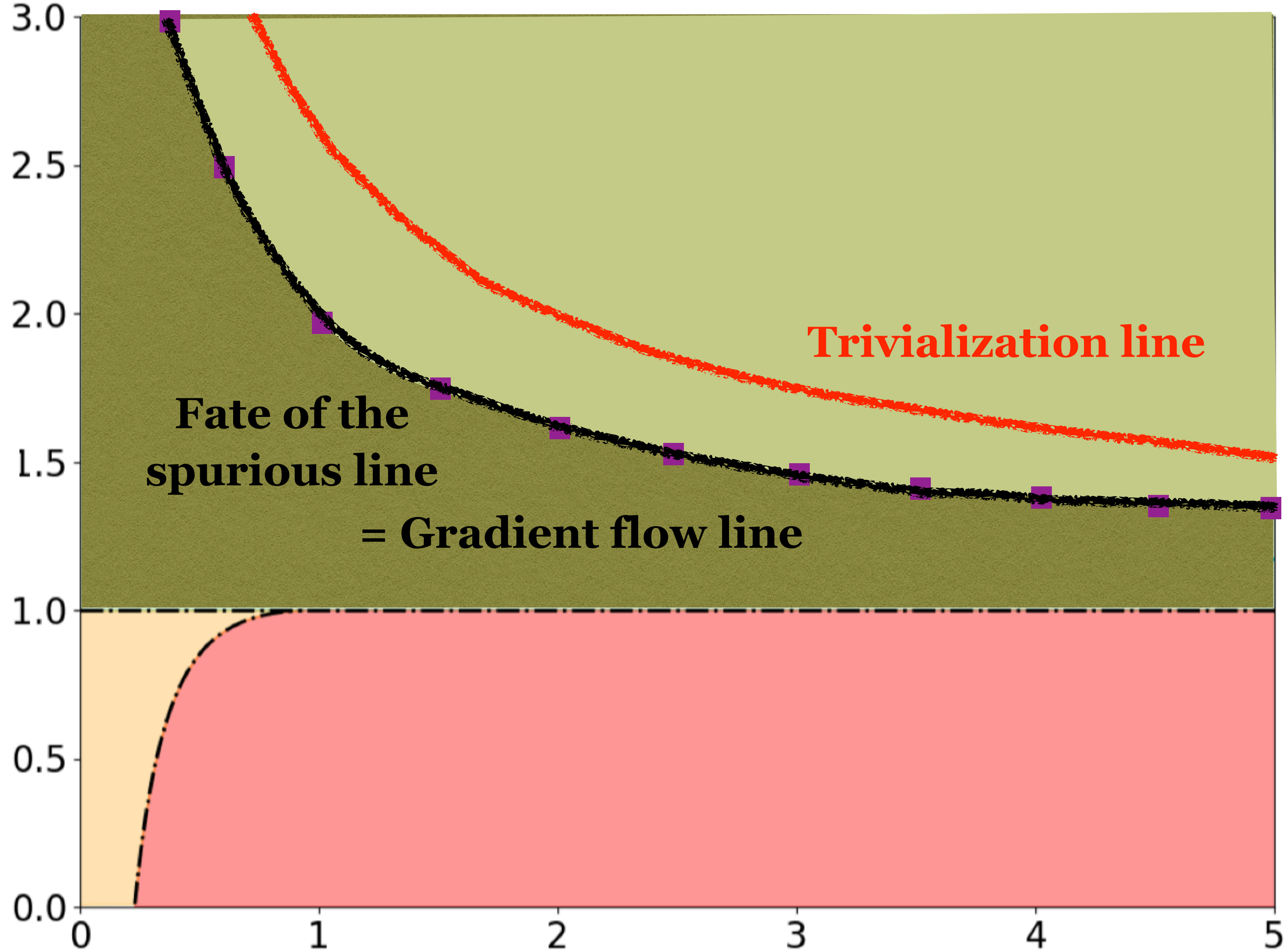
LANDSCAPE ANA



LANDSCAPE ANALYSIS



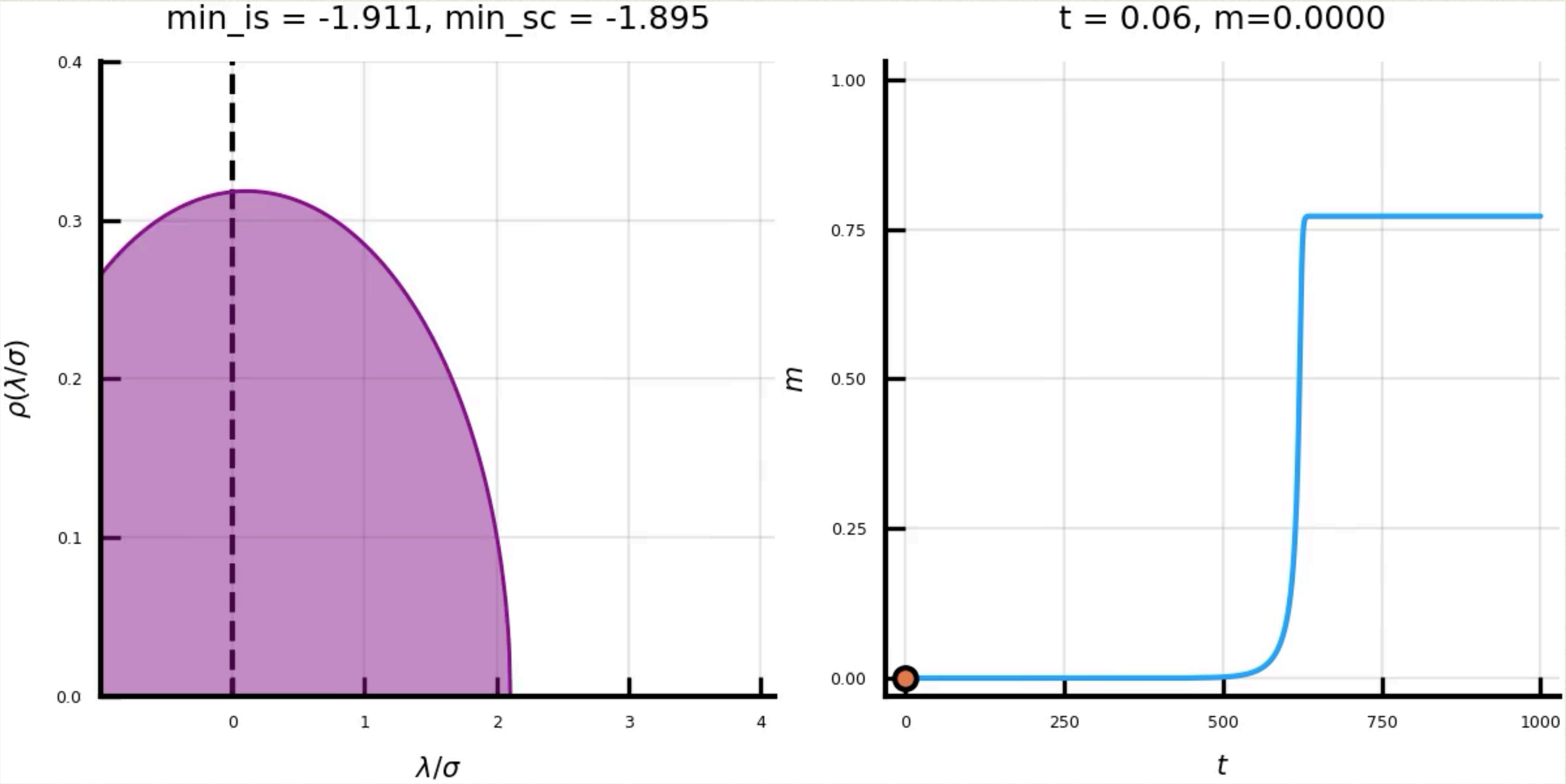
ANALYTICAL SOLUTION
(LANDCSAPE ANALYSIS)



$\lambda_2 = \frac{1}{\Delta_2}$

$\frac{1}{\lambda_p} = \Delta_p$

IN A NUTSHELL



CONCLUSIONS…

- Spherical spike matrix-tensor problem has interesting properties; Many quantities can be computed (O*ptimal performance*s, *energy landscape, performance of AMP, Langevin, Gradient descent…)*

- Observed Gap between Langevin sampling and message passing performances: MCMC not as good as Langevin?

- Minimisation algorithms are observed to work just fine even in presence of (exponentially many) spurious minima

… PERSPECTIVES?

- More on monte-carlo sampling in M. C. Angelini's talk tomorrow
- Other non convex learning and signal processing problems (e.g. Phase retrieval, see Antoine Maillard's talk tomorrow)
- Effect of prior information (see Bruno Loureiro's talk next)
- Neural networks: Single layer perceptron, teacher-student multi-layer deep networks, over-parametrization, etc…
- Non convex setting with other gradient-based algorithms - SGD, Nesterov, momentum, etc ….,
(Recent papers by Zdeborova, Mignacco, Urbani…)

REFERENCES FOR THIS TALK

- Marvels and pitfalls of the Langevin algorithm in noisy high-dimensional inference; Sarao, Biroli, Cammarota, FK, Urbani, Zdeborova, **PRX '19**.

- Passed & Spurious: Descent Algorithms and Local Minima in Spiked Matrix-Tensor Models; Sarao, FK, Urbani, Zdeborova, **ICML'19**, arXiv:1902.00139.

- Who is Afraid of Big Bad Minima? Analysis of Gradient-Flow in a Spiked Matrix-Tensor Model; Sarao, Biroli, Cammarota, FK, Urbani, Zdeborova. **NeurIPS '19.**

… and we want YOU in Switzerland

We are looking for talented postdocs & students!

Apply Now