STATISTICAL PHYSICS OF LEARNING (REVISITED)

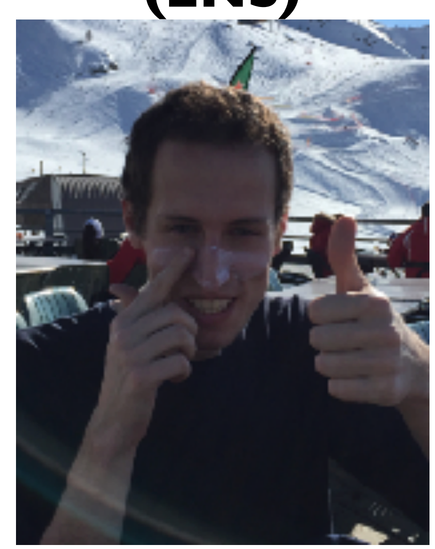
Florent Krzakala

Ecole Normale Supérieure, Paris

Jean Barbier
(EPFL→London)

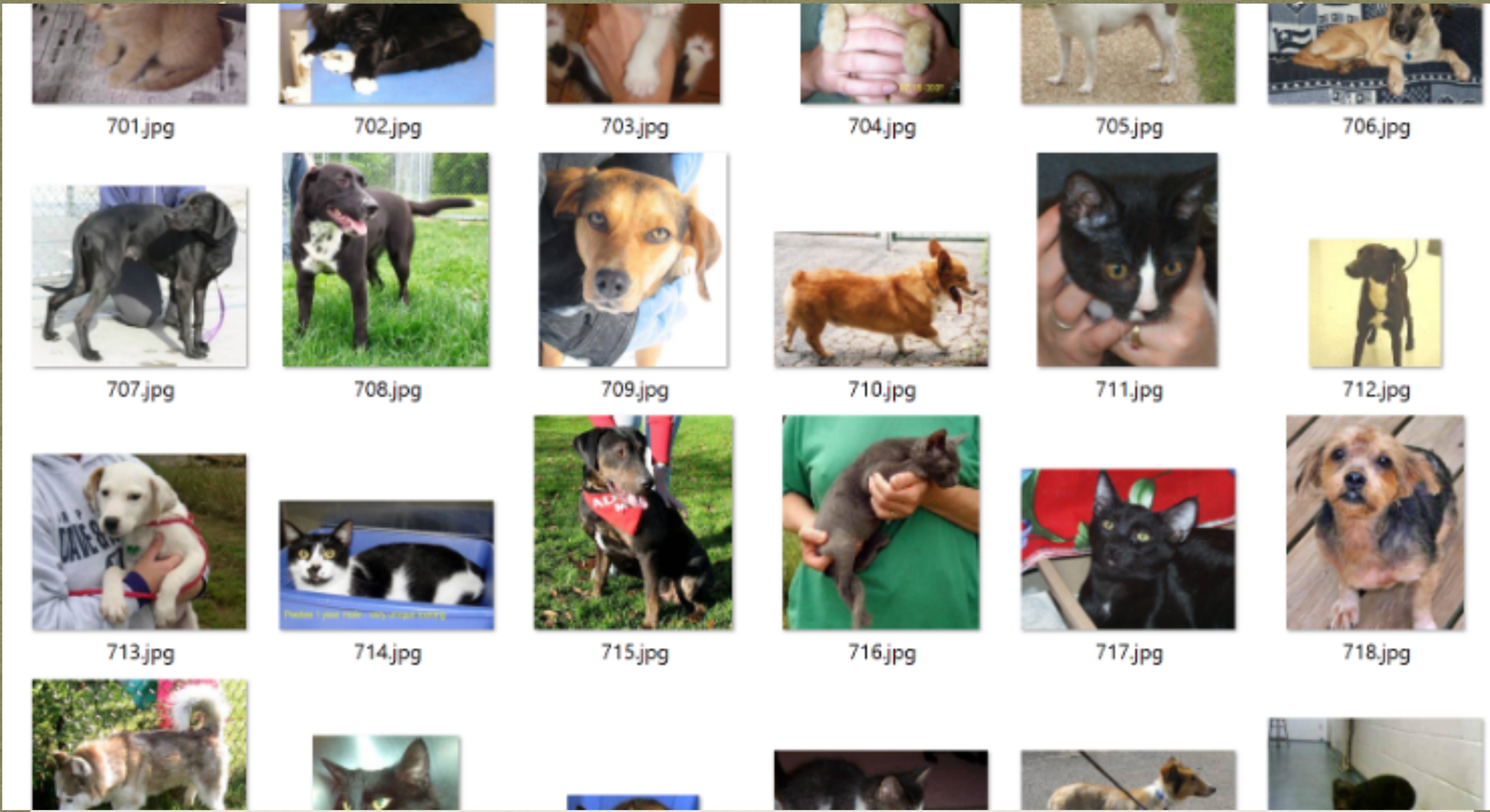
Lenka Zdeborova
(CNRS, Saclay)

Léo Miolane
(ENS)



Nicolas Macris
(EPFL)

LEARNING A RULE



STATIST

PHYSICAL REVIEW A VOLUME 41, NUMBER 12 15 JUN

RAPID COMM

First-order transition to perfect generalization in a neural network with binary synapses

Géza Györgyi*
rgia Institute of Technology, Atlanta, Georgia 30332-0430

VOLUME 76, NUMBER 11 PHYSICAL REVIEW LETTERS 11 MARCH 1996

Mean Field Approach to Bayes Learning in Feed-Forward Neural Networks

Manfred Opper
Institut für Theoretische Physik, Julius-Maximilians-Universität, Am Hubland, D-97074 Würzburg, Germany

Ole Winther
CONNECT, The Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen Ø, Denmark

(Received 6 October 1995)

We propose an algorithm to realize Bayes optimal predictions for feed-forward networks which is based on the Thouless-Anderson-Palmer mean field method developed for the statistical mechanics of disordered systems. We conjecture that our approach will be exact in the thermodynamic limit. The algorithm results in a simple built in leave one out cross validation of the predictions. Simulations for the case of the simple perceptron and the committee machine are in excellent agreement with the results of replica theory.

PACS numbers: 87.10.+e, 64.60.Cn

Computer Physics Communications
lumes 121–122, September–October 1999, Pages 86–93

l physics of neural networks

J. Phys. A: Math. Gen. **22** (1989) 1983–1994. Printed in the UK

**Three unfinished works on the optimal storage capa
of networks**

E Gardner and B Derrida
The Institute for Advanced Studies, The Hebrew University of Jerusale
and Service de Physique Théorique de Saclay†, F-91191 Gif-sur-Yvet

Received 13 December 1988

Abstract. The optimal storage properties of three different neural
studied. For two of these models the architecture of the network is a
interactions, whereas for the third model the output can be an arbit
inputs. Analytic bounds and numerical estimates of the optimal ca
minimal fraction of errors are obtained for the first two models. The
solved exactly and the exact solution is compared to the bounds a
numerical simulations used for the two other models.

A Statistical Approach to Learning and Generalization in Layered Neural Networks

STHER LEVIN, NAFTALI TISHBY, AND SARA A. SOLLA

*A general statistical description of the problem of learning from
xamples is presented. Our focus is on learning in layered net-
orks, which is posed as a search in the network parameter space
r a network that minim*

works with the same parameter space. The common method
of parameter estimation is the maximum likelihood (ML)

Abs
The
studi
com

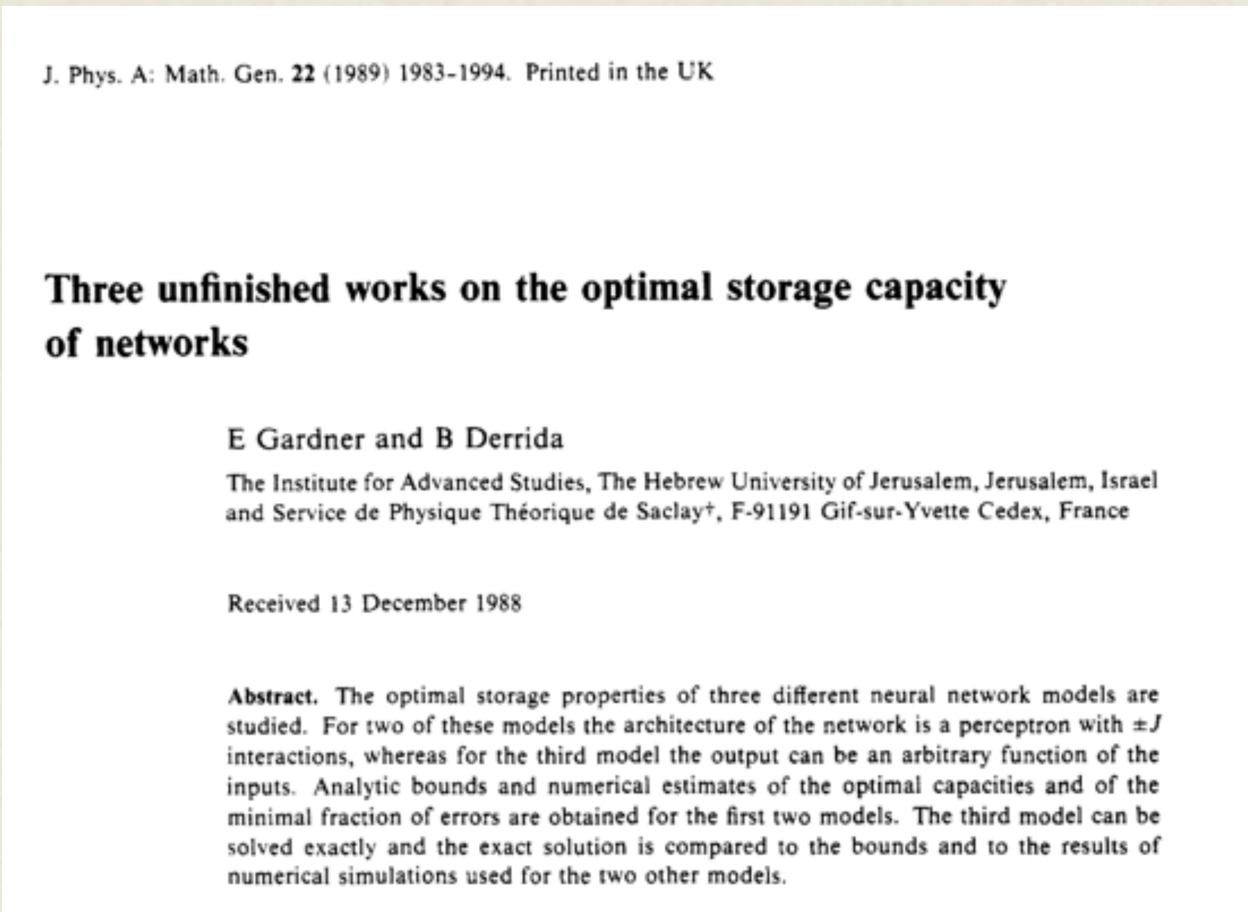
Physics of Biomaterials: Fluctuations, Selfassembly and Evolution pp 309-325 | Cite

Statistical Physics of Neural Networks

Authors Authors and affiliations

David Sherrington

EX:TEACHER-STUDENT PERCEPTRON


J. Phys. A: Math. Gen. **22** (1989) 1983-1994. Printed in the UK

Three unfinished works on the optimal storage capacity of networks

E Gardner and B Derrida

The Institute for Advanced Studies, The Hebrew University of Jerusalem, Jerusalem, Israel and Service de Physique Théorique de Saclay†, F-91191 Gif-sur-Yvette Cedex, France

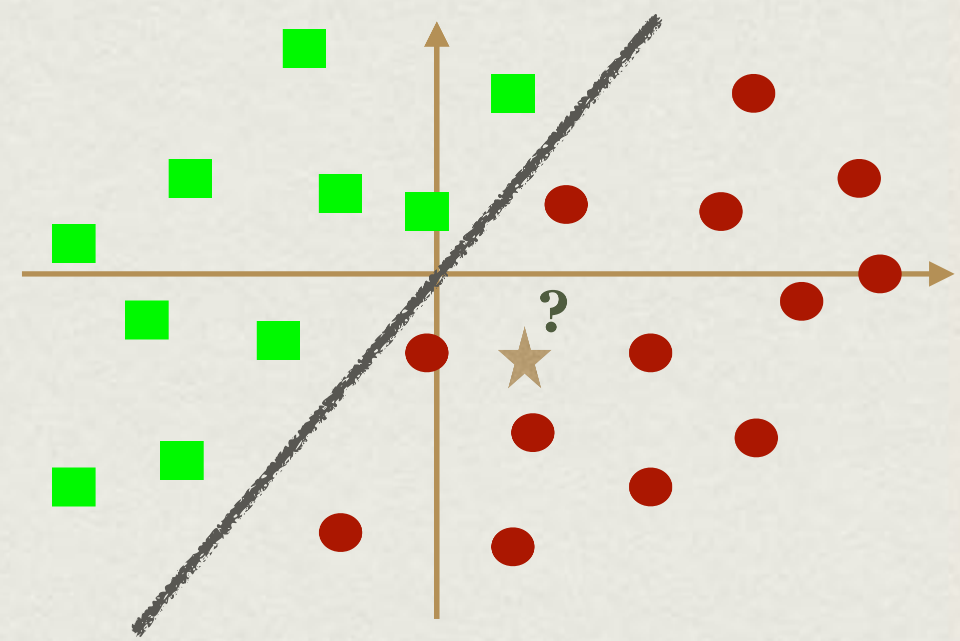
Received 13 December 1988

Abstract. The optimal storage properties of three different neural network models are studied. For two of these models the architecture of the network is a perceptron with $\pm J$ interactions, whereas for the third model the output can be an arbitrary function of the inputs. Analytic bounds and numerical estimates of the optimal capacities and of the minimal fraction of errors are obtained for the first two models. The third model can be solved exactly and the exact solution is compared to the bounds and to the results of numerical simulations used for the two other models.

coordinates of points: $F_{\mu i} \sim \mathcal{N}(0, 1/N)$

hyper-plane parameters: $x_i^* \sim P_X$

labels: $y = \text{sign}(F x^*)$



N=2, M=22, α=M/N=11

$\alpha = O(1), \quad N, M \to \infty$

Goals:

Estimation: Infer x* from M samples/examples of (F_μ, y_μ).
Generalization: Predict labels of new points.

STATISTICAL PHYSICS

a probabilistic process

Generative model

Observation

$$X \xrightarrow{\quad P(Y|X) \quad} Y$$

$P(X)$

Hamiltonian

$$P(X|Y) = \frac{P(Y|X)P(X)}{Z} = \frac{e^{\log\left(P(Y|X)P(X)\right)}}{Z}$$

Posterior probability

Partition sum

TWO QUESTIONS

Many predictions from the heuristic replica method:

- Asymptotic free energy/mutual information Derrida & Gardner 89'
- Optimal overlap & optimal generalization Gyorgyi - Sompolinski & al 90's

▷ can we prove them mathematically rigorously?

What about practical algorithmic performance?

- Not the focus of the physics literature until the early 00's

▷ Can we reach optimal performances with efficient algorithms?

Both answered in this talk.

GENERALIZED LINEAR REGRESSION

component-wise function

$$y = \varphi_\xi(Fx)$$

e.g.: $\varphi_\xi(x) = \mathrm{sign}(z)$

$x_i \in \{\pm 1\}$

labels: $y \in \mathbb{R}^M$

data matrix: $F \in \mathbb{R}^{M \times N}$

regression parameters: $x \in \mathbb{R}^N$

noise: $\xi \in \mathbb{R}^M$

- Goal: Find good fitting parameters x, from examples (F_μ,y_μ).
- **Special cases:** Compressed sensing, signal reconstruction in computed tomography, magnetic resonance imaging, phase retrieval. LASSO, superposition error correcting codes, code-division multiple-access problem, group testing, logistic regression, ...

MAIN RESULTS

Barbier, FK, Macris, Miolane, Zdeborova, arXiv:1708.03395, COLT 2018

Def. free entropy: $f \equiv \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{y,F} \log Z(y,F)$ $\qquad \alpha = \frac{M}{N}$

Theorem 1 (informally): The replica free entropy is correct.

$$f = \sup_{m} \inf_{\hat{m}} f_{RS}(m, \hat{m})$$

$$f_{RS}(m,\hat{m}) = \Phi_{P_X}(\hat{m}) + \alpha \Phi_{P_{\rm out}}(m;\rho) - \frac{m\hat{m}}{2}$$

where

$$\Phi_{P_X}(\hat{m}) \equiv \mathbb{E}_{z,x_0}\left[\ln \mathbb{E}_x \left[e^{\hat{m}xx_0 + \sqrt{\hat{m}}xz - \hat{m}x^2/2}\right]\right]$$

$$\Phi_{P_{\rm out}}(m;\rho) \equiv \mathbb{E}_{v,z}\left[\int {\rm d}y\, P_{\rm out}(y|\sqrt{m}\,v + \sqrt{\rho - m}\,z) \ln \mathbb{E}_w \left[P_{\rm out}(y|\sqrt{m}\,v + \sqrt{\rho-m}\,w)\right]\right]$$

$x, x_0 \sim P_X \qquad z, v, w \sim \mathcal{N}(0,1) \qquad \rho = \mathbb{E}_{P_X}(x^2)$

MAIN RESULTS

Barbier, FK, Macris, Miolane, Zdeborova, arXiv:1708.03395, COLT 2018

Def. free entropy: $f \equiv \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{y,F} \log Z(y,F)$ $\qquad \alpha = \frac{M}{N}$

Theorem 1 (informally): The replica free entropy is correct.

$$f = \sup_{m} \inf_{\hat{m}} f_{RS}(m, \hat{m})$$

$$f_{RS}(m, \hat{m}) = \Phi_{P_X}(\hat{m}) + \alpha \Phi_{P_{\text{out}}}(m; \rho) - \frac{m\hat{m}}{2}$$

Theorem 2 (informally): Optimal error on estimation of x* is:

$$\text{MMSE} = \rho - m^*$$

where m* is the extremizer of f_{RS}. $\qquad \rho = \mathbb{E}_{P_X}(x^2)$

MAIN RESULTS

Barbier, FK, Macris, Miolane, Zdeborova, arXiv:1708.03395, COLT 2018

Def. free entropy: $f \equiv \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_{y,F} \log Z(y,F)$ $\alpha = \frac{M}{N}$

Theorem 1 (informally): The replica free entropy is correct.

$$f = \sup_{m} \inf_{\hat{m}} f_{RS}(m, \hat{m})$$

$$f_{RS}(m,\hat{m}) = \Phi_{P_X}(\hat{m}) + \alpha \Phi_{P_{\text{out}}}(m;\rho) - \frac{m\hat{m}}{2}$$

Theorem 3 (informally): Optimal generalisation error is

$$\mathcal{E}_{\text{gen}} = \mathbb{E}_{v,\xi}\left[f_\xi(\sqrt{\rho}\, v)^2\right] - \mathbb{E}_{v}\,\mathbb{E}_{w,\xi}\left[f_\xi(\sqrt{m^*}\, v + \sqrt{\rho - m^*}\, w)\right]^2$$

where m* is the extremizer of f_{RS}.

$\rho = \mathbb{E}_{P_X}(x^2)$

$v, w \sim \mathcal{N}(0,1)$

$\xi \sim P_\xi$

PROOF IDEA

Notice $$f_{RS}(m,\hat{m}) = \Phi_{P_X}(\hat{m}) + \alpha\Phi_{P_{\text{out}}}(m;\rho) - \frac{m\hat{m}}{2}$$

$\Phi_{P_X}(\hat{m})$ is the free entropy of a scalar denoising problem

$$y' = \sqrt{\hat{m}}x^* + \xi$$

$$\xi \sim \mathcal{N}(0,1)$$

$$x^* \sim P_X$$

$\Phi_{P_{\text{out}}}(m;\rho)$ is the free entropy of a scalar denoising problem

$$\tilde{y} \sim P_{\text{out}}(\tilde{y}|\sqrt{m}\,v + \sqrt{\rho - m}\,z^*)$$

$$v, z^* \sim \mathcal{N}(0,1)$$

$$\rho = \mathbb{E}_{P_X}(x^2)$$

PROOF IDEA

Guerra-Toninelli-like interpolation between the original posterior and N + M independent scalar denoising problems.

Interpolating Hamiltonian:

$$\mathcal{H}_t = -\sum_{\mu=1}^{M} \ln P_{\text{out}}(y_\mu | s_{t,\mu}) + \frac{1}{2}\sum_{i=1}^{N}(y'_i - \sqrt{t\hat{m}}x_i)^2$$

$$s_{t,\mu} = \sqrt{1-t}[Fx]_\mu + \sqrt{\int_0^t m(t')dt'}v_\mu + \sqrt{\int_0^t (\rho - m(t'))dt'}z_\mu$$

$$f_N = f_N(t=1) - \int_0^1 f'_N(t)dt$$

PROOF IDEA

$$f_N = f_{\text{replica}}(m, \hat{m}) + \int_0^1 \mathcal{R}(m(t), \hat{m}, t)dt$$

with $\mathcal{R}(m(t), \hat{m}, t)$ a complicated function

Adaptive interpolation:
Choose interpolation path m(t) to cancel the integral
(Crucial contribution by Barbier & Macris, *Probab. Theory Relat. Fields '08*)

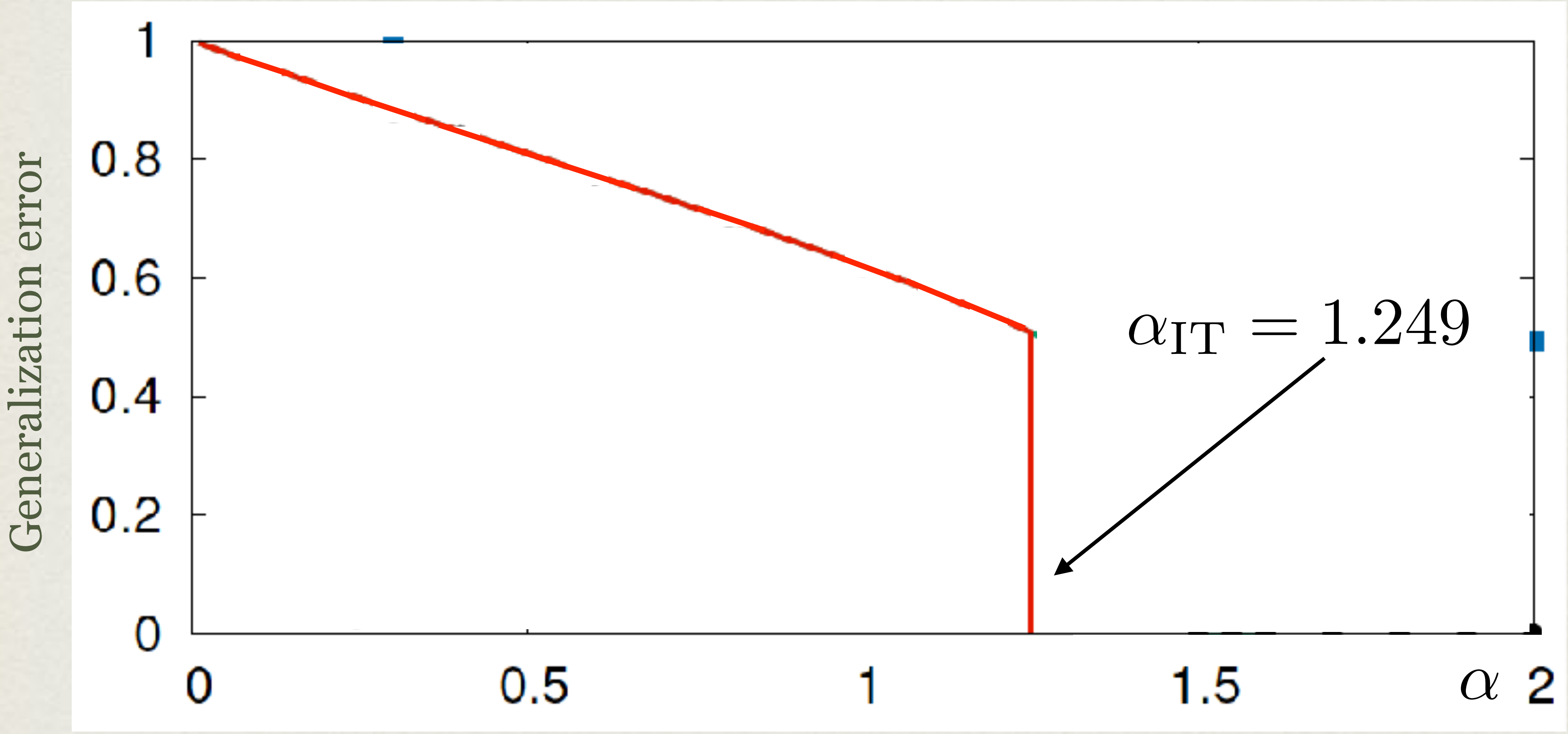
Key Ingredient: (Nishimori/Bayes theorem): Under expectations ground truth x* is exchangeable for a sample from P(x|y,F).

$$\mathbb{E}_{x^*,y}\mathbb{E}_{P(x|y)}[g(y, x^{(1)}, x^*)] = \mathbb{E}_y\mathbb{E}_{P(x|y)}[g(y, x^{(1)}, x^{(2)})]$$

□

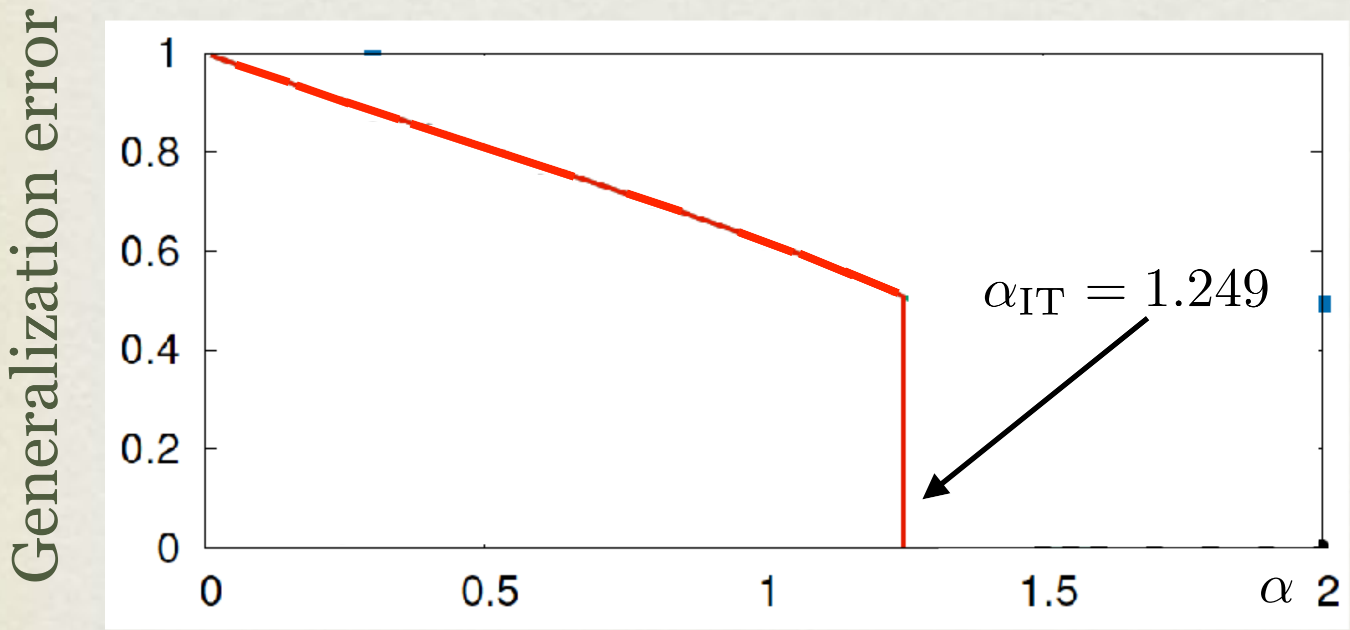
EX: BINARY PERCEPTRON TRANSITION TO PERFECT LEARNING

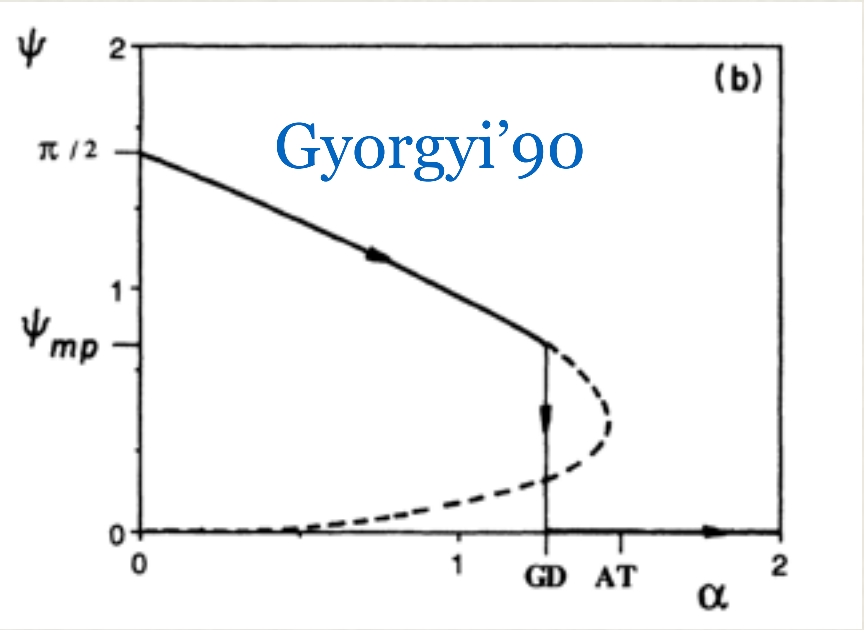
$$y = \text{sign}(Fx^*) \qquad P_X(x) = \frac{1}{2}[\delta(x-1) + \delta(x+1)]$$



EX: BINARY PERCEPTRON TRANSITION TO PERFECT LEARNING

$$y = \text{sign}(Fx^*) \qquad P_X(x) = \frac{1}{2}[\delta(x-1) + \delta(x+1)]$$



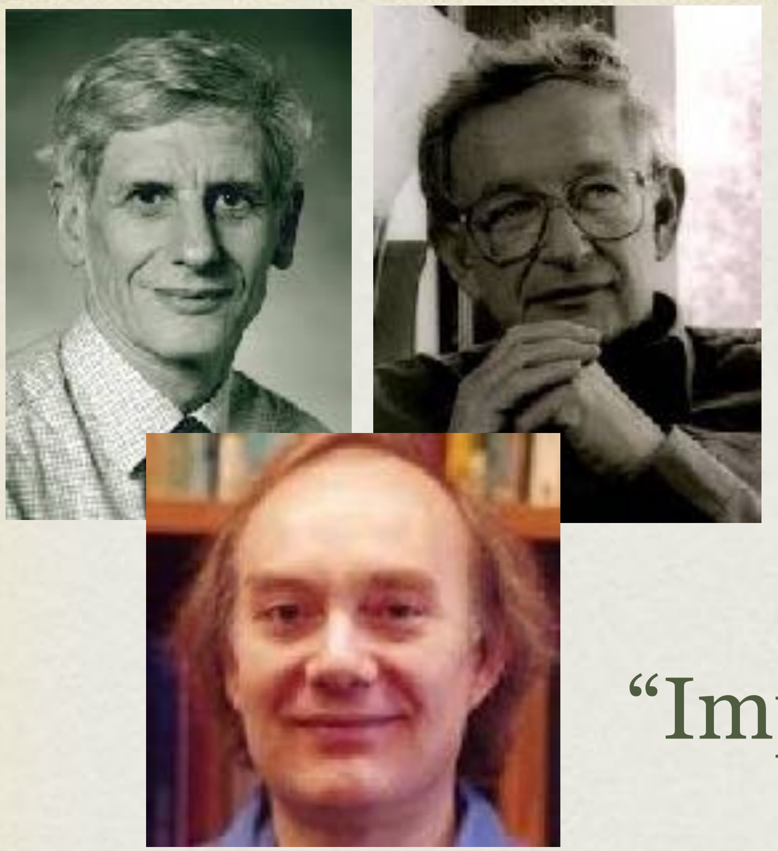


CLOSING A 28 YEARS OLD CONJECTURE…

… AND SOLVING MANY NEW PROBLEMS:
COMPRESSED SENSING, PHASE RETRIEVAL,
SUPERPOSITION CODES, LOGISTIC REGRESSION…

WHAT ABOUT EFFICIENT ALGORITHMS ?

A GENERIC APPROACH: THE TAP EQUATIONS



Solution of'solvable model of a spin glass'
DJ **Thouless**, PW **Anderson**, RG **Palmer** - Philosophical Magazine, 1977 - Taylor & Francis
Abstract The Sherrmgton-Kirkpatrick model of a spin glass is solved by a mean field technique which is probably exact in the limit of infinite range interactions. At and above T c the solution is identical to that obtained by Sherrington and Kirkpatrick (1975) using the n→ ...

"Improved" mean-field equations for spin glasses

$$H_N(\sigma, J) = -\frac{1}{\sqrt{N}} \sum_{(i,j)} J_{ij}\sigma_i\sigma_j - h\sum_i \sigma_i,$$

$$m_i^{t+1} = \tanh\left[h + \sum_j \beta J_{ij} m_j^{t} - \beta^2 \sum_j J_{ij}^2 (1 - (m_j^{t})^2) m_i^{t}\right]$$

THOULESS-ANDERSON-PALMER ACCORDING TO BOLTHAUSEN

An Iterative Construction of Solutions of the TAP Equations for the Sherrington–Kirkpatrick Model

Erwin Bolthausen*

Institute of Mathematics, Universität Zürich, Zürich, Switzerland. E-mail: eb@math.uzh.ch

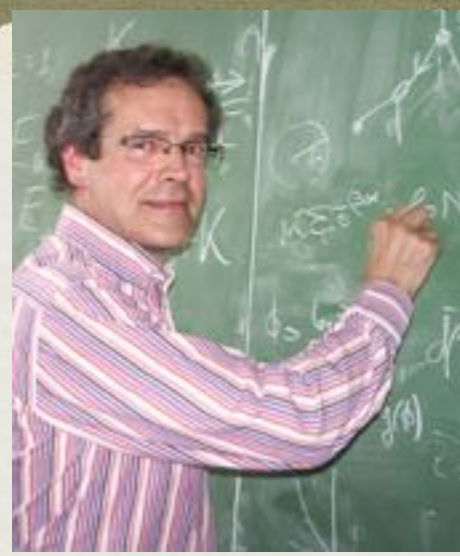
Received: 1 June 2012 / Accepted: 12 June 2013
Published online: 27 December 2013 – © Springer-Verlag Berlin Heidelberg 2013

Abstract: We propose an iterative scheme for the solutions of the TAP-equations in the Sherrington–Kirkpatrick model which is shown to converge up to and including the de Almeida–Thouless line. The main tool is a representation of the iterations which reveals an interesting structure of them. This representation does not depend on the temperature parameter, but for temperatures below the de Almeida–Thouless line, it contains a part which does not converge to zero in the limit.

Convergence of the equations is now proven!

$$m_i^{t+1} = \tanh\left[h + \sum_j \beta J_{ij} m_j^{t} - \beta^2 \sum_j J_{ij}^2 (1 - (m_j^{t})^2) m_i^{t-1} \right]$$

TAP EQUATIONS FOR PERCEPTRON



M. Mézard ('89)

J. Phys. A: Math. Gen. **22** (1989) 2181–2190. Printed in the UK

The space of interactions in neural networks: Gardner's computation with the cavity method

Marc Mézard

Laboratoire de Physique Théorique de l'Ecole Normale Supérieure†, 24 rue Lhomond, 75231 Paris Cedex 05, France

Received 16 January 1989

Abstract. Gardner's computation of the number of N-bit patterns which can be stored in an optimal neural network used as an associative memory is derived without replicas, using the cavity method. This allows for a unified presentation whatever the basic measure in the space of coupling constants, but above all it gives the clear physical content of the assumption of replica symmetry. TAP equations are also derived.

4. TAP equations

The previous results were derived by averaging over many samples, within the assumption of the existence of one single pure state. However, as explained by Mézard *et al* (1987) the cavity method can be used to study one single sample (always in the thermodynamic limit). In the present section we shall derive TAP-like equations (Thouless *et al* 1977) which are mean-field equations valid for one given sample, within one pure state.

Considering a sample with N couplings J_i and p patterns $\{\xi_i^\mu\}$, we shall need the following order parameters:

$$m_j = \langle J_j \rangle_{N,p} \qquad x^\mu_{N,p} = (1 - e^{-\beta}) P^\mu_{N,p}(K) \tag{30}$$

$$m_i = f_1\left(\frac{1}{\sqrt{N}} \sum_\mu \xi_i^\mu x_\mu - (B - A) m_i\right) \qquad \forall i = 1, \ldots, N$$

$$x^\mu = g_1\left(\frac{1}{\sqrt{N}} \sum_i \xi_i^\mu m_i - (q_1 - q_0) x_\mu\right) \qquad \forall \mu = 1, \ldots, p$$

where the parameters q_0, q_1, A and B, are given by

$$q_0 = \frac{1}{N} \sum_i m_i^2$$

$$q_1 = \frac{1}{N} \sum_i f_2\left(\frac{1}{\sqrt{N}} \sum_\mu \xi_i^\mu x^\mu - (B - A) m_i\right)$$

$$A = \frac{1}{N} \sum_\mu (x^\mu)^2$$

$$B = \frac{1}{N} \sum_\mu g_2\left(\frac{1}{\sqrt{N}} \sum_i \xi_i^\mu m_i - (q_1 - q_0) x^\mu\right)$$

APPROXIMATE MESSAGE PASSING
THE MODERN AVATAR OF TAP

$$
\begin{cases}
V^t &= \overline{\mathbf{v}^t} \\
\boldsymbol{\omega}^t &= \boldsymbol{\Phi}\mathbf{a}^{t-1}/\sqrt{n} - V^t\mathbf{g}^{t-1} \\
g^t_\mu &= g_{\rm out}(\omega^t_\mu, Y_\mu, V^t) \quad \forall\ \mu = 1,\dots m \\
\Sigma^t &= \left(\alpha\, \overline{g^2_{\rm out}(\boldsymbol{\omega}^t, \mathbf{Y}, V^t)}\right)^{-1} \\
\mathbf{R}^t &= \mathbf{a}^{t-1} + \boldsymbol{\Phi}^\intercal\mathbf{g}^t/(\Sigma^t\sqrt{n}) \\
a^t_i &= \eta(\Sigma^t, R^t_i) \quad \forall\ i = 1,\dots n \\
v^t_i &= \Sigma^t\,\partial_R\eta(\Sigma^t, R)|_{R=R^t_i} \quad \forall\ i = 1,\dots n
\end{cases}
$$

Mean and variance
of the marginals:

$$a_i^{t+1} = f(A^t, B_i^t)$$

$$v_i^{t+1} = \left(\frac{\partial f}{\partial B}\right)(A^t, B_i^t)$$




$$g_{\rm out}(\omega, y, V) \equiv \frac{\int {\rm d}z P_{\rm out}(y|z)\,(z-\omega)\,e^{-\frac{(z-\omega)^2}{2V}}}{V\int {\rm d}z P_{\rm out}(y|z) e^{-\frac{(z-\omega)^2}{2V}}}$$

$$\eta(\Sigma, R) = \frac{\int {\rm d}x\, x\, P_X(x)\, e^{-\frac{(x-R)^2}{2\Sigma}}}{\int {\rm d}x\, P_X(x)\, e^{-\frac{(x-R)^2}{2\Sigma}}}$$

Message Passing Algorithms for Compressed Sensing
https://arxiv.org › cs ▾
by DL Donoho - 2009 - Cited by 1092 - Related articles
Jul 21, 2009 - Authors: David L. **Donoho**, Arian **Maleki**, Andrea **Montanari**. (Submitted on 21 Jul ...
From: Andrea **Montanari** [view email] [v1] Tue, 21 Jul 2009 ...

Generalized approximate message passing for estimation with random linear mixing
S Rangan - Information Theory Proceedings (ISIT), 2011 IEEE ..., 2011 - ieeexplore.ieee.org
We consider the estimation of a random vector observed through a linear transform followed by a componentwise probabilistic measurement channel. Although such linear mixing estimation problems are generally highly non-convex, Gaussian approximations of belief ...
★ 99 Cited by 469 Related articles All 9 versions

RIGOROUS STATE EVOLUTION

Bolthausen ’06, Bayati-Montanari ’10, Rangan ’10, Donoho-Javamart-Montanari ‘12

Define: $m^t \equiv \frac{1}{p}\sum_{i=1}^{p} x_i^* a_i^t$ then $\mathrm{MSE}(t) = \rho - m^t$

m^t in the AMP algorithm ($n, p \to \infty, \alpha = \Theta(1)$) evolves as:

$$m^{t+1} = 2\partial_{\hat{m}}\Phi_{P_X}(\hat{m}^t)$$

$$\hat{m}^t = 2\alpha\partial_m\Phi_{P_{\text{out}}}(m^t;\rho)$$

Recall the RS free entropy?

$$f_{RS}(m,\hat{m}) = \Phi_{P_X}(\hat{m}) + \alpha\Phi_{P_{\text{out}}}(m;\rho) - \frac{m\hat{m}}{2}$$

AMP is simply trying to extremize the replica function!

WHEN DOES AMP WORK?

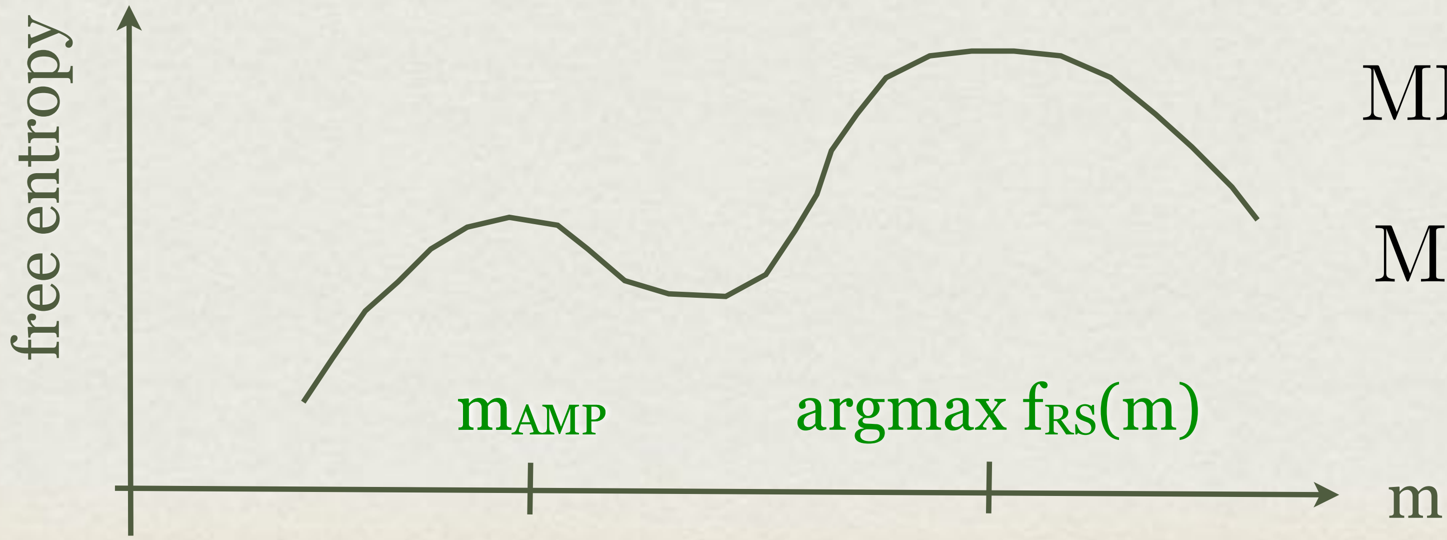
Barbier, FK, Macris, Miolane, Zdeborova, arXiv:1708.03395, COLT 2018

$$f_{RS}(m,\hat{m}) = \Phi_{P_X}(\hat{m}) + \alpha\Phi_{P_{\rm out}}(m;\rho) - \frac{m\hat{m}}{2}$$

$$f_{RS}(m) = \inf_{\hat{m}} f_{RS}(m,\hat{m})$$

- AMP-MSE given by the local maximum of the free entropy reached starting from small m/large MSE.

- Optimal MSE (MMSE) given by the global maximum of the free entropy.

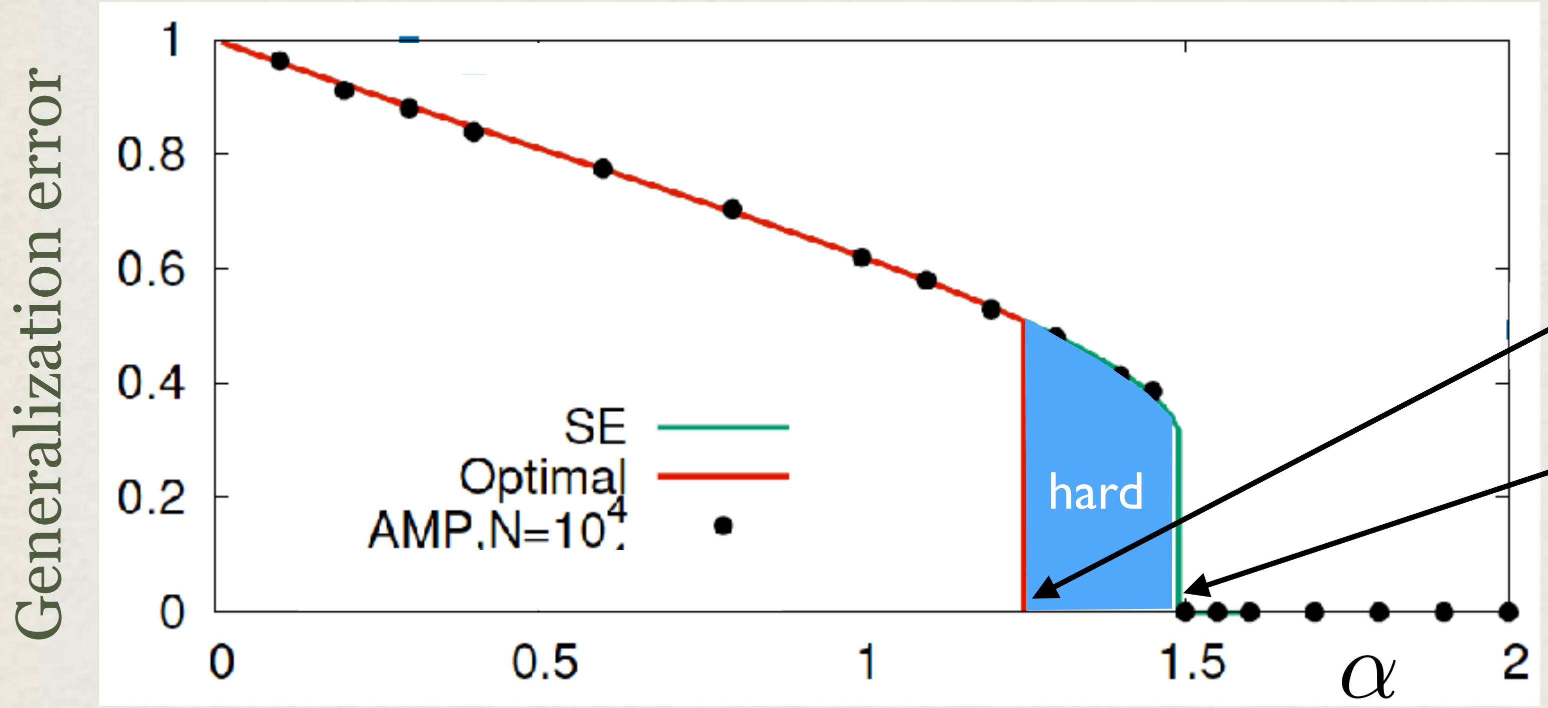


$$\mathrm{MMSE} = \rho - \mathrm{argmax} f_{RS}(m)$$

$$\mathrm{MSE_{AMP}} = \rho - m_{\rm AMP}$$

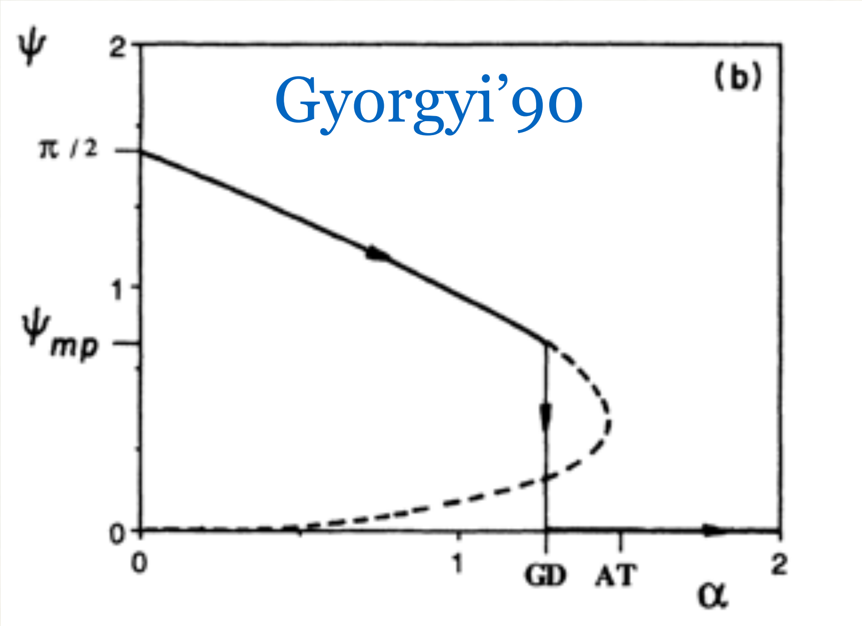
EX: BINARY PERCEPTRON

$$y = \text{sign}(Fx^*) \qquad P_X(x) = \frac{1}{2}[\delta(x-1) + \delta(x+1)]$$



$\alpha_{\text{IT}} = 1.249$

$\alpha_{\text{Alg}} = 1.493$



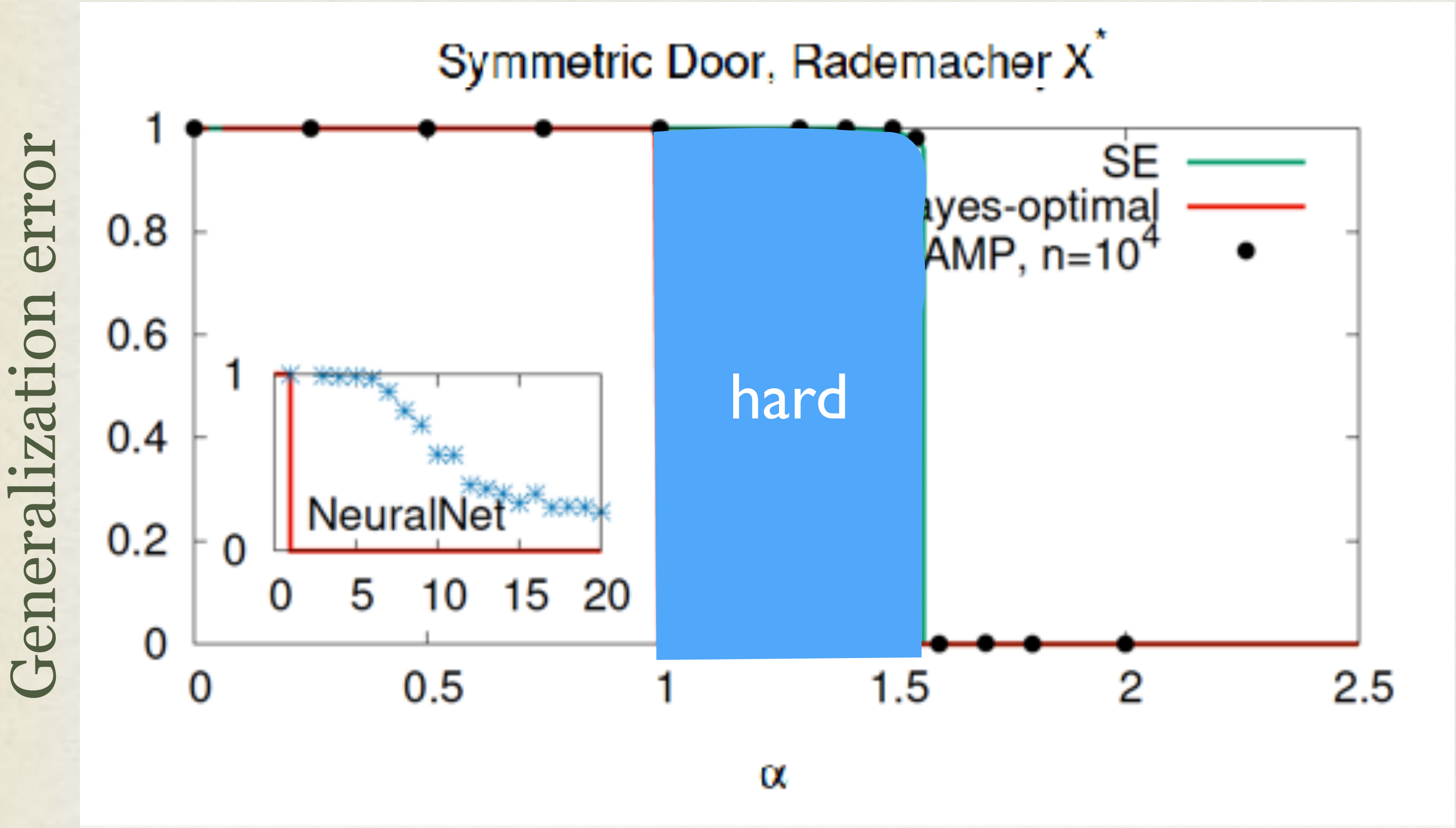
- GAMP is optimal, out of the hard phase.
- Redemption of the "un-physical" branch.

THIS HAPPENS ALL THE TIME!

SYMMETRIC BINARY PERCEPTRON

$$y = \text{sign}(|Fx^*| - K)$$

$$P_X(x) = \frac{1}{2}[\delta(x-1) + \delta(x+1)]$$



$$\alpha_{\text{IT}} = 1$$

$$\alpha_{\text{Alg}} = 1.566$$

K chosen so that P(y=1)=0.5

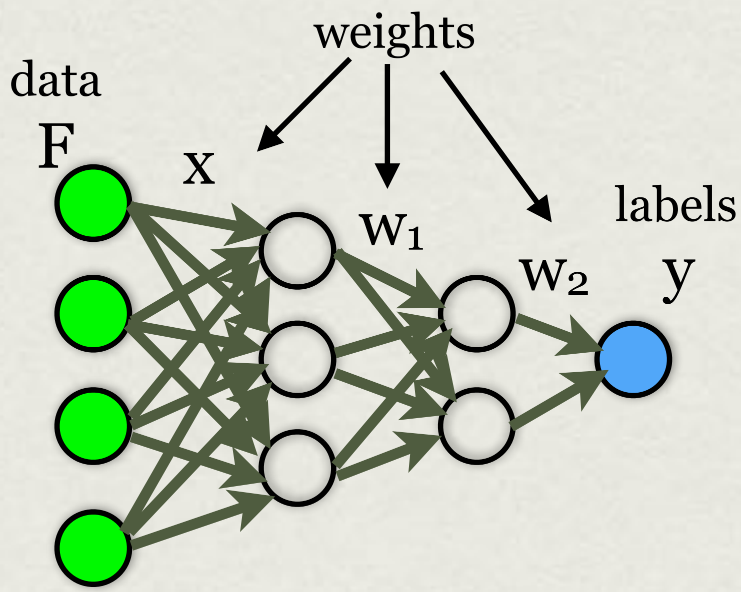
from: Barbier, FK, Macris, Miolane, Zdeborova arXiv:1708.03395

MULTI-LAYER NETWORK

Committee machine

Quenched teacher-student model studied in Schwarze'92.
Proof and approximate message passing (Aubin, Maillard, Barbier, Macris, FK, Zdeborova '18)

- p input units
- K hidden units
- output unit
- n training samples



L=3 layers

x learned, w fixed

Limit: $\alpha = \frac{M}{N} = O(1) \quad K = O(1) \quad M, N \to \infty$

MULTI-LAYER NETWORK

$$y_\mu = \text{sign}\Big[\sum_{l=1}^{K} \text{sign}\big(\sum_{i=1}^{p} F_{\mu i} x^*_{il}\big)\Big]$$

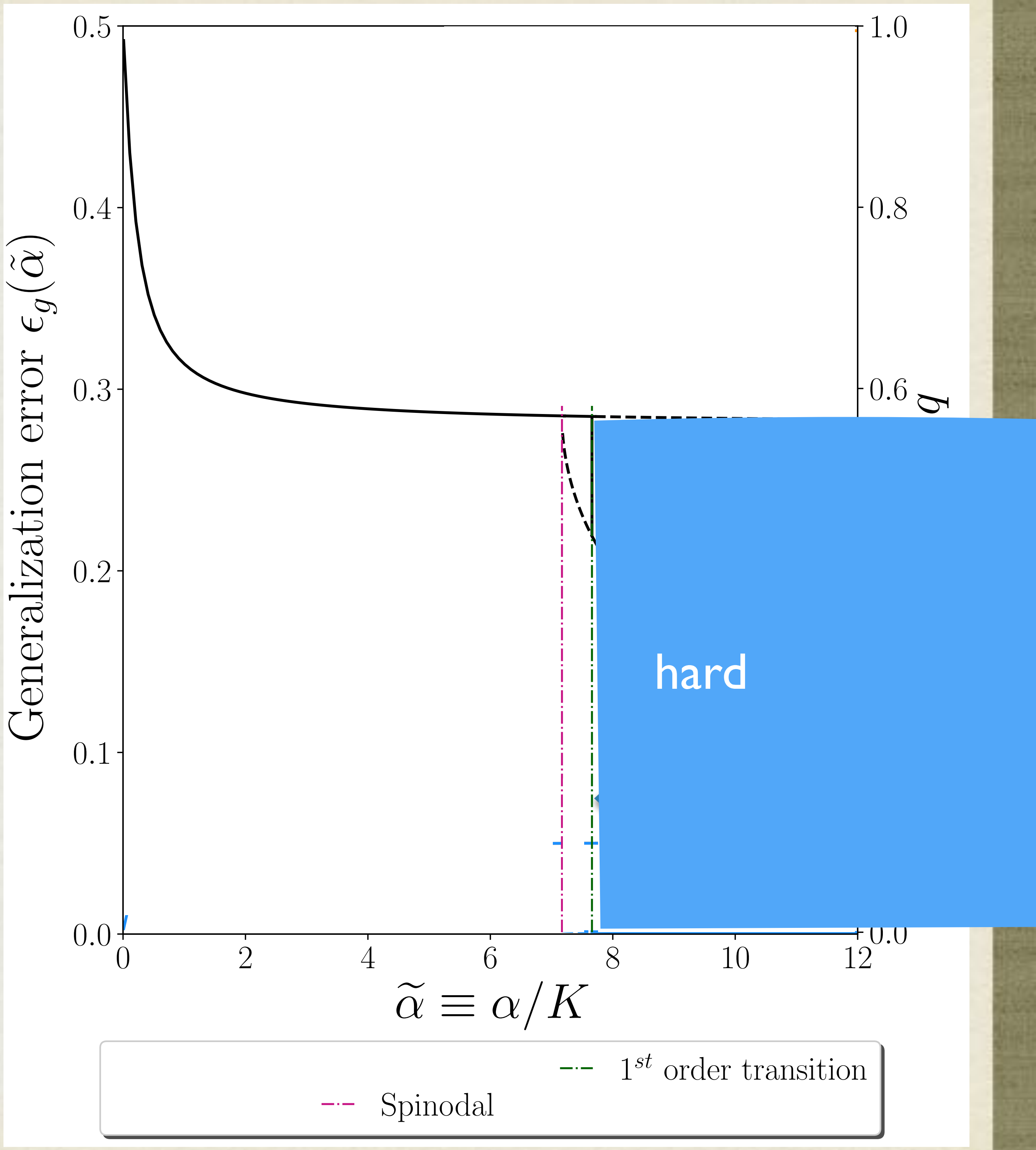
$$K \gg 1$$

- Large algorithmic gap:
  - IT threshold:

$$\alpha_{\text{IT}} = 7.65K$$

  - Algorithmic threshold

$$\alpha_{\text{Algo}} = O(K^2)$$



HARD PHASE EVERYWHERE

Hard phase = spinodal region of first order phase transitions.

Fact : So far all known polynomial algorithms fail in the hard phase!

Conjecture:

AMP/TAP achieves the lowest error among all polynomial algorithms (asymptotically)

Hard phase identified in:

- stochastic block model
- dense planted sub-matrix;
- low-rank tensor completion;
- compressed sensing;
- planted constraint satisfaction;
- Gaussian mixture clustering;
- low-density parity check error correcting codes;
- sparse principal component analysis;
- generalised linear regression;
- dictionary learning;
- blind source separation;
- learning in binary perceptron;
- phase retrieval; ...

CONLUSIONS

- Proof of the replica formulas and prediction for many learning/inference problems (here perceptron, but similar results holds for generalised linear models, matrix and tensor factorisation, simple neural nets …)
- A mean-field type algorithm acheive optimal prediction in polytime…
- … unless metastability appears, in which case we believe nothing works!
- Many more challenging mathematical physics problems in learning and computer science (see Lenka Zdeborova tomorrow at 15.45)

Shameless advertising:

Many Postdoc positions opened in Ecole Normale in Paris

1794

ENS

ÉCOLE NORMALE SUPÉRIEURE

References

The committee machine: Computational to statistical gaps in learning a two-layers neural network
B. Aubin, A. Maillard, J. Barbier, FK, N. Macris and L. Zdeborová arXiv:1806.05451

Fundamental limits of detection in the spiked Wigner model
A. El Alaoui, FK, M. I. Jordan arXiv:1806.09588

Phase Transitions, Optimal Errors and Optimality of Message-Passing in Generalized Linear Models
J. Barbier, FK, N. Macris, L. Miolane, L. Zdeborová, *COLT 2018*

Statistical and computational phase transitions in spiked tensor estimation
T. Lesieur, L. Miolane, M. Lelarge, FK, L. Zdeborová, *ISIT 2017*

Information-theoretic thresholds from the cavity method
A. Coja-Oghlan, FK, W. Perkins, L. Zdeborova *STOC 2017 & Advances in Mathematics 2018*

The Mutual Information in Random Linear Estimation
J. Barbier, M. Dia, N. Macris, FK, Allerton 2017

Mutual information for symmetric rank-1 matrix estimation: A proof of the replica formula
J. Barbier, M. Dia, N. Macris, FK, T. Lesieur, L. Zdeborova, *NIPS 2017*

Mutual Information in Rank-One Matrix Estimation
FK, J. Xu, L. Zdeborová, *ITW 2016*

Thank you for your attention!