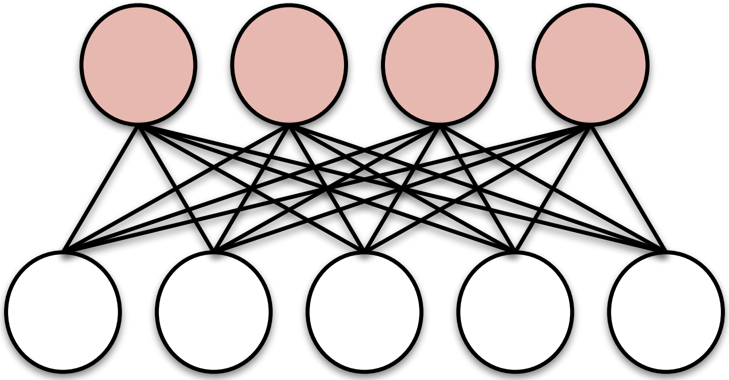
Inferring effective couplings with Restricted Boltzmann Machines



Alfonso Navas
Predoctoral Researcher
Collaborators:
Aurélien Decelle, Cyril Furtlehner and Beatriz Seoane
Complutense University of Madrid
Department of Theoretical Physics

Talk Outline

1. Introduction
2. … From the RBM to the Ising Model
3. Results of Simple Numerical Experiments
4. Conclusion and further directions

Talk Outline

1. **Introduction**
2. … From the RBM to the Ising Model
3. Results of Simple Numerical Experiments
4. Conclusion and further directions

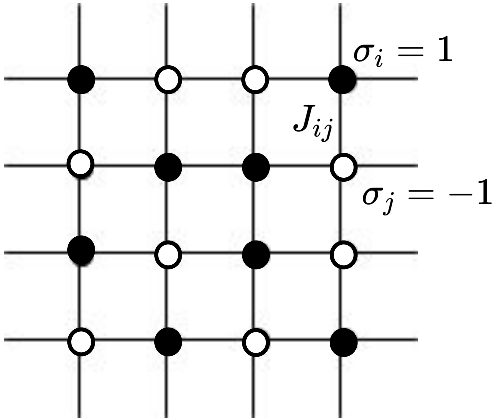
Introduction: The *inverse* Ising problem

Hamiltonian of the system

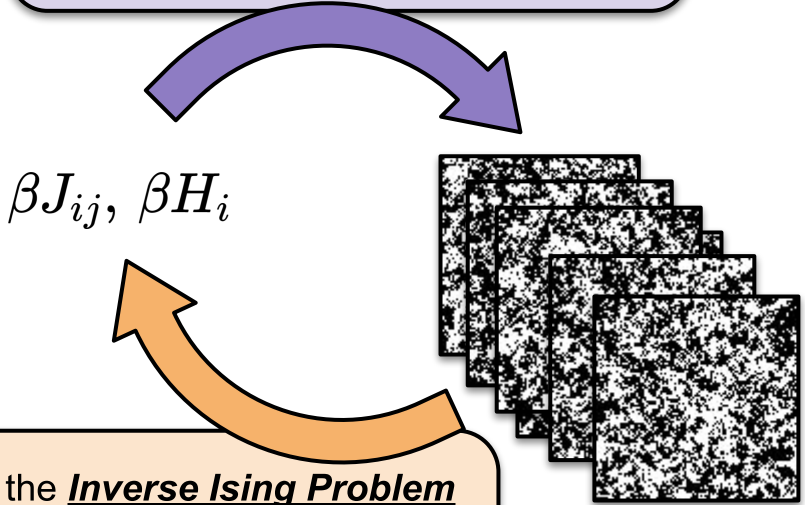
$$\mathcal{H}_{\text{Ising}}(\sigma) = -\sum_{ij} J_{ij}\sigma_i\sigma_j - \sum_i H_i\sigma_i$$

Boltzmann Distribution

$$p(\sigma) = \frac{1}{\mathcal{Z}} e^{-\beta\mathcal{H}_{\text{Ising}}(\sigma)}, \; \beta = \frac{1}{k_B T}$$



The ***forward problem*** consists in **deriving the macroscopic properties of a system from a simple microscopic model.**

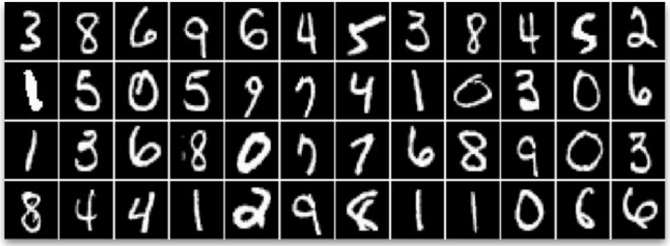


On the ***<u>Inverse Ising Problem</u>*** we want **to infer the parameters from observed spins configurations**.

Introduction: Inverse problems in Machine Learning

Given a Data set,

$$X = \{x^{(1)},\, x^{(2)},\ \ldots x^{(M)}\}$$



We adjust the parameters of a model such that **the empirical distribution of the data set fits the distribution of the model:**

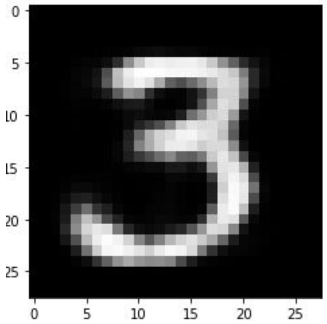
$$p_{\text{data}}(x) \sim p_\theta(x) = \frac{e^{-E_\theta(x)}}{\mathcal{Z}_\theta}$$

When the probability distribution of the model is given by a Boltzmann distribution, one has an ***Energy Based Model.***

$E_\theta(x)$ ⇨ *Energy function*

Effective model of the data
which can be used for:
modelling, interpretability or generating.

Example:
(Generating)



Which Energy function should we use?

We can try a solution such that **the magnetizations and correlations of the model match the corresponding values of the data** (i.e. ***Boltzmann Learning***).

$$\langle x_i \rangle_{\mathrm{model}} = \langle x_i \rangle_{\mathrm{data}}, \ \langle x_i x_j \rangle_{\mathrm{model}} = \langle x_j x_j \rangle_{\mathrm{data}}$$

By imposing the above constraints and using ***maximum entropy principles***, it is possible to obtain the following energy function

$$E_\theta(x) = -\sum_{ij} J_{ij} x_i x_j - \sum_i H_i x_i$$

Such an energy function is indeed the Ising model Hamiltonian!

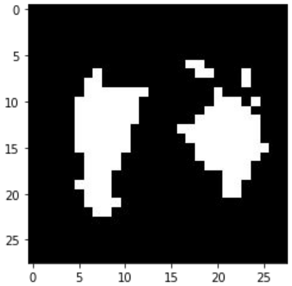
Can this energy function encode any data set?

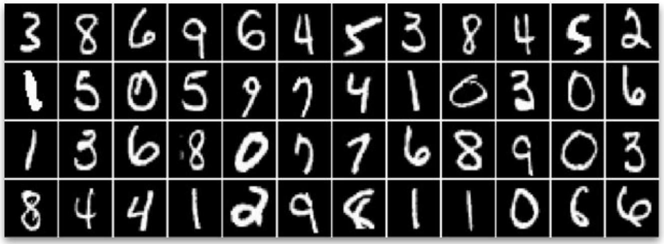
$$E_\theta(x) = -\sum_{ij} J_{ij} x_i x_j - \sum_i H_i x_i$$

$$p_\theta(x) = \frac{e^{-E_\theta(x)}}{\mathcal{Z}_\theta}$$

This energy function cannot includes beyond 2-body correlations!

Example:





… We could try to include higher order correlations by adding more parameters:

$$E_\theta(x) = -\sum_i h_i x_i - \sum_{ij} J^{(2)}_{ij} x_i x_j - \sum_{ijk} J^{(3)}_{ijk} x_i x_j x_k - \sum_{ijkl} J^{(4)}_{ijkl} x_i x_j x_k x_l + \dots$$

the number of parameters quickly diverges!

…Any solution?

Talk Outline

1. Introduction
2. **… From the RBM to the Ising Model**
3. Results of Simple Numerical Experiments
4. Conclusion and further directions

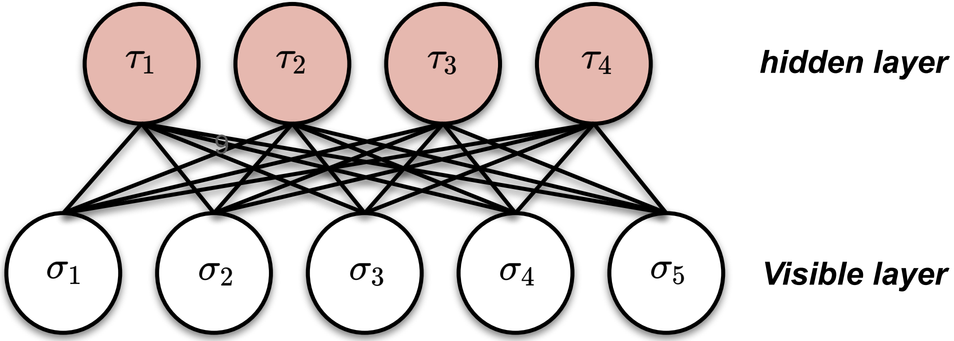
The Restricted Boltzmann Machine

Hamiltonian of the RBM:

$$\mathcal{H}(\sigma,\tau) = -\sum_{i=1}^{N_h}\sum_{j=1}^{N_v}\tau_i w_{ij}\sigma_j - \sum_{j=1}^{N_v}\eta_j\sigma_j - \sum_{i=1}^{N_h}\theta_i\tau_i$$

Boltzmann Distribution

$$p(\sigma,\tau) = \frac{e^{-\mathcal{H}(\sigma,\tau)}}{\mathcal{Z}},\ \beta = 1$$



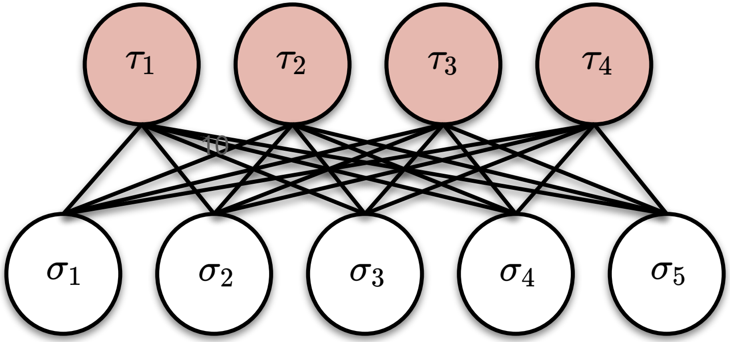
Let us marginalize over the hidden nodes:

$$p(\sigma) = \frac{1}{\mathcal{Z}}\sum_{\tau} e^{-\mathcal{H}(\sigma,\tau)} = \frac{1}{\mathcal{Z}} e^{-\mathcal{H}(\sigma)}$$

$$\Rightarrow \mathcal{H}(\sigma) = -\sum_j \eta_j\sigma_j - \sum_i \ln\cosh\left(\sum_j w_{ij}\sigma_j + \theta_i\right)$$

Adding *latent variables (hidden layer)* allows us to encode higher order correlations!

RBM for Inverse Problems?



As the relative importance of higher order couplings in inverse problems remains unexplored,

We can use the latent model learned by an RBM to fill this gap!

RBM effective Hamiltonian:

$$\mathcal{H}(\sigma) = -\sum_j \eta_j \sigma_j - \sum_i \ln\cosh\left(\sum_j w_{ij}\sigma_j + \theta_i\right)$$

Thus, we will expand the RBM energy function as generalized Ising model Hamiltonian

Generalized Ising Hamiltonian:

$$\mathcal{H}(\sigma) = -\sum_j H_j \sigma_j - \sum_{j_1 > j_2} J_{ij} \sigma_{j_1} \sigma_{j_2} - \sum_{j_1 > j_2 > j_3} J_{j_1 j_2 j_3} \sigma_{j_1} \sigma_{j_2} \sigma_{j_3} + \ldots$$

…from the RBM to the Ising Model

Step 1:* ***Expansion of the Hamiltonian***

$$\mathcal{H}(\sigma) = -\sum_j \eta_j \sigma_j - \sum_i \ln\cosh\left(\sum_j w_{ij}\sigma_j + \theta_i\right)$$

$$= -\sum_j \eta_j \sigma_j - \sum_{\sigma'} \prod_j \delta_{\sigma_j \sigma'_j} \sum_i \ln\cosh\left(\sum_j w_{ij}\sigma'_j + \theta_i\right)$$

$$\delta_{\sigma_j \sigma'_j} = \frac{1}{2}\left(1 + \sigma_j \sigma'_j\right) \rightarrow = -\sum_j \eta_j \sigma_j - \frac{1}{2^{N_v}} \sum_{\sigma'} \prod_j \left(1 + \sigma_j \sigma'_j\right) \sum_i \ln\cosh\left(\sum_j w_{ij}\sigma'_j + \theta_i\right)$$

<u>*Step 1*:*</u> ***Expansion of the Hamiltonian***

$$\mathcal{H}(\sigma) = -\sum_j \eta_j \sigma_j - \frac{1}{2^{N_v}} \sum_{\sigma'} \prod_j (1+\sigma_j \sigma'_j) \sum_i \ln\cosh\left(\sum_j w_{ij}\sigma'_j + \theta_i\right)$$
$$= -\sum_j H_j \sigma_j - \sum_{j_1>j_2} J_{ij}\sigma_{j_1}\sigma_{j_2} - \sum_{j_1>j_2>j_3} J_{j_1 j_2 j_3}\sigma_{j_1}\sigma_{j_2}\sigma_{j_3} + \dots$$

By comparing term by term, we finally find

$$H_j = \frac{1}{2^{N_v}} \sum_{\sigma'} \sum_i \sigma'_j \ln\cosh\left(\sum_j w_{ij}\sigma'_j + \theta_i\right) + \eta_j$$

$$J_{j_1\dots j_n} = \frac{1}{2^{N_v}} \sum_{\sigma'} \sum_i \sigma'_{j_1} \dots \sigma'_{j_n} \ln\cosh\left(\sum_j w_{ij}\sigma'_j + \theta_i\right)$$

…. These expressions cannot be exactly computed!

Step 2:* ***Numerical approximation of found expressions***

Let us introduce the random variable

$$X_i^{(j_1 \dots j_n)} = \sum_{\mu=n+1}^{N_v} w_{ij_\mu} s_{j_\mu}$$

Each s_{j_μ} is a random variable defined over the binary support $\{-1, 1\}$

Then, we rewrite our expressions in terms of expected values of such variables

$$H_j = \eta_j + \sum_i \mathbb{E}_{X_i^{(j)}} \left[\ln \frac{\cosh\left(\theta_i + w_{ij} + X_i^{(j)}\right)}{\cosh\left(\theta_i - w_{ij} + X_i^{(j)}\right)} \right]$$

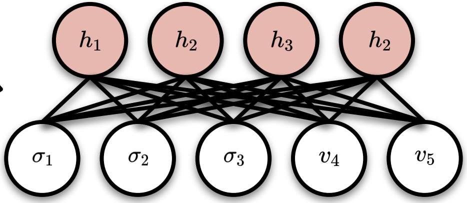
$$J_{j_1 \dots j_n} = \frac{1}{2^n} \sum_i \mathbb{E}_{X_i^{(j_1 \dots j_n)}} \left[\sum_{\sigma'_{j_1}} \dots \sum_{\sigma'_{j_n}} \sigma'_{j_1} \dots \sigma'_{j_n} \ln\cosh\left(\sum_{\mu=1}^{n} w_{ij_\mu} \sigma'_{j_\mu} + X_i^{(j_1 \dots j_n)} + \theta_i \right) \right]$$

These expressions can be easily implemented as numeric integrals using the Central Limit Theorem :)

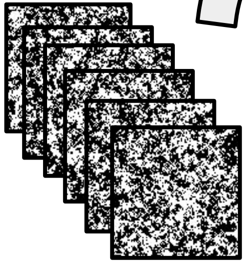
Talk Outline

1. Introduction
2. … From the RBM to the Ising Model
3. **Results of Simple Numerical Experiments**
4. Conclusion and further directions

What did we do?



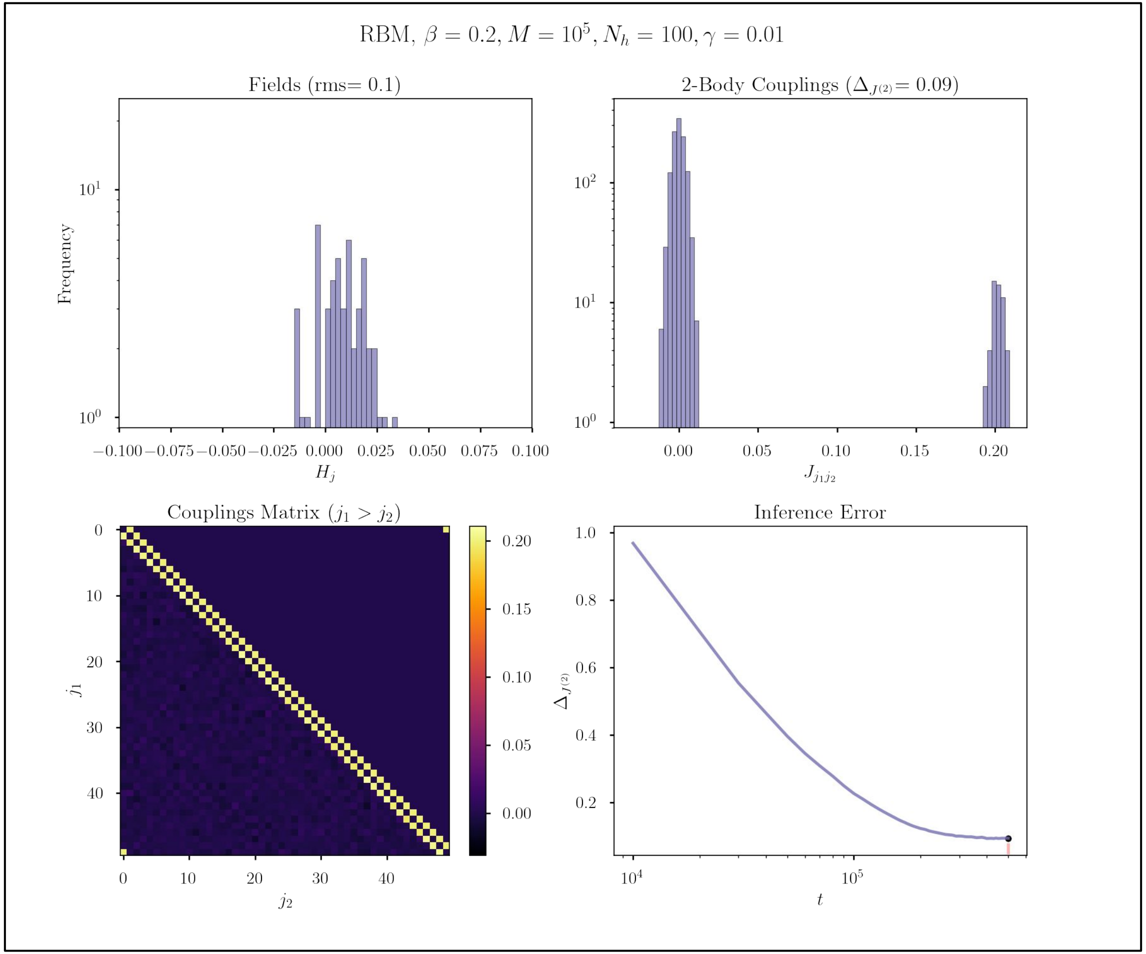
2. Trained RBM with generated samples



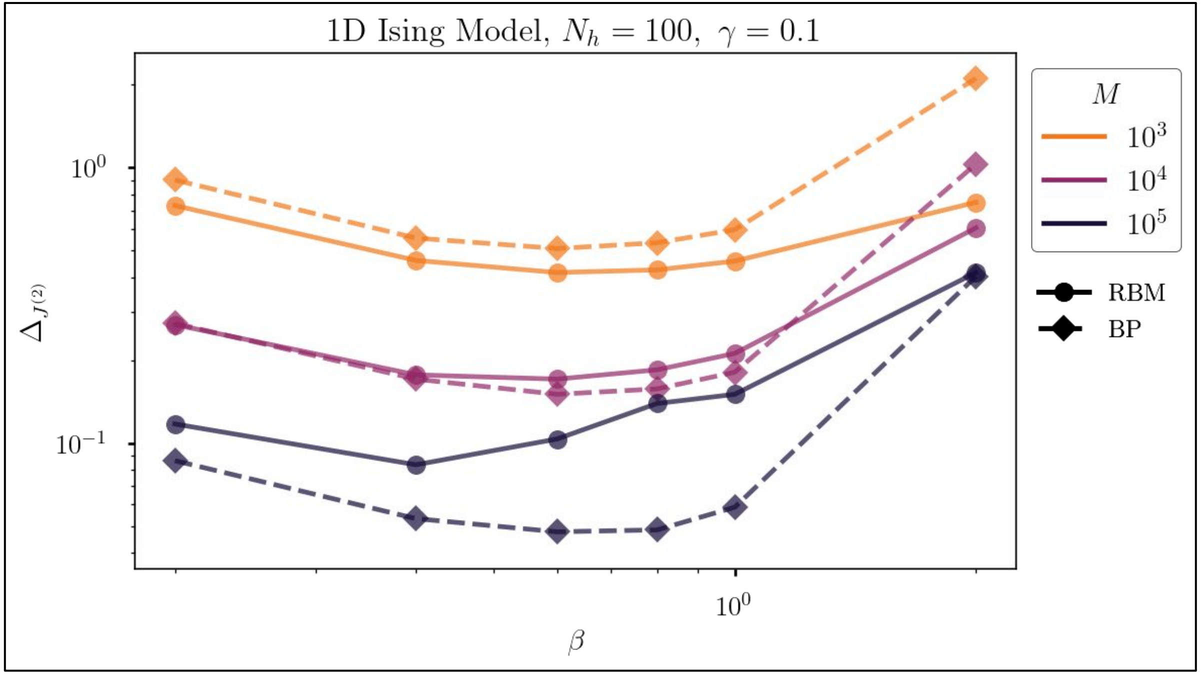
1. Generated Ising model samples

3. Inferred the effective spins model out of the trained RBM models

Learning the 1D Ising Model



Temperature dependence for 1D Inference



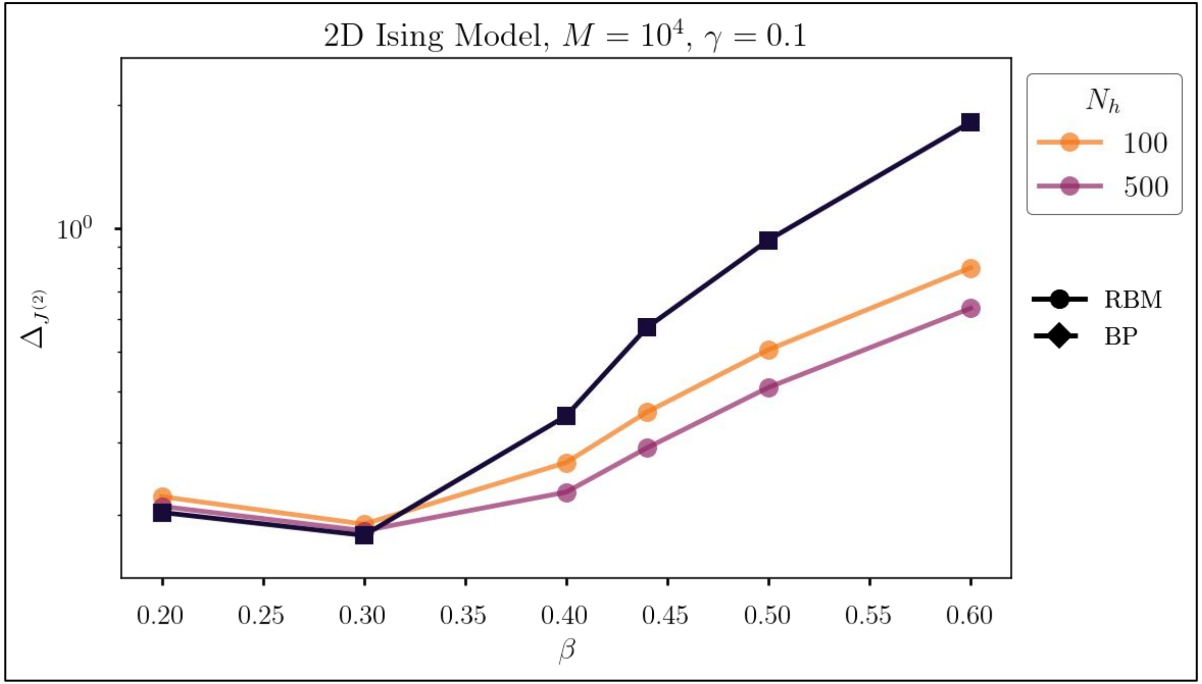
The ***Bethe-Peierls (BP) solution*** for inverse Ising problem is **exact when the graph of couplings is a tree.**

$$J_{ij}^{\mathrm{BP}} = -\frac{1}{2}\arcsin\left[2\left(C^{-1}\right)_{ij}\right]$$

Therefore, **we expect that BP performs particularly well for a ferromagnetic 1D Ising model**

RBMs can outperform state-of-the-art methods at
low temperature regimes and **when training data set is small**.

Temperature dependence for 2D Inference



Increasing the number of hidden nodes seems to improve the inference at low temperature

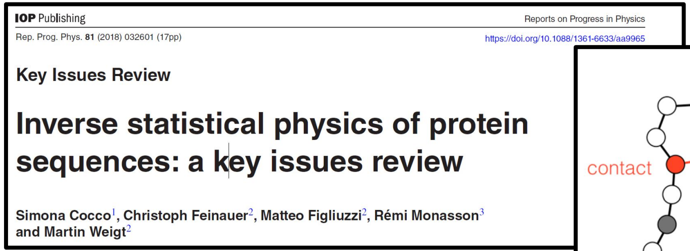
We have to evaluate if the decrease in the low temperature regimes is due to out-of-equilibrium training

Conclusions and Perspectives

- We introduced a method that allows us to extract the effective model learned by an RBM trained with Ising Model **(i.e., we solve an Ising inverse problem using RBMs)**.

- The next steps to be considered:

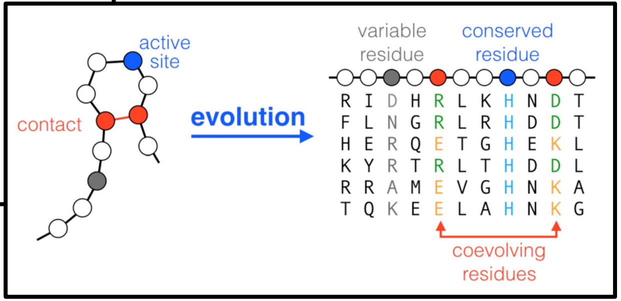
  - **Development of efficient training methods for RBMs.**
  - **Extension of a such inference method to other models** (e.g., Potts Model)


IOP Publishing Reports on Progress in Physics
Rep. Prog. Phys. 81 (2018) 032601 (17pp) https://doi.org/10.1088/1361-6633/aa9965

Key Issues Review

Inverse statistical physics of protein sequences: a key issues review

Simona Cocco[1], Christoph Feinauer[2], Matteo Figliuzzi[2], Rémi Monasson[3] and Martin Weigt[2]



Thank you

Any questions?