

# Kinetic Ising Models Accounting for Hidden Activity

Angel Poc-López

February 2022

## 1 Introduction

The Kinetic Ising model allows to model the behavior of  $N$  interacting neurons over time. At time  $t$ , the probability of activation of a neuron  $s_i$  is:

$$P(s_{i,t}|\mathbf{s}_{t-1}) = \frac{e^{s_i(t)h_i(t)}}{2 \cosh h_i(t)}. \quad (1)$$

Here, the activation probability is driven by the effective fields  $\mathbf{h}_t$ , composed of a bias term  $\mathbf{H} = \{H_i\}$  and couplings to units at the previous time step  $\mathbf{J} = \{J_{ij}\}$ ,

$$h_{i,t} = H_i + \sum_j J_{ij}s_j(t-1). \quad (2)$$

We can observe how the parameters governing the system ( $\mathbf{J}$  and  $\mathbf{H}$ ) are fixed over time. However, this might be too rigid for systems with a very complex dynamic.

## 2 Kinetic Ising Model with Hidden Units

In [1], Hertz et al. propose to use a variable  $H_i(t)$  to study the non-stationary case. In contrast to this, we propose to introduce a new variable  $b_i(t)$  (hidden units or neurons) as part of the effective field of each neuron, maintaining the parallel update rule of Eq. 1. In this frame, we can see how the effective field of each neuron  $h_i(t)$  is subsuming the activity of each of the hidden units by means of the weighting parameter  $\mathbf{M}$ :

$$h_i(t) = H_i + \sum_g M_{ig}b_g(t-1) + \sum_j J_{ij}s_j(t-1). \quad (3)$$

Each hidden unit  $b_i(t)$  is a neural network itself, which has connections with visible and hidden units in the previous time step:

$$b_i(t) = \tanh \left( \sum_j K_{ij}s_j(t-1) + \sum_k L_{ik}b_k(t-1) \right). \quad (4)$$

In Fig. 1 we can observe a simple schema of the architecture of the system.

Our aim is to compose a generative model able to reproduce the neural activity of a given dataset. In order to achieve this goal, we will perform a gradient ascent on the likelihood:

$$\mathcal{L} = \sum_{it} s_i(t)h_i(t) - \log 2 \cosh h_i(t) \quad (5)$$

In the following section we show the mathematical derivation required for the gradient ascent.

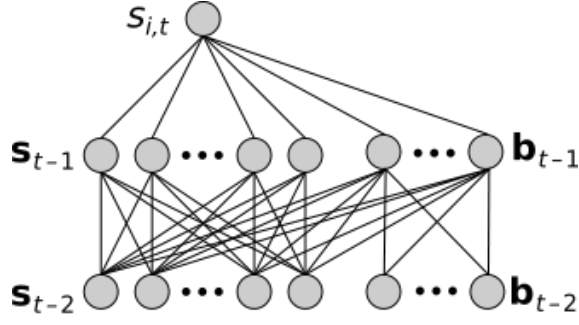


Figure 1: Architecture of the network.

## 2.1 Derivatives

First, we develop the derivative of the likelihood w.r.t.  $\mathbf{H}$  and  $\mathbf{J}$ . These derivatives are equal to the ones in the standard Kinetic Ising model:

$$\frac{\partial \mathcal{L}}{\partial H_n} = \frac{\partial \sum_{it} [s_i(t) h_i(t) - \log 2 \cosh h_i(t)]}{\partial H_n} \quad (6)$$

$$= \sum_t [s_n(t) - \tanh(h_n(t))], \quad (7)$$

$$\frac{\partial \mathcal{L}}{\partial J_{nm}} = \frac{\partial \sum_{it} [s_i(t) h_i(t) - \log 2 \cosh h_i(t)]}{\partial J_{nm}} \quad (8)$$

$$= \sum_t [s_n(t) s_m(t-1)] - \sum_{it} \left[ \frac{\partial \log 2 \cosh h_i(t)}{\partial J_{nm}} \right] \quad (9)$$

$$= \sum_t [s_n(t) s_m(t-1)] - \sum_{it} \left[ \frac{\frac{\partial 2 \cosh h_i(t)}{\partial J_{nm}}}{2 \cosh h_i(t)} \right] \quad (10)$$

$$= \sum_t [s_n(t) s_m(t-1)] - \sum_{it} \left[ \frac{\frac{\partial 2 \cosh h_i(t)}{\partial h_i(t)} \frac{\partial h_i(t)}{\partial J_{nm}}}{2 \cosh h_i(t)} \right] \quad (11)$$

$$= \sum_t [s_n(t) s_m(t-1)] - \sum_t [\tanh(h_n(t)) s_m(t-1)] \quad (12)$$

$$= \sum_t [(s_n(t) - \tanh h_n(t)) (s_m(t-1))]. \quad (13)$$

Secondly, we derive with respect to  $\mathbf{M}$ , resulting in a very similar equation:

$$\frac{\partial \mathcal{L}}{\partial M_{nm}} = \frac{\partial \sum_{it} [s_i(t) h_i(t) - \log 2 \cosh h_i(t)]}{\partial M_{nm}} \quad (14)$$

$$= \sum_t [s_n(t) b_m(t-1)] - \sum_{it} \left[ \frac{\partial \log 2 \cosh h_i(t)}{\partial M_{nm}} \right] \quad (15)$$

$$= \sum_t [(s_n(t) - \tanh h_n(t)) (b_m(t-1))]. \quad (16)$$

Following we derive w.r.t.  $\mathbf{K}$ :

$$\frac{\partial \mathcal{L}}{\partial K_{nm}} = \sum_{it} \left[ \frac{\partial (s_i(t) h_i(t) - \log 2 \cosh h_i(t))}{\partial h_i(t)} \frac{\partial h_i(t)}{\partial K_{nm}} \right] \quad (17)$$

$$= \sum_{it} \left[ (s_i(t) - \tanh h_i(t)) \frac{\partial h_i(t)}{\partial K_{nm}} \right] \quad (18)$$

$$= \sum_{it} \left[ (s_i(t) - \tanh h_i(t)) \frac{\partial \left( \sum_g M_{ig} b_g(t-1) \right)}{\partial K_{nm}} \right] \quad (19)$$

$$= \sum_{it} \left[ (s_i(t) - \tanh h_i(t)) \left( \sum_g M_{ig} \frac{\partial b_g(t-1)}{\partial K_{nm}} \right) \right], \quad (20)$$

Having into account that to complete the derivative, we also need to compute the derivative of  $b_g(t)$  w.r.t.  $\mathbf{K}$ :

$$\frac{\partial b_g(t)}{\partial K_{nm}} = \frac{\partial \tanh \left( \sum_j K_{gj} s_j(t-1) + \sum_k L_{gk} b_k(t-1) \right)}{\partial K_{nm}} \quad (21)$$

$$= \left( \delta_{gn} s_m(t-1) + \sum_k L_{gk} \frac{\partial b_k(t-1)}{\partial K_{nm}} \right) (1 - b_g^2(t)) \quad (22)$$

Subsequently, we derive the likelihood w.r.t.  $\mathbf{L}$  in a very similar way:

$$\frac{\partial \mathcal{L}}{\partial L_{nm}} = \sum_{it} \left[ \frac{\partial (s_i(t) h_i(t) - \log 2 \cosh h_i(t))}{\partial h_i} \frac{\partial h_i}{\partial L_{nm}} \right] \quad (23)$$

$$= \sum_{it} \left[ (s_i(t) - \tanh h_i(t)) \frac{\partial h_i}{\partial L_{nm}} \right] \quad (24)$$

$$= \sum_{it} \left[ (s_i(t) - \tanh h_i(t)) \left( \sum_g M_{ig} \frac{\partial b_g(t-1)}{\partial L_{nm}} \right) \right], \quad (25)$$

Where the derivative of  $b_g(t)$  w.r.t.  $\mathbf{L}$  is:

$$\frac{\partial b_g(t)}{\partial L_{nm}} = \frac{\partial \tanh \left( \sum_j K_{gj} s_j(t-1) + \sum_k L_{gk} b_k(t-1) \right)}{\partial L_{nm}} \quad (26)$$

$$= \left( \delta_{gn} b_m(t-1) + \sum_k L_{gk} \frac{\partial b_k(t-1)}{\partial L_{nm}} \right) (1 - b_g^2(t)) \quad (27)$$

Finally, to avoid a cold start setting  $b(t=0)$  to 0, we define the derivative of the likelihood w.r.t.  $\mathbf{b}(t=0)$

$$\frac{\partial \mathcal{L}}{\partial b_z(0)} = \frac{\partial \sum_{it} [s_i(t) h_i(t) - \log 2 \cosh h_i(t)]}{\partial b_z(0)} \quad (28)$$

$$= \frac{\partial \sum_i [s_i(1) h_i(1) - \log 2 \cosh h_i(1)]}{\partial b_z(0)} + \frac{\partial \sum_i [s_i(2) h_i(2) - \log 2 \cosh h_i(2)]}{\partial b_z(0)} \quad (29)$$

$$= \sum_i (s_i(1) - \tanh h_i(1)) M_{iz} + \sum_i (s_i(2) - \tanh h_i(2)) \left( \sum_g M_{ig} (1 - b_g^2(1)) L_{gz} \right) \quad (30)$$

### 3 Results

To test our method, we compare the performance of our method in contrast to the standard Kinetic Ising model. We are interested in a scenario in which there is a model generating data but only a portion of these data is visible to the models. To reproduce this, we first simulate a model with  $N=10$  total neurons for 5000 steps where we only collect the data from  $V=6$  visible neurons. Subsequently, we try to infer the parameters that maximize the likelihood both with a standard Kinetic Ising model and with a Kinetic Ising model with a unique hidden unit. We assess the inference process doing a simulation of the network every 250 iterations for 6500 time-steps. With the data generated in the simulations, we compute the mean ( $\mathbf{m}$ ), same time correlations ( $\mathbf{C}$ ) and delayed correlations ( $\mathbf{D}$ ).

In order to make sure that the results were not biased by the initialization of each of the algorithms, we have repeated the experiments several times randomly modifying several aspects driven with a seed. These aspects include: the parameters  $\mathbf{H}$  and  $\mathbf{J}$  of the original model, the random selection of the neurons out of the original set and the initialization of the parameters of the new models.

In all the figures displayed in the appendix A we can observe the same behavior. If we look at the bottom part of the figures we will see that the Mean Squared Error (MSE) of the all the statistical moments  $\mathbf{m}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  computed after the simulations is always lower or equal in the models with a hidden unit. This is more remarkable in most of the cases if we observe the MSE of the mean and the delayed correlations. The results can be reproduced using the script `sim_fit_HI2.py` and modifying the seed.

### References

- [1] John Hertz, Yasser Roudi, and Joanna Tyrcha. Ising models for inferring network structure from spike data, 2011.

# A Figures

## A.1 Seed 722

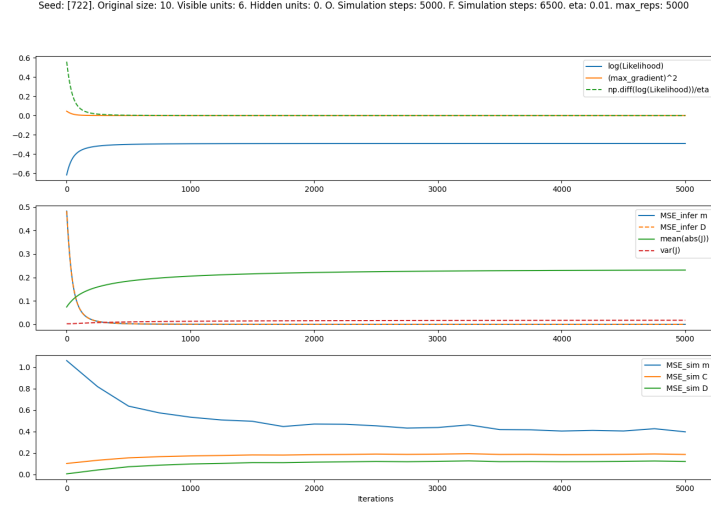


Figure 2: Seed: 722. Standard Ising model. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

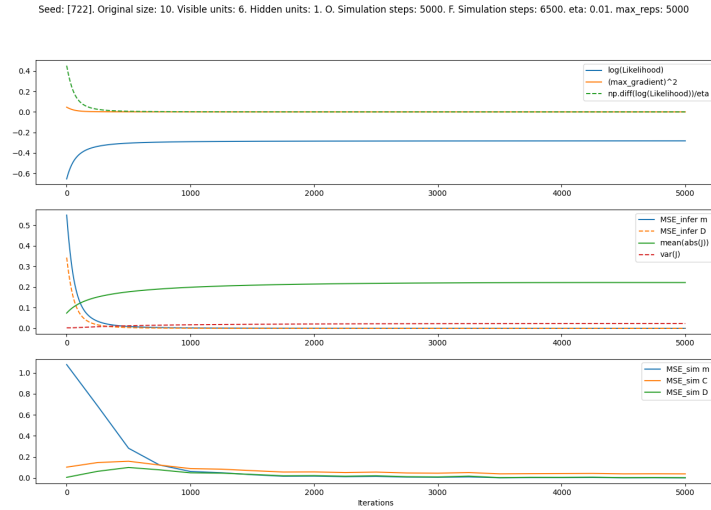


Figure 3: Seed: 722. Kinetic Ising model with 1 hidden unit. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

## A.2 Seed 1185

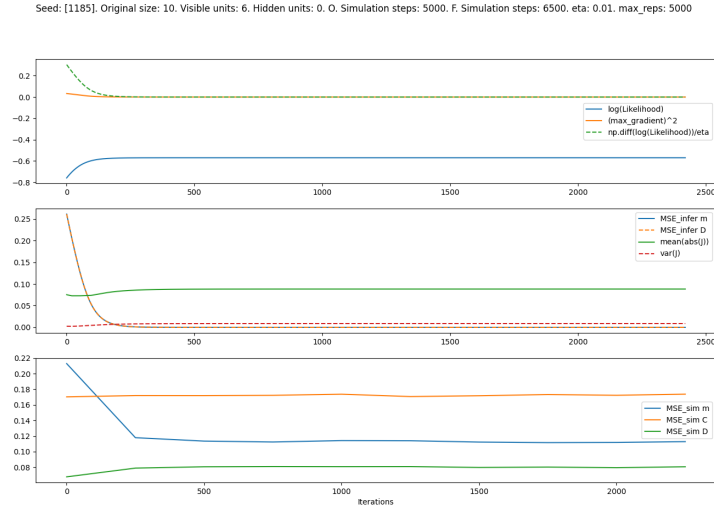


Figure 4: Seed: 1185. Standard Ising model. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

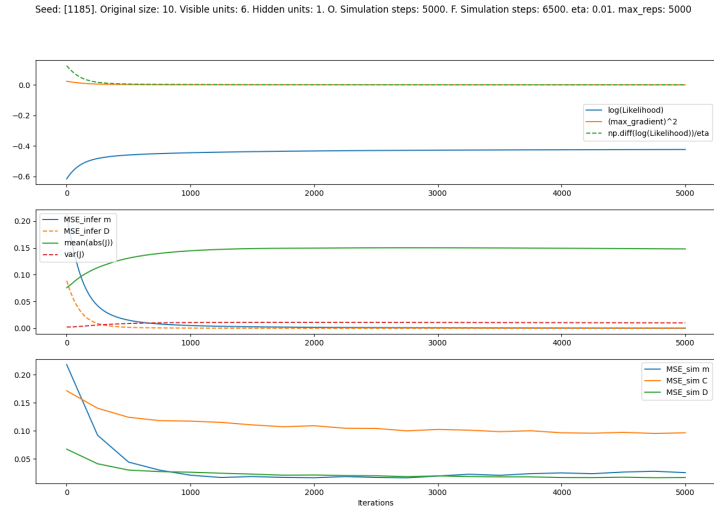


Figure 5: Seed: 1185. Kinetic Ising model with 1 hidden unit. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

### A.3 Seed 2178

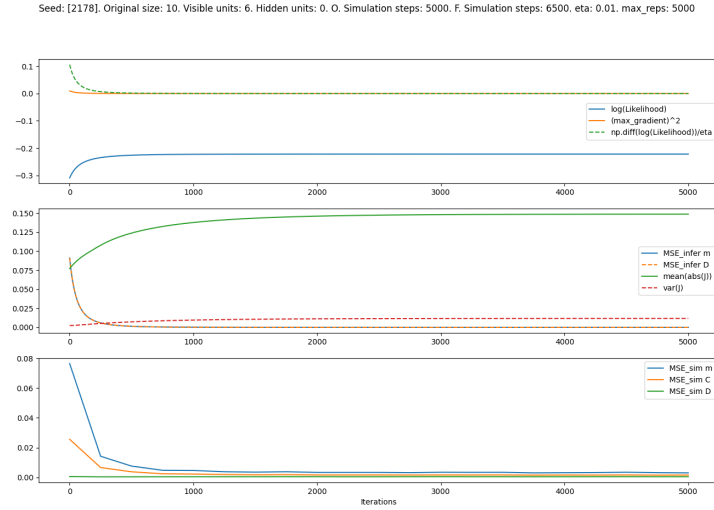


Figure 6: Seed: 2178. Standard Ising model. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

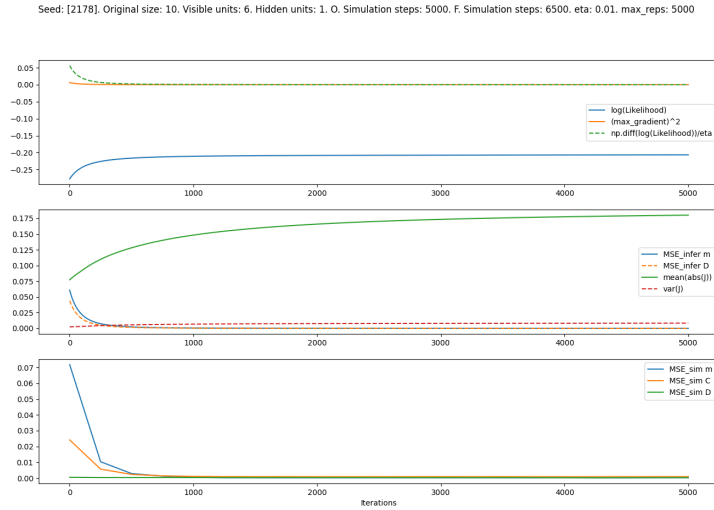


Figure 7: Seed: 2178. Kinetic Ising model with 1 hidden unit. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

## A.4 Seed 2692

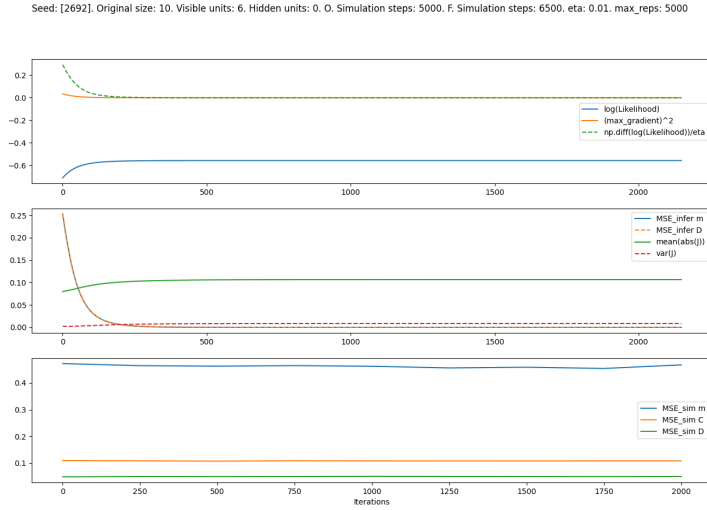


Figure 8: Seed: 2692. Standard Ising model. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

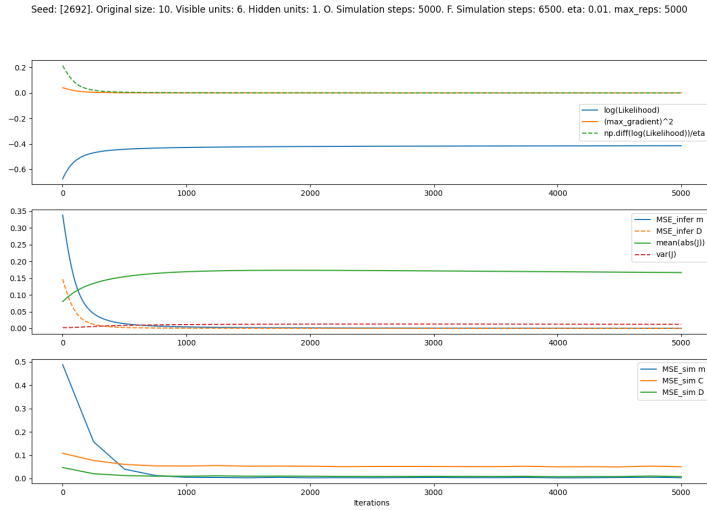


Figure 9: Seed: 2692. Kinetic Ising model with 1 hidden unit. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.



## A.5 Seed 3262

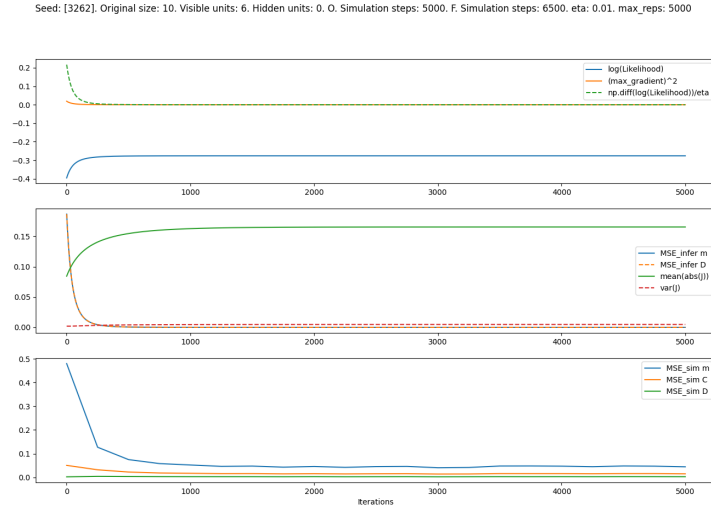


Figure 10: Seed: 3262. Standard Ising model. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of m and D during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of m, D and D after the simulations.

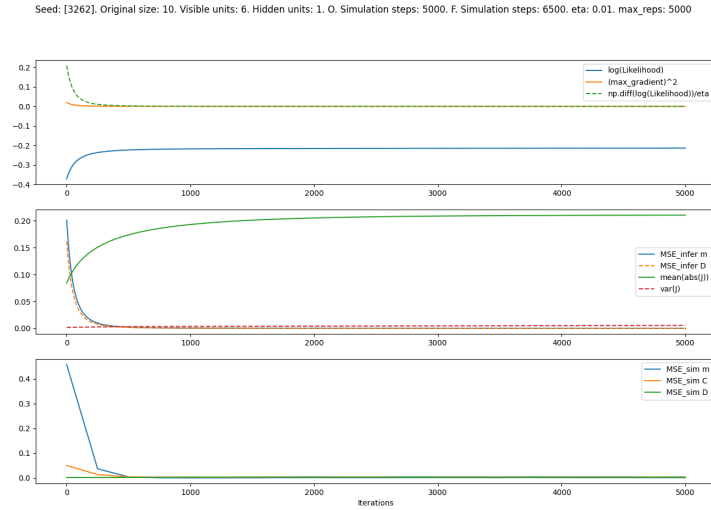


Figure 11: Seed: 3262. Kinetic Ising model with 1 hidden unit. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of m and D during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of m, D and D after the simulations.

## A.6 Seed 3813

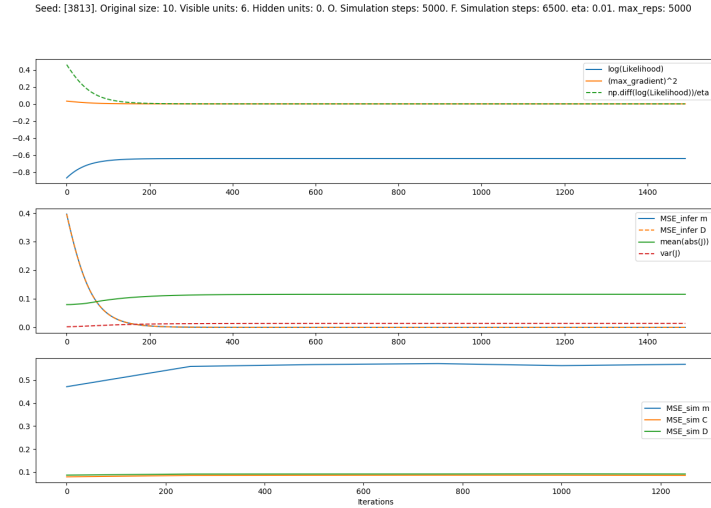


Figure 12: Seed: 3813. Standard Ising model. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

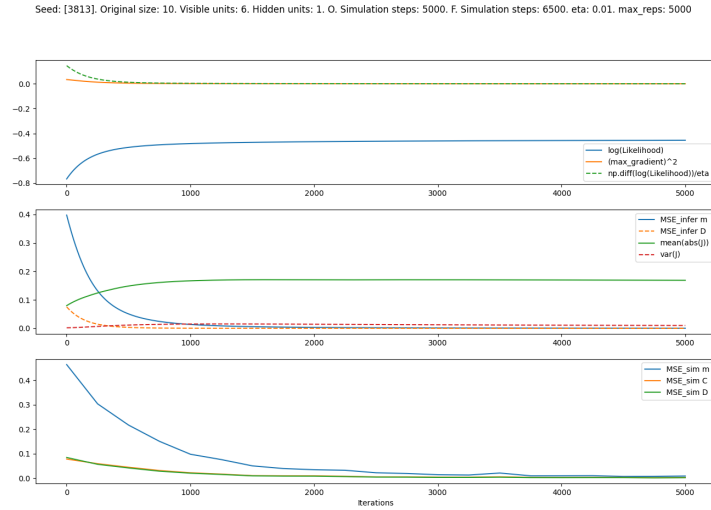


Figure 13: Seed: 3813. Kinetic Ising model with 1 hidden unit. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

## A.7 Seed 3988

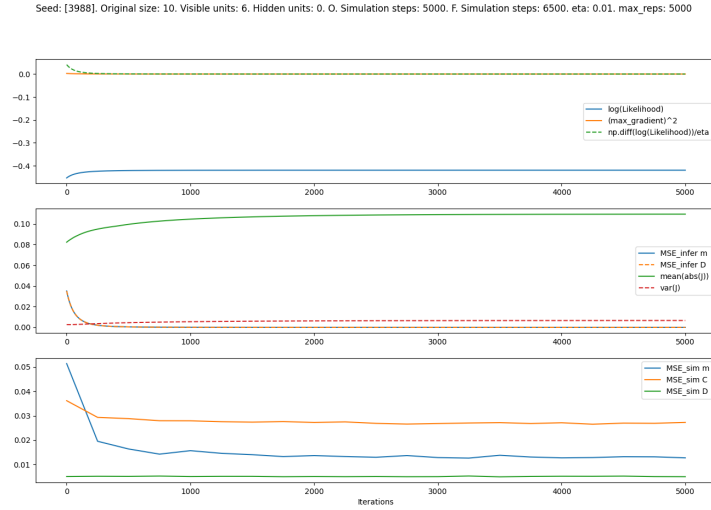


Figure 14: Seed: 3988. Standard Ising model. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

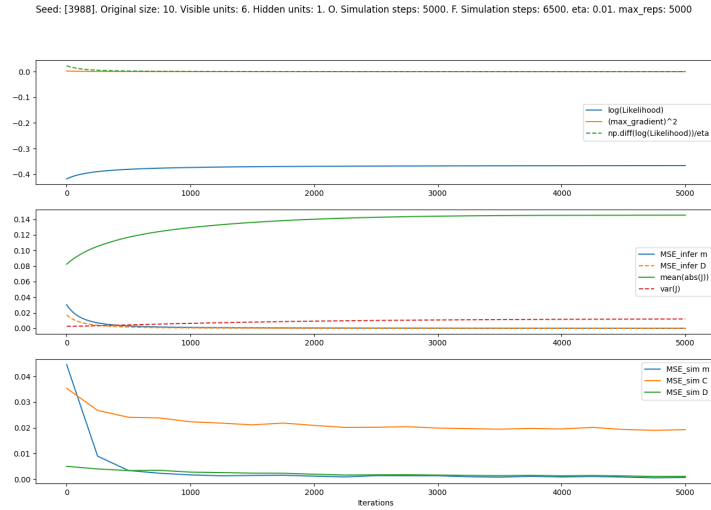


Figure 15: Seed: 3988. Kinetic Ising model with 1 hidden unit. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of  $m$  and  $D$  during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of  $m$ ,  $D$  and  $D$  after the simulations.

## A.8 Seed 4075

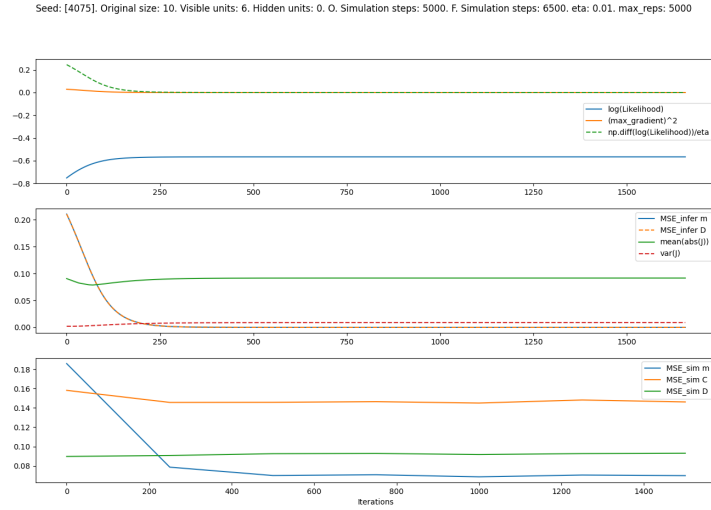


Figure 16: Seed: 4075. Standard Ising model. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of m and D during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of m, D and D after the simulations.

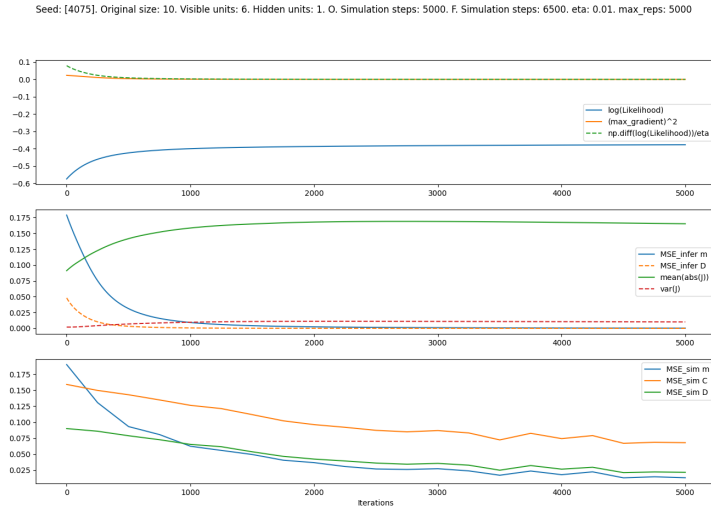


Figure 17: Seed: 4075. Kinetic Ising model with 1 hidden unit. Top: logarithm of the likelihood, square of the maximum gradient and difference of the likelihood between consecutive time-steps. Middle: MSE of m and D during inference and evolution of the behavior of  $\mathbf{J}$ . Bottom: MSE of m, D and D after the simulations.