BP for learning networks with hidden units

C. Battistin[2], Y. Roudi[1,2], J. Tyrcha[3], J. Hertz[1]

1. NORDITA, KTH Royal Institute of Technology and Stockholm University, Stockholm, Sweden
2. Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, NTNU, Trondheim, Norway
3. Department of Mathematics, Stockholm University, Stockholm, Sweden

NORDITA

Norwegian Centre of Excellence

NETADIS Statistical Physics Approaches to Networks Across Disciplines

THE KAVLI FOUNDATION

1. The model

Kinetic Ising model provides an useful platform for studying learning and inference in non-equilibrium models with potential applications to high-throughput data analysis.

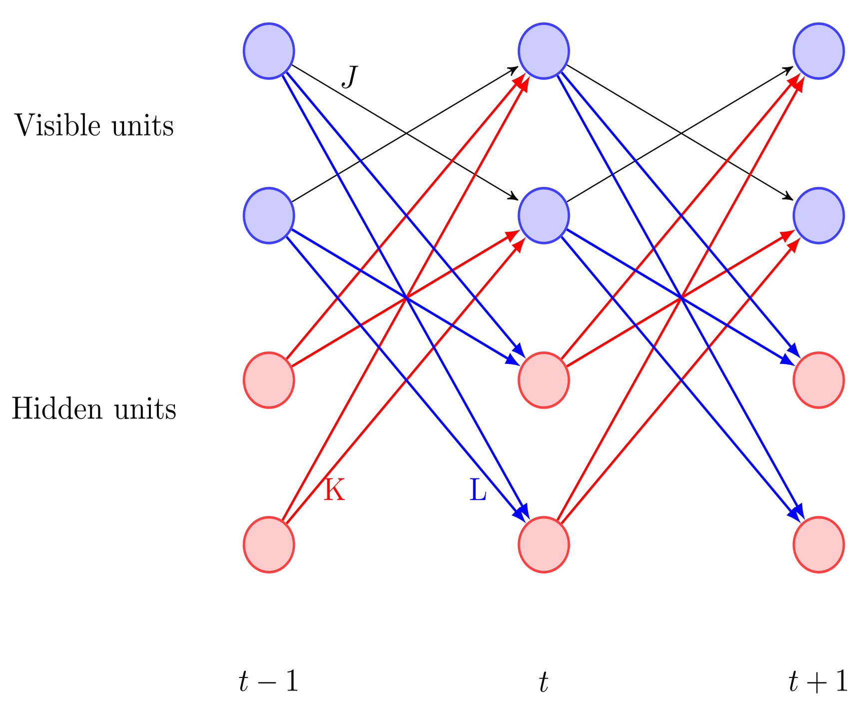
The aim of this work is to provide an efficient algorithm for inferring the statistics of hidden units and to study its performances in reconstructing partially observed networks with sparse connectivity.

In particular we considered a network of N_v observed and N_h hidden binary units, whose synchronous dynamics is defined by the transition probability:

$$\mathrm{P}[\mathbf{s},\boldsymbol{\sigma}] = \prod_t \frac{\exp\left[\sum_i s_i(t+1)h_i(t)\right]}{\prod_i 2\cosh\left[h_i(t)\right]} \times \frac{\exp\left[\sum_a \sigma_a(t+1)b_a(t)\right]}{\prod_a 2\cosh\left[b_a(t)\right]}$$

where $\sigma_a(t)$ ($a=1,\dots,N_h$) are the states of hidden units, $s_i(t)$ ($i=1,\dots,N_v$) are the visible states, while:

$$\begin{aligned} h_i(t) &= \sum_j J_{ij}s_j(t) + \sum_b K_{ib}\sigma_b(t) \\ b_a(t) &= \sum_j L_{aj}s_j(t) \end{aligned}$$



The likelihood of the observed time series **s** is factorizable in independent traces over σ(t):

$$\mathrm{P}_t[\mathbf{s}] = \sum_{\boldsymbol{\sigma}} \frac{\exp\left[\sum_i s_i^+\left(\sum_j J_{ij}s_j + \sum_b K_{ib}\sigma_b\right) + \sum_{a,j}\sigma_a L_{aj}s_j^-\right]}{\prod_i 2\cosh\left[\sum_j J_{ij}s_j + \sum_b K_{ib}\sigma_b\right]\prod_a 2\cosh\left[\sum_j L_{aj}s_j^-\right]}$$

Introducing the replicas

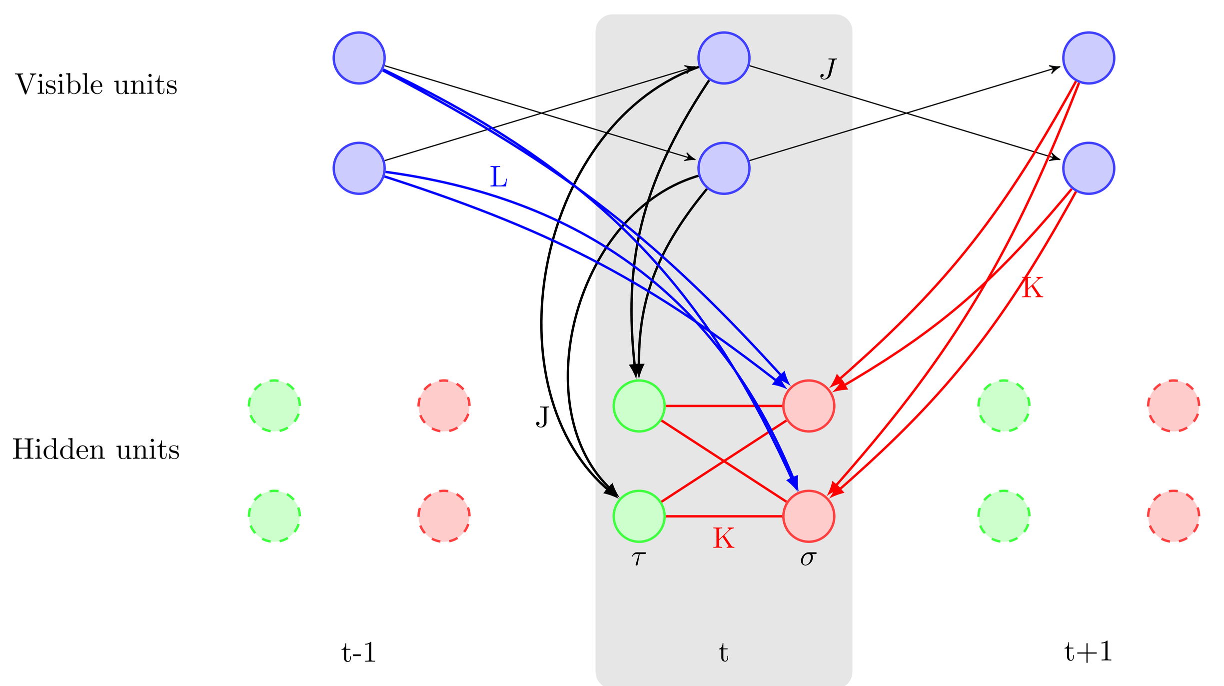
Using the relation:

$$(2\cosh[f_i])^n = \sum_{\tau}\exp\left[\sum_{\alpha=1}^{n}\sum_i \tau_i^\alpha f_i\right]$$

where:

$$f_i = \sum_j J_{ij}s_j + \sum_b K_{ib}\sigma_b$$

we rephase our problem in to an equilibrium one at each time step, at the cost of introducing $n \to -1$ replicas of N_v hidden units τ.



Under replica-symmetry assumption, gradient ascent learning provides us the following update rules for the couplings:

$$\begin{aligned} \Delta J_{ij} &\propto s_i^+ s_j - m_i s_j \\ \Delta K_{ib} &\propto s_i^+\mu_b - \langle \tau_i\sigma_b\rangle \\ \Delta L_{aj} &\propto \mu_a s_j^- - \tanh\left[\sum_b L_{bj}s_j^-\right]s_j^- \end{aligned}$$

where $m_i = \langle\tau_i\rangle$ and $\mu_a = \langle\sigma_a\rangle$.

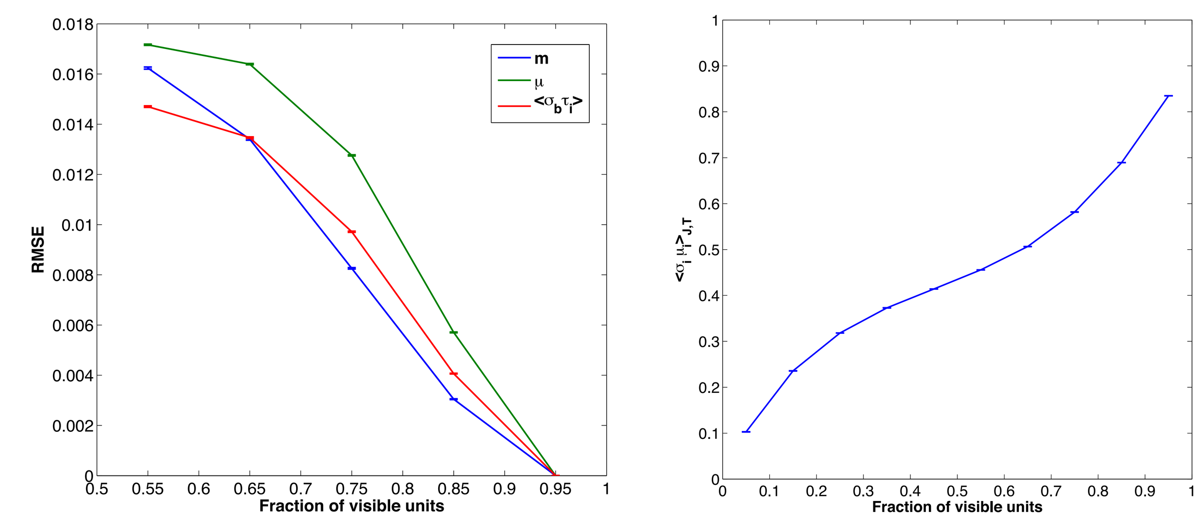
SusP equations:

$$\begin{aligned} \mu_a &= \tanh\left[\sum_j L_{aj}s_j^- + s_j^+K_{ja} - \tanh^{-1}\left[t_{ja}m_j^a\right]\right] \\ m_i &= \tanh\left[\sum_j J_{ij}s_j + \sum_b \tanh\left[t_{ib}\mu_b^i\right]\right] \\ \langle\tau_i\sigma_a\rangle &= \frac{\partial m_i}{\partial b_a^-} + m_i\mu_a \end{aligned}$$

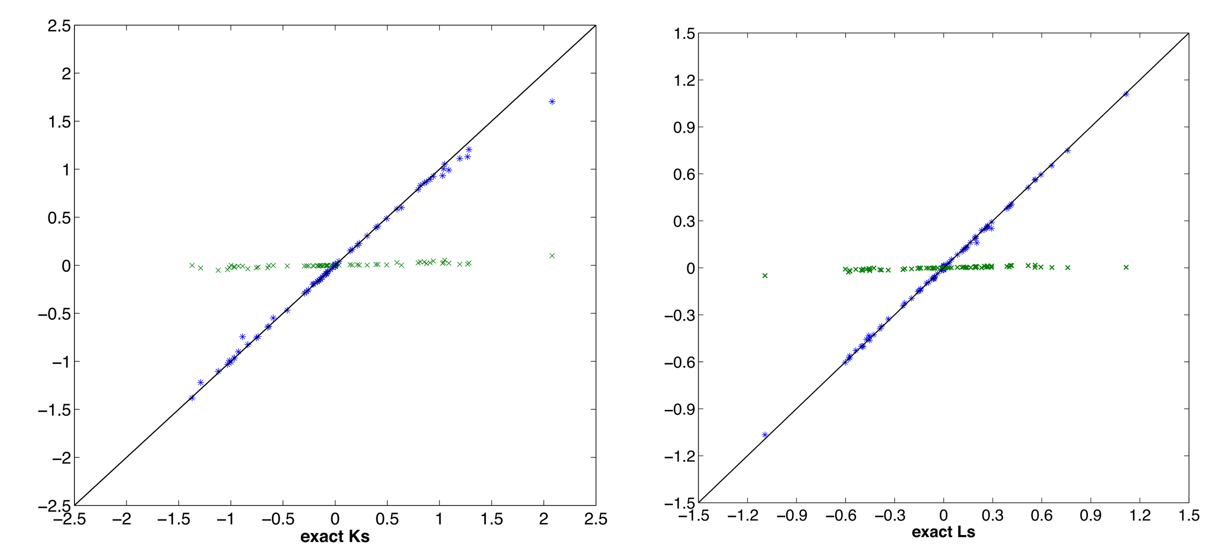
where $t_{ja} = \tanh[K_{ja}]$ and $b_a^- = \sum_j L_{aj}s_j^-$

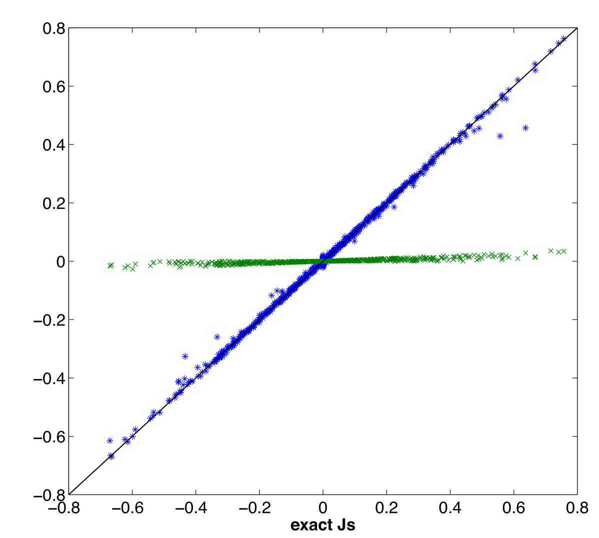
2. Learning and inference

Exponential drop of the RMSE between replica and message passing magnetizations and correlations in the fraction of visible units (left).
Inflection point at N_v/N=0.5 of percentage of correct inferred magnetizations signs (right).



N_v+N_h=20, T=10^2 time steps, 10^5 realizations of the couplings; sparsness(c)=0.1, g=1.



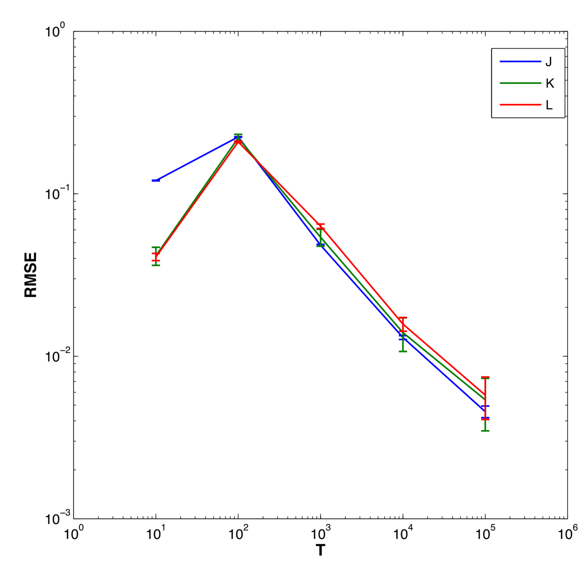


Strong agreement between reconstructed and true couplings.

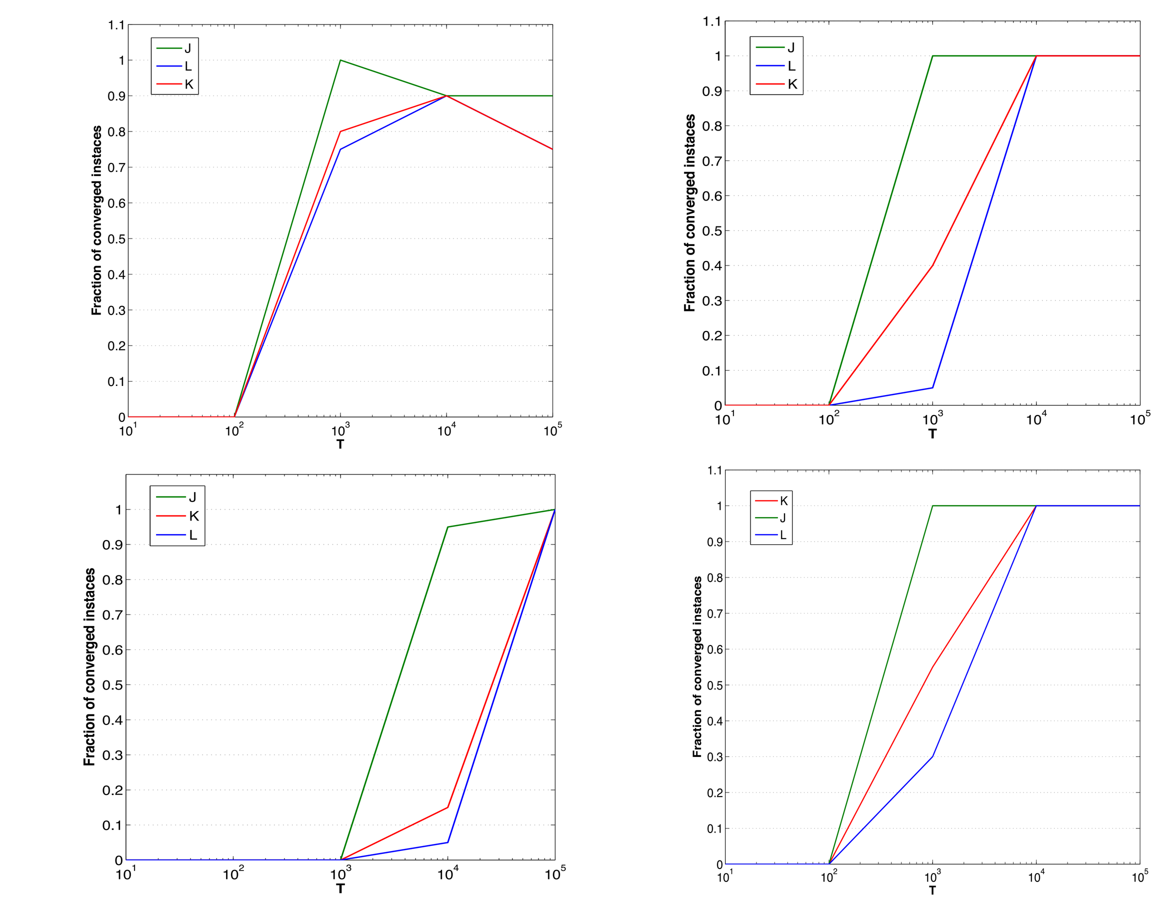
$$J \sim g/\sqrt{N_v c}$$
$$L \sim g/\sqrt{N_v c}$$
$$K \sim g/\sqrt{N_h c}$$

N_v=80, Nh=8 , g=1, c=0.1, T=10^5; green=initial value of the couplings; blue=reconstructed couplings;

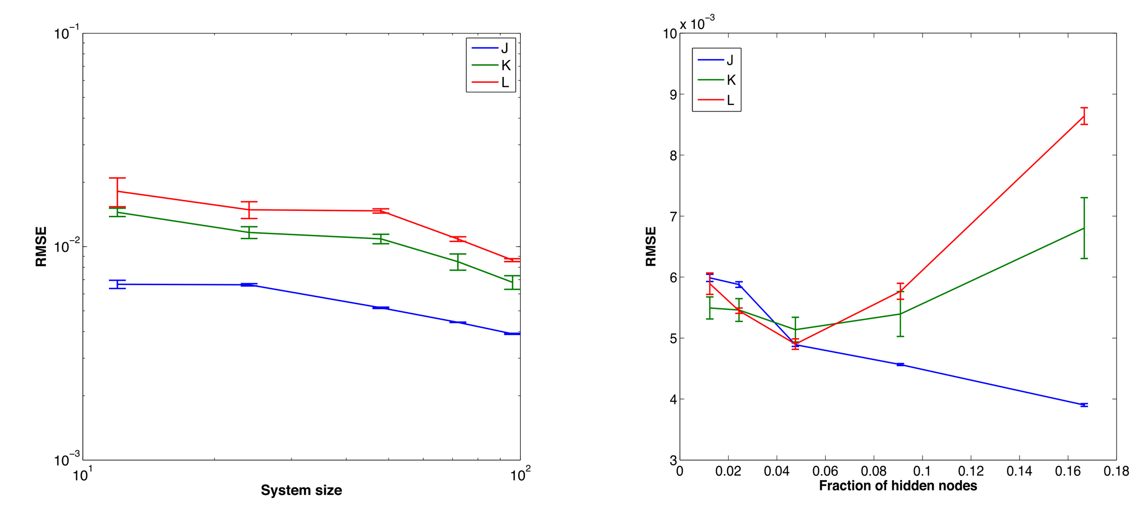
The normalized RMSE scales as $1/\sqrt{T}$ with the data length.



Strong connectivity in fully connected networks make reconstructed couplings diverge for long data sets.



N_v=20, N_h=4, 20 realizations of the couplings. Clockwise from top left: g=1, c=1.; g=0.5,c=1.; g=1., c=0.1; g=0.1,c=1.



g=1, c=0.1, T=10^5 ; N_h/N=0.25 (left), N_v=80 (right);

3. Conclusions and further developments

- The algorithm shows good performances on sparse networks with strong couplings. Analogous results have been obtained for binary connectivity.
- Regularization of the learning could prevent reconstructed couplings from exploding.
- Time consuming and unstable SusP two point function can be replaced with independent spins approximation.
- It would be interesting to establish how the RMSE scales with the fraction of visible units in the network.
- We will compare the performances of BP with mean field methods and use mean field magnetizations as starting values for the hidden magnetizations.

Supported by the Kavli Foundation and the Norwegian Research Council Centre of Excellence Grant