Inference of Kinetic Ising Model on Sparse Graphs

Pan Zhang

Abstract Based on dynamical cavity method, we propose an approach to the inference of kinetic Ising model, which asks to reconstruct couplings and external fields from given time-dependent output of original system. Our approach gives an exact result on tree graphs and a good approximation on sparse graphs, it can be seen as an extension of Belief Propagation inference of static Ising model to kinetic Ising model. While existing mean field methods to the kinetic Ising inference e.g., naïve mean-field, TAP equation and simply mean-field, use approximations which calculate magnetizations and correlations at time t from statistics of data at time $t - 1$, dynamical cavity method can use statistics of data at times earlier than $t - 1$ to capture more correlations at different time steps. Extensive numerical experiments show that our inference method is superior to existing mean-field approaches on diluted networks.

Keywords Inverse Ising model · Dynamical cavity method · Bethe approximation

1 introduction

Inference problem, which aims at learning internal structure of complex systems from empirical output data, has been studied for a long time. Among many models, inverse Ising model, which is a basic pair-wise model in physics, has been widely studied for inference problem in many fields.

A lot of study [1–4] has been carried out on the (static) inverse Ising problem, which asks to infer couplings and external fields of Ising model from configurations sampled from Boltzmann distribution. Efficient algorithms based on mean-field approximations and message passing have been developed to address the inference task. Recently it receives special attention in neural network reconstruction of retinas based on multi-electrode recordings [5], and gene-gene interaction reconstruction [2]. In static inverse Ising model, empirical data are M equilibrium configurations $\{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_M\}$ that sampled from Boltzmann distribution

P. Zhang (✉)
Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy
e-mail: pan.zhang@polito.it

$\mathcal{P} = \frac{1}{Z} e^{-H(\boldsymbol{\sigma})}$, where energy $H(\boldsymbol{\sigma})$ is a function of couplings and external fields. When M is very large, posterior distribution of H peaks on the maximum point so that one can find the couplings by maximizing the log-likelihood $\mathcal{L} = -\langle H \rangle - \log Z$, where $\langle H \rangle$ denotes averaged energy. Thanks to the convexity of log-likelihood, the exact method, so-called Boltzmann machine, computes exactly the correlations and magnetizations and match them with empirical ones. However, computing magnetizations and correlations exactly for large systems is unreachable, so many efforts have been paid to study approximate approaches based on different approximations, including naïve mean-field approximation, TAP equation, small correlation expansion and message passing schemes.

Boltzmann distribution is a strong restriction to system to be studied. Many real systems like biological systems, especially with time-dependent external stimuli, do not have this good property. *Inverse kinetic Ising model*, which asks to reconstruct couplings and time-dependent external fields from dynamic output of real system, obviously meets wider needs in real system inference. In inverse kinetic Ising model, exact reconstruction uses all empirical configurations in learning and time-complexity is proportional to the number of configurations. When number of configurations is large, one needs efficient approximate methods. In last years, Naïve Mean-field and TAP approaches have been adapted from static Ising model in case of weak couplings [6]. On densely connected fully asymmetric networks, an exact reconstruction method is proposed for asymmetric SK model [7, 8].

However, biological systems are often sparsely connected, so it should be valuable to adapt message passing [9] method which work well in sparse-connected static Ising model inference into kinetic Ising model inference. Another problem of existing mean-field methods is that they use approximations which predict quantities of time $t + 1$ based on time t, we term these approximations *one-step approximations*. There are two cases that one-step approximations works well. In first case, experimental data are drawn from stationary state where $t \to \infty$ and in second case there is no feedback loops in system e.g. the fully asymmetric model. But in practice it is difficult to ensure these two conditions, so it should be useful to develop a method that use statistics at not only one time step before, but earlier time steps to do the construction. In what follows, we show how to use Bethe approximation in inference of kinetic Ising model, in context of dynamic cavity method. This can be seen as an extension of Belief-Propagation inference in static inverse Ising problems [9].

The paper is organized as follows. Firstly, in Sect. 2, we give definition of kinetic Ising model. In Sect. 3 we introduce exact inference and existing mean-field approaches. In Sect. 4 we discuss dynamical cavity methods and inference of kinetic Ising model. Section 5 contains some numerical results to compare performance of inference by dynamical cavity method with other mean-field approximations. The last section contains conclusions and some discussions.

2 Kinetic Ising Model

We consider random graphs with N vertices and average connectivity C. In this paper we focus on sparse graphs that $C \ll N$. Connections of graphs are defined by matrix $c_{ij} = 0, 1$. $c_{ij} = 1$ means there is a directed edge from vertex i to vertex j and $c_{ij} = 0$ means there is no such edge. If $c_{ij} = c_{ji} = 1$, the edge between i and j is symmetric, otherwise it is asymmetric. We set each edge to be symmetric with probability P_s and asymmetric with probability $1 - P_s$. Obviously, with $P_s = 1$, graph is undirected and with $P_s = 0$, graph is fully asymmetric.

On each vertex i there is a spin $\sigma_i = \{+1, -1\}$ associated, and on each edge $j \to i$ there is a coupling J_{ij} associated. If $c_{ji} = 0$, coupling J_{ij} is set to 0 otherwise value of

J_{ij} is generated randomly according to Gaussian distribution with zero mean and variance $J/\sqrt{C}$. Dynamics of spins are defined by configurations at different discrete time steps from time 0 to time T: $\underline{\sigma}(0), \ldots, \underline{\sigma}(T)$. At each time step, value of spin i is updated based on configuration of last time according to

$$W\big[\sigma_i(t+1)\big|\boldsymbol{\sigma}(t)\big] = \frac{\exp[\beta\sigma_i(t+1)h_i(t)]}{2\cosh\beta h_i(t)}, \tag{1}$$

where β denotes the inverse temperature and local field at time t is expressed as $h_i(t) = \sum_{j\in\partial i} J_{ij}\sigma_j(t) + \theta_i(t)$, with $\theta_i(t)$ denoting external fields acting on spin i at time t. We consider in this paper the parallel dynamics, that all spins updated synchronously, evolution of configuration is written as

$$W\big[\boldsymbol{\sigma}(t+1)\big|\boldsymbol{\sigma}(t)\big] = \prod_{i=1}^{N} W\big[\sigma_i(t+1)\big|\boldsymbol{\sigma}(t)\big]. \tag{2}$$

We argue that it would be not difficult to extend our approach from parallel dynamics to sequential dynamics following what has been done for mean-field methods in [10].

Direct problem of kinetic Ising model asks to predict dynamical behavior of spins at an arbitrary time t, given couplings, external fields and initial state of network. Calculating full distribution of system $P(\boldsymbol{\sigma}(t))$ is a very difficult task, so most theories focused on computing macroscopic observables e.g. magnetization at a time $m(t) = \sum_{\sigma_i(t)} \sigma_i(t)p(\sigma_i(t))$ and correlation at different times $c_{ij}(t, t') = \sum_{\sigma_i(t),\sigma_j(t')} \sigma_i(t)\sigma_j(t')p(\sigma_i(t), \sigma_j(t'))$. This direct problem have been studied for long time especially in the context of attractor neural networks [11, 12]. Many approaches has been proposed to study ensemble averaged macroscopic quantities, e.g. dynamical replica method [13], generating functional method [14] and dynamical cavity method [15].

Inverse problem of kinetic Ising model is the problem we would like to study in this paper. Unlike direct problem, instead of couplings and external fields, experimental output of original model (set of configurations) is given, one is asked to infer couplings and external fields from those experimental data. The experimental data are obtained by running parallel dynamics of Ising model according to Eq. (1) on graphs for R realizations of time length T paths (trajectories), denoted by $\{\sigma_i^r(t)\}$ with $r \in [1, R]$ and $t \in [1, T]$. In this paper, we set T to 10.

3 Exact Reconstruction and Mean-Field Approximations

Given experimental configurations, probability of observing these samplings as a function of couplings is written as:

$$
\begin{aligned}
P\big(\boldsymbol{J}_{ij}; \boldsymbol{\theta}_i(t)\big) &= \prod_{r=1}^{R}\prod_{t=1}^{T-1} W\big[\boldsymbol{\sigma}(t+1)\big|\boldsymbol{\sigma}(t)\big] \\
&= \prod_{r=1}^{R}\prod_{t=1}^{T-1}\prod_{i=1}^{N} \exp[\beta\sigma_i^r(t+1)h_i^r(t) - \log 2\cosh\beta h_i^r(t)].
\end{aligned} \tag{3}
$$

Then, log-likelihood of observing data is defined as:

$$\mathcal{L}\big(\boldsymbol{J}_{ij};\boldsymbol{\theta}_i(t)\big) = \log\big(P\big(\boldsymbol{J}_{ij};\boldsymbol{\theta}_i(t)\big)\big)$$

$$= \sum_{r=1}^{R}\sum_{t=1}^{T}\left[\beta\sum_{i=1}^{N}\sigma_i^r(t+1)\left(\sum_{j\in\partial i}J_{ij}\sigma_j^r(t)+\theta_i(t)\right)\right.$$

$$\left.-\sum_{i=1}^{N}\log 2\cosh\beta\left(\sum_{j\in\partial i}J_{ij}\sigma_j^r(t)+\theta_i(t)\right)\right]. \tag{4}$$

To find the most-likely couplings and external fields, one needs to maximize the log-likelihood. At the maximum point of log-likelihood, by setting derivatives of $\mathcal{L}$ with respect to J_{ij} and θ_i to zero, one has following equations:

$$\big\langle\sigma_i^r(t+1)\big\rangle_R = \left\langle\tanh\beta\left(\sum_{k\in\partial i}J_{ik}\sigma_k^r(t)+\theta_i(t)\right)\right\rangle_R, \tag{5}$$

$$\big\langle\big\langle\sigma_i^r(t+1)\sigma_j^r(t)\big\rangle_T\big\rangle_R = \left\langle\left\langle\tanh\left[\beta\left(\sum_{k\in\partial i}J_{ik}\sigma_k^r(t)+\theta_i(t)\right)\right]\sigma_j^r(t)\right\rangle_T\right\rangle_R, \tag{6}$$

where $\langle\cdot\rangle_T$ denotes taking average over time and $\langle\cdot\rangle_R$ denotes taking average over realizations.

Let us use $m_i^{data}(t)=\langle\sigma_i^r(t)\rangle_R$ and $c_{ij}^{data}(t,t-1)=\langle\langle\sigma_i^r(t)\sigma_j^r(t)-1\rangle_T\rangle_R$ to denote experimental magnetization at time t and correlation at time t and $t-1$ respectively, and use $m_i^J(t)$ and $c_{ij}^J(t+1,t)$ to denote magnetization and correlations predicted by Ising model. Equations (5), (6) show that at max-likelihood point, $m_i^J(t+1)$ and $c_{ij}^J(t+1,t)$ are calculated using configurations at time t, so we term it *configurations based one-step reconstruction*. The reconstruction can be carried out by using gradient descent learning starting from an initial couplings and fields:

$$\theta_i = \theta_i + \eta\big(\big\langle m_i^{data}(t)\big\rangle_t - \big\langle m_i^J(t)\big\rangle_t\big), \tag{7}$$

$$J_{ij} = J_{ij} + \eta\big(\big\langle c_{ij}^{data}(t+1,t)\big\rangle_t - \big\langle c_{ij}^J(t+1,t)\big\rangle_t\big) \tag{8}$$

where constant η is learning rate. Note that in above scheme, one has to scan all configurations to compute the average on every learning step. It is time consuming and not realistic when amount of experimental data is huge. To overcome this computational problem, several mean-field approaches, e.g. Naïve mean-field, TAP and simply mean-field method, are proposed. They use different approximations to compute macroscopic observables, say average magnetizations and correlations, and use those macroscopic observables, instead of all configurations, to do inference.

Naïve mean-field method [6] applies naïve mean-field approximation

$$m_i^J(t+1) = \tanh\left[\beta\sum_{j\in\partial i}J_{ij}m_j(t)\right], \tag{9}$$

and compute one-step delayed correlation as

$$c_{ij}^J(t+1,t) = m_j(t)m_i^J(t+1) + \big(1-\big(m_i^J(t+1)\big)^2\big)\sum_k J_{ik}\big[c_{kj}(t,t)-m_k(t)m_j(t)\big]. \tag{10}$$

As an extension of naïve mean-field, TAP method takes the expansion on couplings and outperforms naïve TAP method in case of weak couplings, the approximation is written as:

$$m_i^J(t+1) = \tanh\left[\beta\sum_{j\in\partial i}J_{ij}m_j(t) - m_i^J(t+1)\sum_j J_{ij}^2\big(1-m_j^2(t)\big)\right], \tag{11}$$

and delayed correlation is expressed as

$$
\begin{aligned}
c_{ij}^J(t+1,t) = {} & m_j(t)m_i^J(t+1) + \left[1 - \left(m_i^J(t+1)\right)^2\right] \\
& \times \beta\left[1 - \left[1 - \left(m_i^J(t+1)\right)^2\right]\sum_k J_{ij}^2(1-m_k^2(t))\right] \\
& \times \sum_k J_{ik}\left[c_{kj}(t,t) - m_k(t)m_j(t)\right].
\end{aligned} \tag{12}
$$

Another recently proposed approach, simply Mean-field method [7, 8], which assumes Gaussian distribution of the local field of each spin, computes magnetization as

$$
m_i^J(t+1) = \int \frac{dx}{2\pi} e^{x^2/2} \tanh\beta\left(\sum_{j\in\partial a} J_{ij}m_j(t) + x\sqrt{\sum_j J_{ij}^2(1-m_j^2(t))}\right), \tag{13}
$$

and delayed correlation as

$$
\begin{aligned}
c_{ij}^J(t+1,t) = {} & m_j(t)m_i^J(t+1) \\
& + \int \frac{dx}{2\pi} e^{x^2/2}\left[1 - \tanh^2\beta\left(\sum_{j\in\partial a} J_{ij}m_j(t) + x\sqrt{\sum_j J_{ij}^2(1-m_j^2(t))}\right)\right] \\
& \times \sum_k J_{ik}\left[c_{kj}(t,t) - m_k(t)m_j(t)\right].
\end{aligned} \tag{14}
$$

In above listed mean-field approximations, predictions of $m_i^J(t+1)$ and $c_{ij}^J(t+1,t)$ are made by using magnetizations and correlations at time t, this kind of reconstruction can be termed *statistics based one-step reconstruction*. In some special cases, e.g. fully asymmetric networks, observables at time t are sufficient to predict observables at time $t+1$ since effect of feedback loops are not present. But In general cases, to predict observables at time $t+1$ one needs to use not only observables at time t but also at earlier time steps up to $t=0$.

4 Dynamical Cavity Reconstruction

Dynamical cavity method, originally proposed in [15], together with dynamical replica analysis and generating functional analyses, are powerful tools in analyzing dynamics of networks. The advantage of dynamical cavity method with respect to other two methods is that it can be applied on single instances to compute observables for every node.

From Eq. (2), we can write the probability of observing a path as:

$$
P\left(\boldsymbol{\sigma}^{[0,T]}\right) = \prod_{t=0}^T W\left[\boldsymbol{\sigma}(t+1)\big|\boldsymbol{\sigma}(t)\right]P\left(\boldsymbol{\sigma}(0)\right), \tag{15}
$$

where $P(\boldsymbol{\sigma}(0))$ denotes initial probability distribution of spin configuration at time 0. It is difficult to study directly the full path distribution, so dynamical cavity method focuses on marginal probability of a path that spin i evolves from time 0 to time t conditioning on the local fields $\theta_i^{[0,t]}$:

$$
P_i\left(\sigma_i^{[0,t]}\big|\theta_i^{[0,t]}\right) = \sum_{\substack{\sigma_j^{[0,T]} \\ j\neq i}} P\left(\{\sigma_i^{[0,t]}; \sigma_j^{[0,T]}\}\big|\boldsymbol{\theta}^{[0,t]}\right). \tag{16}
$$

By applying Bethe approximation, one can derive following iterative cavity equations (we refer to literature [15] for details of derivations):

$$P_{i \to k}\big(\sigma_i^{[0,t]}\big|\theta_i^{[0,t-1]} + J_{ki}\sigma_k^{[0,t-1]}\big) = \sum_{\sigma_j^{[0,t-1]}} \prod_{j \in \partial i \setminus k} P_{j \to i}\big(\sigma_j^{[0,t-1]}\big|\theta_j^{[0,t-1]} + J_{ji}\sigma_i^{[0,t-1]}\big)$$

$$\times \prod_{t'=1}^{t-1} \frac{e^{\beta\sigma_i^{t'}[\sum_{j \in \partial i} J_{ij}\sigma_j^{t'-1} + \theta_i^{t'-1}]}}{2\cosh\beta[\sum_{j \in \partial i} J_{ij}\sigma_j^{t'-1} + \theta_i^{t'-1}]} P_i\big(\sigma_i^0\big),$$

(17)

where $P_{i \to k}\big(\sigma_i^{[0,t]}\big|\theta_i^{[0,t-1]} + J_{ki}\sigma_k^{[0,t-1]}\big)$ is cavity message denoting (cavity) marginal probability of a path that spin i evolves from time 0 to time t with its neighbor spin k removed from the graph.

Above equation is exact on tree graphs and a good approximation on sparse graphs. After a fixed point of cavity equation is reached, observables e.g. marginal probabilities, magnetizations and correlations can be computed as functions of cavity messages. However, solving Eq. (17) is very time-consuming for large t because computational complexity is proportional to 2^{2t}. So in practice one needs approximations to reduce computational complexity. Here we adopt the most simple approximation, one-time approximation [15], which is also named time-factorization approximation [16]:

$$P_{i \to k}\big(\sigma_i^{[0,t]}\big|J_{ki}\sigma_k^{[0,t-1]}\big) = \prod_{t'=0}^{t} P_{i \to k}\big(\sigma_i^{t'}\big|J_{ki}\sigma_k^{t'-1}\big). \tag{18}$$

This approximation ignores correlations among different time steps before time $t-2$, and makes summation over variables before time $t-2$ possible. Then, from Eq. (17), one arrives at

$$P_{i \to k}\big(\sigma_i^k\big|J_{ki}\sigma_k^{t-1}\big) = \sum_{\sigma_i^{t-2}} \sum_{\sigma_j^{t-1}} \prod_{j \in \partial i \setminus k} P_{j \to i}\big(\sigma_j^{t-1}\big|J_{ji}\sigma_i^{t-2}\big) W\Big[\sigma_i^t\Big|\sum_j J_{ij}\sigma_j^{t-1}\Big] P_i\big(\sigma_i^{t-2}\big),$$

(19)

and expression of magnetizations and delayed correlations can be derived from above equation

$$m_i^J(t+1) = \sum_{\sigma_i(t+1)} \sum_{\sigma_i^{t-1}} \sum_{\sigma_j^t} \prod_{j \in \partial i} P_{j \to i}\big(\sigma_j^t\big|J_{ji}\sigma_i^{t-1}\big) W\Big[\sigma_i^t\Big|\sum_j J_{ij}\sigma_j^t\Big] P_i\big(\sigma_i^{t-1}\big)\sigma_i(t+1),$$

(20)

$$c_{ij}^J(t+1,t) = \sum_{\sigma_k^t} \sum_{\sigma_j^t} \prod_{k \in \partial i} P_{k \to i}\big(\sigma_k^t\big|J_{ki}\sigma_i^{t-1}\big) P_{j \to i}\big(\sigma_j^t\big|J_{ji}\sigma_i^{t-1}\big) \tanh\Big(\beta\sum_{k \in \partial i} J_{ik}\sigma_k^t\Big)\sigma_j^t.$$

(21)

With $m_i^J(t+1)$ and $c_{ij}^J(t+1,t)$ obtained, a standard procedure to perform inference is using gradient descent to learn couplings from differences between experimental and predicted correlations, and learn external fields from differences between experimental and predicted correlations.

$$\theta_i^{Ising} = \theta_i^{Ising} + \eta\Delta m_i, \tag{22}$$

$$J_{ij}^{Ising} = J_{ij}^{Ising} + \eta\Delta C_{ij}, \tag{23}$$

Fig. 1 Evolution of inference error ΔJ and difference of one-step delayed correlations of dynamical cavity inference as function of iterating time for networks with $N = 100$, $C = 5.0$, $P_s = 0$ and different coupling strength. Number of realizations used in experimental data R is 5×10^4. Networks are fully asymmetric with $J = 0.2$, 0.9 and 2.0 in *top*, *middle* and *bottom* panel respectively
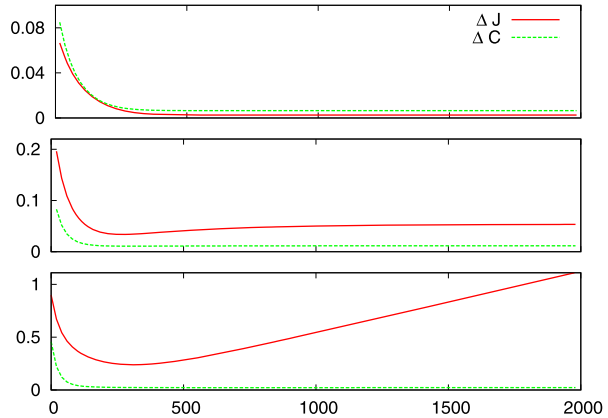
where

$$\Delta m_i = \frac{1}{T} \sum_{t=1}^{T} \left[m_i^{data}(t) - m_i^J(t) \right], \tag{24}$$

$$\Delta C_{ij} = \frac{1}{T-1} \sum_{t=2}^{T} \left[C_{ij}^{data}(t, t-1) - C_{ij}^J(t, t-1) \right]. \tag{25}$$

Note that there are two ways to do the inference using magnetizations and delayed correlations, one way as used in this paper is to learn couplings by gradient descent, and the other one is to solve directly the couplings by matching exactly the experimental data, which usually requires inversion of a correlation-based matrix [6, 7]. Usually gradient descent learning works slower than matrix inversion method and the convergence of learning method may depend on initial couplings and learning rate. However, in matrix inversion method, sometimes it is difficult to find a set of couplings and external fields that exactly matches the noisy experimental data, while the learning method is able to converge to a fixed point close to the true experimental data (as shown in Fig. 1).

5 Performance and Comparison with Mean-Field Methods

In this section we make comparative analysis between dynamical cavity method and mean-field methods on inverse kinetic Ising problem. Before we compare performance of those methods, the first thing we are interested in is the convergence properties of gradient descent learning used in our scheme. In Fig. 1 we plot the evolution of average inference error at each time step

$$\Delta J = \frac{1}{\sqrt{NC}} \sqrt{\sum_{\langle ij \rangle} \left(J_{ij}^{Ising} - J_{ij}^{true} \right)^2} \tag{26}$$

and average difference of one-step delayed correlation

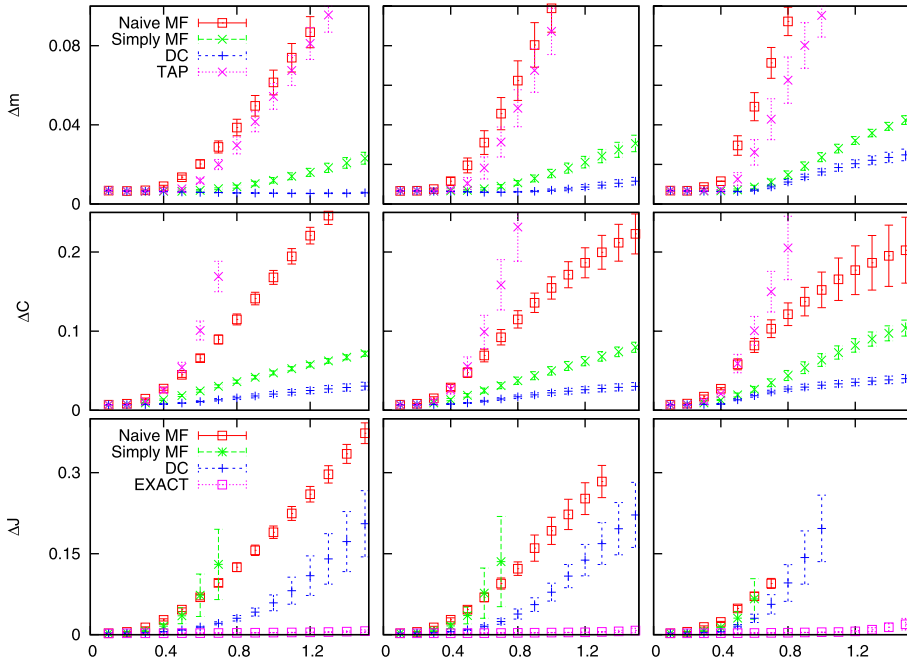$$\Delta C = \frac{1}{\sqrt{NC}} \sqrt{\sum_{\langle ij \rangle} \left(C_{ij}^J - C_{ij}^{data} \right)^2} \tag{27}$$

Fig. 2 Difference of magnetizations Δm, one-step delayed correlations ΔC and inference error ΔJ given by naïve mean-field, TAP, simply mean-field and dynamical cavity (DC) method of Ising model with Gaussian distributed couplings with zero mean and variance J. X *axes* denote coupling strength J. Size of the graph is $N = 100$, average degree is 5. Fraction of symmetric edges $P_s = 0$ in *left 3 figures*, $P_s = 0.5$ in *middle 3 figures* and in the *right ones* $P_s = 1.0$. Data are averaged over 10 realizations. Some data of inference errors are not present because of un-convergence of gradient descent learning

given by dynamical inference in the gradient descent process. In Fig. 1 we recorded evolution of ΔJ and ΔC in three gradient descent process on the same network with different coupling strengths. We can see that in three panels of Fig. 1, all correlation differences converge in the learning, but evolutions of ΔJ are much different. In case of weak couplings (e.g. $J = 0.2$ at left panel), ΔJ decreases monotonously with difference of correlation till converges to the error level characterized by noise in experimental data. In case of $J = 0.9$, ΔJ decreases to a minimum value then start increasing and finally converges to a point larger than the minimum value. When J increases to 2.0, ΔJ keeps increasing after reaching the minimum point and finally goes beyond the initial error and diverges. Coupling strength plays a role of inverse temperature, a larger coupling strength gives equivalently low temperature in the model and dynamics of the model becomes more difficult to predict due to stronger correlations. Consequently, approximations we made in dynamical method turns worse with stronger couplings and results to worse convergence of gradient descent learning and larger inference error.

We have compared convergence properties of gradient descent learning with other approximations on the same network (data not shown), results show that with naïve mean-field, TAP and simply mean-field approximations, ΔJ stops convergence with couplings weaker than 0.9 because they give even worse approximations than dynamical cavity method on larger couplings (see also Fig. 2). However exact reconstruction does not have this converging problem as it does not use approximations so inference error only depends on quality of

data (e.g. number of realizations used in computing experimental data), and is not a function of coupling strength.

To evaluate performance of approximations, especially how it is influenced by coupling strength, one method is to compare accuracy of approximations in direct problem, e.g., by applying approximations on graphs with fixed couplings and external fields to compute magnetization and correlation and compare result with experimental data. Error in correlations are characterized by Eq. (27) and now ΔC is not a function of time t since couplings are fixed. Error of magnetization is defined in the same way by difference of magnetization given by approximations and that in the experimental data:

$$\Delta m = \frac{1}{\sqrt{N}} \sqrt{\sum_{\langle i \rangle} \left(m_i^J - m_i^{data} \right)^2}. \tag{28}$$

Result are plotted in Fig. 2, where graphs have same number of nodes and average connectivity but different degree of symmetry. Note that number of samplings used in computing experiment data is large enough, and noise in experimental data can be ignored compared with error made in approximations, so a smaller Δm and ΔJ indicates a lower error made by the approximation. Top and middle panels of Fig. 2 show that in direct problem, all approximations work better in asymmetric networks than in symmetric networks. This is because in asymmetric networks, feedback correlations are not present so correlation ignored by approximations are fewer than in symmetric networks. Among all method, dynamical method works the best in this direct problem, which indicates that Bethe approximations is more accurate than mean-field approximations in diluted networks. More over, TAP method outperforms naïve mean-field only with weak couplings, because the expansion is carried out with weak couplings, and it is more difficult to find a solution in Eq. (11) with stronger couplings.

To evaluate performance of inference based on these approximations, we computed inference error ΔJ given in Eq. (26) by different approximations on same set of graphs, results are plotted in bottom panel of Fig. 2. TAP works badly and stops converging on networks with very weak coupling strength, so I did not plot the inference error of TAP. The missing points in the figure indicate that with value of that coupling strength, gradient descent learning does not converge. Figures show that all methods perform worse when degree of symmetry increases, as in the direct problems. Dynamical cavity method converges with larger J and gives smaller inference error at same J than Naïve mean-field and simply mean-field. Simply mean-field outperforms naïve mean-field only with weak couplings and stops converging with weaker couplings. One interesting point is that, with degree of symmetry of network increases, the difference of ΔC and ΔJ between dynamical cavity method and simply mean-field becomes larger while the difference between naïve mean-field and dynamical cavity becomes smaller.

We saw from Figs. 1 and 2 that, performances of dynamical cavity method is much worse than exact reconstruction. This is because due to computational expenses, networks we tested here are rather small, so loop effect can not be ignored on these networks (see e.g. [17]). If we increase the sparsity of the graph, the error of dynamical cavity method should be decreased. To illustrate at this point, in Fig. 3 we plot inference errors of dynamical cavity method for networks with same average connectivity but different system sizes. The figure shows that with N increasing, inference error approaches the error of exact inference. We expect that with $N \to \infty$, inference error of dynamical cavity method gives same result to that of exact reconstruction in fully asymmetric networks. Note that in contrast to dynamical cavity method, mean-field method does not benefits from increasing system size because

Fig. 3 Inference error of dynamical cavity (DC) inference, naïve mean-field and exact method for networks with different system size N. Average connectivity of network is 5, $P_s = 0$, $J = 1.5$ and data are averaged over 10 realizations
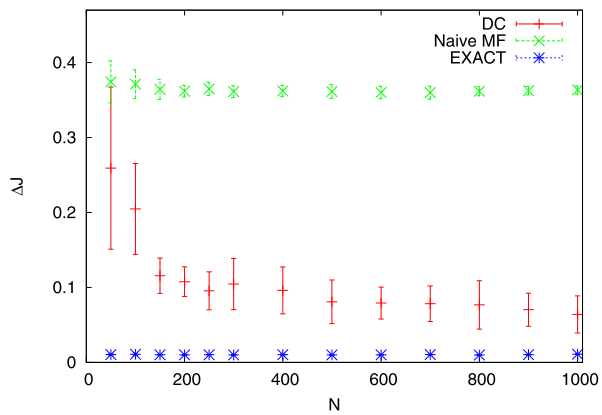
mean-field approximation does not become better with sparsity increases, and as a consequence, performance between dynamical cavity method and mean-field methods becomes larger for sparse networks with larger system sizes.

6 Conclusion and Discussion

In this paper we presented how to use dynamical cavity method to infer kinetic Ising model, and compared the performance on sparse graphs with existing mean-field methods. Results show that in sparse graphs dynamical cavity method works better than other mean-field approximations and performance increases with sparsity of graph increases. In computing cavity marginals, we use the simplest one-time approximation to simplify the cavity equations, and statistical quantities at time $t + 1$ are computed from quantities at time t and time $t - 1$, which is different from mean-field approximations which do predictions by using only quantities at time t. We believe that dynamical cavity inference could benefit a lot from developing more efficient approximations in computing cavity equations to collect correlations between more time steps. The networks discussed in this paper have known topology. It is difficult to apply dynamical cavity method to networks with unknown structure since it is so expensive to run cavity equations on fully-connected networks. For inference problems which asks to reconstruct both network topology and coupling strengths, one can use another method e.g. simply mean-field to reconstruct the structure then refine the couplings by cavity method.

References

1. Cocco, S., Leibler, S., Monasson, R.: Proc. Natl. Acad. Sci. USA **206**, 14058 (2009)
2. Weigt, M., White, R.A., Szurmant, H., Hoch, J.A., Hwa, T.: Proc. Natl. Acad. Sci. USA **106**, 67 (2009)
3. Ackley, D.H., Hinton, G.E., Sejnowski, T.J.: Cogn. Sci. **9**, 147 (1985). ISSN 0364-0213
4. Roudi, Y., Tyrcha, J., Hertz, J.: Phys. Rev. E **79**, 051915 (2009)
5. Schneidman, E., Berry, M.J., Segev, R., Bialek, W.: Nature **440**, 1007 (2006)
6. Roudi, Y., Hertz, J.: Phys. Rev. Lett. **106**, 048702 (2011)
7. Mezard, M., Sakellariou, J.: J. Stat. Mech. **2011**, L07001 (2011)

8. Sakellariou, J., Roudi, Y., Mezard, M., Hertz, J.: arXiv:1106.0452 (2011)
9. Braunstein, A., Ramezanpour, A., Zecchina, R., Zhang, P.: Phys. Rev. E **83**, 056114 (2011)
10. Zeng, H.-L., Aurell, E., Alava, M., Mahmoudi, H.: Phys. Rev. E **83**, 041135 (2011)
11. Coolen, A.C.C.: arXiv:cond-mat/0006011 (2000)
12. Hatchett, J.P.L., Wemmenhove, B., Castillo, I.P., Nikoletopoulos, T., Skantzos, N.S., Coolen, A.C.C.:
    J. Phys. A, Math. Gen. **37**, 6201 (2004)
13. Coolen, A.C.C., Laughton, S.N., Sherrington, D.: Phys. Rev. B **53**, 8184 (1996)
14. Sommers, H.-J.: Phys. Rev. Lett. **58**, 1268 (1987)
15. Neri, I., Bolle, D.: J. Stat. Mech. **2009**, P08009 (2009)
16. Aurell, E., Mahmoudi, H.: J. Stat. Mech. **2011**, P04014 (2011)
17. Zhang, P., Chen, Y.: Physica A **387**, 4441 (2008)