Inference by belief propagation in composite systems

Etienne Mallard and David Saad
Neural Computing Research Group, Aston University, Birmingham B4 7ET, United Kingdom
(Received 19 March 2008; published 8 August 2008)

We devise a message passing algorithm for probabilistic inference in composite systems, consisting of a large number of variables, that exhibit weak random interactions among all variables and strong interactions with a small subset of randomly chosen variables; the relative strength of the two interactions is controlled by a free parameter. We examine the performance of the algorithm numerically on a number of systems of this type for varying mixing parameter values.

DOI: 10.1103/PhysRevE.78.021107 PACS number(s): 02.50.−r, 89.75.−k, 02.60.Pn, 75.10.Nr

I. INTRODUCTION

Complexity has been identified as a key research area of significant future demand in a variety of fields from telecommunication and *ad hoc* networks to biological systems, transport, and social networks [1,2]. Among the main characteristics of complex systems are their heterogeneous structure, nonlinearity, and large scale, which makes it difficult to investigate them using traditional methods. Current research activities mostly focus on large scale simulations, for instance of interacting agents, or on numerical solutions of coupled nonlinear deterministic or stochastic differential equations.

The sensitivity of most numerical methods to model parameters and external observations, the sheer scale of the systems studied, and the range of interactions involved pose significant difficulties when it comes to reliable numerical modeling and analysis of such systems. Providing robust, principled, and reliable algorithms for obtaining solutions in specific instantiations of such systems is considered very difficult due to their large scale, nonlinearity, and inherent multilevel interactions.

The general approach that we advocate for understanding such systems is based on approximate and distributive probabilistic methods that are local, scale well (linearly or at most quadratically) with the system size, accommodate variable and measurement uncertainties, and readily provide confidence levels for the inferred variables. These take the form of message passing techniques such as belief propagation (BP) that have been developed independently within the physics [3], computer science [4,5] and information theory [6] communities. The main advantage of these methods is their moderate growth in computational cost, with respect to the systems size, due to the local nature of the calculation when applied to problems that can be mapped onto sparse graphs. They have been proved to be exact on polytrees and provide good approximations as long as the number of short loops in the corresponding graph is small.

Different approaches, based on mean-field approximations, have been suggested for obtaining solutions in the case of densely connected systems, where the number of connections is large and of the same order as the number of variables [while the connection strength is relatively weak, $O(N^{-1/2})$ where N is the system size] [3,7,8]. These highly successful methods heavily rely on the assumptions that the system is very large, densely connected, and the interactions weak, through a sequence of approximations.

Until recently, message passing techniques were deemed unsuitable for inference in densely connected systems due to the inherently high number of short loops in the corresponding graphical representation, and the large number of connections per node, which result in a high computational cost. Both properties are considered prohibitive to the use of conventional message passing techniques. However, various methods have been recently suggested [9–12] for a message passing based algorithm in densely connected systems; they rely on replacing individual messages by averages sampled from a Gaussian distribution of some mean and variance that are modified iteratively.

The problem we focus on here is that of composite systems with large numbers of elements and multilevel interactions, which represents a particular manifestation of a complex system. These are particularly difficult and challenging systems to analyze since although principled approaches have been devised separately for systems with very dense [9] or very sparse [4,5] interacting elements, they typically fail for the composite multilevel systems. In this paper we show how recent advances in the development of message passing techniques can give rise to a new algorithm which accommodates messages from both sparse and dense components of the same graph.

The motivation for studying the specific system considered here is that it is among the simplest composite models involving a combination of dilute and dense couplings. In addition, such models are likely to feature in various engineering applications. For example, dense weak interactions are likely to emerge, in addition to the designed sparse and strong interactions, in high-density integrated circuits. Analyzing the effects of the emerging weak couplings may be highly useful for minimizing their impact. Alternatively, such systems may be engineered deliberately with a combination of dilute (strong) and dense (weak) interactions either to make them more robust or to exploit specific properties of the composite system. In multiuser channel coding, for instance, this may make the communication process more robust against different types of noise, desynchronization or malicious attacks [13]. Only recently, a special case of the system studied here was suggested as a model for studying the resilience of networks against attacks [14]. A similar system to the one studied here was recently analyzed at a macroscopic level, using the replica method [15], complementing

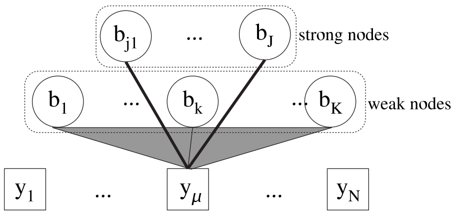

FIG. 1. Graphical representation of the composite inference problem, characterized by a high number of weak links between all variables and an evidential node, and only a few strong couplings to randomly selected variables.

the microscopic treatment of specific instances introduced here.

The remainder of the paper is organized as follows. In Sec. II we present the model to be studied, while the message passing equations will be derived in Sec. III. Experiments aimed at examining the new algorithm will be described in Sec. IV. We will conclude with a summary and future research directions in Sec. V.

II. MODEL

While BP-based algorithms for inference in sparsely [4,5] and densely [9–11] connected systems have been introduced, no method has been devised for inference in composite systems that comprise both weak (but densely connected) and strong (sparse) interactions.

The model we focus on here is based on N noisy measurements y_μ, $\mu \in \{1 \cdots N\}$ of K interacting variables (bits or spins) b_k, $k \in \{1,2,\ldots,K\}$. The model comprises two components, the first represents *weak* interactions between all variables while the second represents a few (J) *strong* interactions with a few, randomly chosen, variables. The random binary interactions themselves $s_{\mu i} \in \{-1,1\}$ are chosen with equal probability of taking the values ± 1. The measurements, corrupted by Gaussian noise of zero mean and standard deviation σ_0 take the form

$$y_\mu = \frac{\gamma}{\sqrt{J}} \sum_{j=1}^{J} s_{\mu i_j} b_{i_j} + \frac{1-\gamma}{\sqrt{K}} \sum_{k=1}^{K} s_{\mu k} b_k + \sigma_0 n_\mu, \qquad (1)$$

where $\{i_1, i_2, \ldots, i_j\}$ are a set of randomly selected (fixed) indices for each evidential node (measurement) μ for the given system, $0 \le \gamma \le 1$ is a coefficient that regulates the ratio of dense and sparse components, and the coefficients $1/\sqrt{K}$ and $1/\sqrt{J}$ normalize the strength of both components to $O(1)$. Figure 1 provides a graphical representation of the model investigated. Notice that the sum over nodes with weak couplings includes *all* variables and does not separate those with strong interactions; as there are only $O(1)$ such variables, their contribution to the sum over nodes with weak couplings in negligible.

This model represents a special case of a composite system where different levels of interaction coexist. In particular, the strength of the interactions are defined in such a way as to keep both contributions of similar order even in large systems. We believe the same approach can be easily extended to more complex connectivity and interaction profiles.

The suggested composition of strong and weak couplings has been chosen as it is arguably the simplest choice. Rescaling the couplings with respect to the choice of γ may also be sensible although both choices have their pros and cons in terms of the scaling properties they provide.

III. ALGORITHM

We employ the Bayesian scheme to infer the values of the various variables given the composite interactions. The aim is to find the local maximum *a posteriori* (MAP) variable values, also termed marginal posterior maximizer (MPM), $P(b_k|\{y_\mu\}, \forall \mu)$, on the basis of the observation and prior belief in the values of the variables.

A. Messages

Belief propagation algorithms are based on the derivation of messages, local conditional probabilities, to be passed from variable to evidential nodes and vice versa, using a closed set of approximate equations. In the case considered here there are two different types of messages from and to variable nodes that are strongly or weakly interacting with (or, correspondingly, sparsely and densely connected to) a particular evidential node.

The messages from variable to factor, or evidential, nodes are defined separately from those passed from factor to variable nodes:

$$P^{t+1}(y_\mu|b_k,\{y_{\nu\neq\mu}\}) = \sum_{b_{l\neq k}} P(y_\mu|\boldsymbol{b}) \prod_{l\neq k} P^t(b_l|\{y_{\nu\neq\mu}\}), \qquad (2)$$

$$P^t(b_k|\{y_{\nu\neq\mu}\}) = \alpha^t_{\mu k} \prod_{\nu\neq\mu} P^t(y_\nu|b_k,\{y_{\sigma\neq\nu}\}), \qquad (3)$$

where $t=1,2,\ldots$ is an iteration (time) index. There is also a normalization constant $\alpha^t_{\mu k}$ due to the two constraints $\Sigma_{y_\mu=\pm 1} P^t(y_\mu|b_k,\{y_{\nu\neq\mu}\})=1$ and $\Sigma_{b_k=\pm 1} P^t(b_k|\{y_{\nu\neq\mu}\})=1$. The marginalized posterior at tth update is evaluated from $P^t(y_\mu|b_k,\{y_{\nu\neq\mu}\})$ as $P^t(b_k|\boldsymbol{y}) = \alpha_k \Pi_{\mu=1}^N P^t(y_\mu|b_k,\{y_{\nu\neq\mu}\})$, where α_k is again a normalization constant.

Since b_k is a binary variable, one can with no loss of generality, parametrize the conditional probability distributions as

$$P^t(y_\mu|b_k,\{y_{\nu\neq\mu}\}) \propto (1+\hat{m}^t_{\mu k} b_k)/2,$$

$$P^t(b_k|\{y_{\nu\neq\mu}\}) = (1+m^t_{\mu k} b_k)/2,$$

$$\text{and} \quad P^t(b_k|\boldsymbol{y}) = (1+m^t_k b_k)/2$$

using the parameters $m^t_{\mu k}$ (magnetization), m^t_k, and $\hat{m}^{t+1}_{\mu k}$. This simplifies the expressions by writing recursive equations for the two sets of parameters

$$\hat{m}^{t+1}_{\mu k} = \frac{\sum_{\boldsymbol{b}} b_k P(y_\mu|\boldsymbol{b}) \prod_{l\neq k} [(1+m^t_{\mu l} b_l)/2]}{\sum_{\boldsymbol{b}} P(y_\mu|\boldsymbol{b}) \prod_{l\neq k} [(1+m^t_{\mu l} b_l)/2]}, \qquad (4)$$

$$m_{\mu k}^{t} = \tanh\left(\sum_{\nu \neq \mu} \tanh^{-1} \hat{m}_{\nu k}^{t} \right). \tag{5}$$

Employing these variables, the approximated posterior average of b_k at the tth update can be computed as $m_k^t = \tanh(\sum_{\mu=1}^{N} \tanh^{-1} \hat{m}_{\mu k}^{t})$. After convergence, the inferred value becomes

$$b_k = \mathrm{sgn}(m_k^t). \tag{6}$$

The two cases of strong and weak links will be considered separately when the messages $m_{\mu k}$, from variable k to evidential node μ, are calculated. If k is part of the strongly connected local neighborhood of evidential node μ, the equations will be based on the sparse graph model [4,5], whereas the approach used for the densely connected case [9–11] will be used when variable node k is weakly connected to μ.

B. Strongly interacting nodes

In calculating the messages $\hat{m}_{\mu k}$ and $m_{\mu k}$ one should consider separately the contribution made by nodes that interact strongly and weakly with the particular factor node examined. The summation over variables is decomposed to two separate sums, over the strongly and weakly interacting variables,

$$\hat{m}_{\mu k}^{t+1} = \frac{\sum_{\boldsymbol{b}_s} \sum_{\boldsymbol{b}_w} b_k P(y_\mu | \boldsymbol{b}) \prod_{l \neq k} [(1 + m_{\mu l}^{t} b_l)/2]}{\sum_{\boldsymbol{b}_s} \sum_{\boldsymbol{b}_w} P(y_\mu | \boldsymbol{b}) \prod_{l \neq k} [(1 + m_{\mu l}^{t} b_l)/2]},$$

$$m_{\mu k}^{t} = \tanh\left(\sum_{\nu \neq \mu} \tanh^{-1} \hat{m}_{\nu k}^{t} \right).$$

Considering $\Delta_\mu = \frac{1-\gamma}{\sqrt{K}} \sum_{l=1}^{K} s_{\mu l} b_l + \frac{\gamma}{\sqrt{J}} \sum_{j=1}^{J} s_{\mu i_j} b_{i_j}$, the first term represents a sum over a large number of independent variables while the second can be summed exactly to provide a certain value at each iteration t. To evaluate Δ_μ we employ the central limit theorem: Δ_μ obeys a Gaussian distribution $\mathcal{N}(\langle \Delta_\mu^t \rangle, (1 - Q_\mu^t)_\gamma)$, where $(1 - Q_\mu^t)_\gamma \equiv (1-\gamma)^2 (1 - Q_\mu^t)$,

$$\langle \Delta_\mu^t \rangle = \frac{1-\gamma}{\sqrt{K}} \sum_{l=1}^{K} s_{\mu l} m_{\mu l}^{t} + \frac{\gamma}{\sqrt{J}} \sum_{j=1}^{J} s_{\mu i_j} b_{i_j} \tag{7}$$

and

$$Q_\mu^t = (1/K) \sum_{l=1}^{K} (m_{\mu l}^t)^2, \tag{8}$$

which is well approximated by

$$Q^t = (1/K) \sum_{l=1}^{K} (m_l^t)^2. \tag{9}$$

Using the Gaussian nature of the noise and the distribution of Δ_μ one obtains at each iteration t

$$P(y_\mu | \boldsymbol{b}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_0^2 + (1 - Q^t)_\gamma}} \exp\left[-\frac{(y_\mu - \langle \Delta_\mu^t \rangle)^2}{2[\sigma_0^2 + (1 - Q^t)_\gamma]} \right] \tag{10}$$

and the corresponding messages

$$\hat{m}_{\mu k}^{t+1} = \frac{\sum_{\boldsymbol{b}_s} b_k P(y_\mu | \boldsymbol{b}) \prod_{l \in S(\mu)/k} [(1 + m_{\mu l}^{t} b_l)/2]}{\sum_{\boldsymbol{b}_s} P(y_\mu | \boldsymbol{b}) \prod_{l \in S(\mu)/k} [(1 + m_{\mu l}^{t} b_l)/2]}, \tag{11}$$

where $S(\mu)/k$ represents all the variables connected through strong links to factor μ, except node k. Note that dependence on the variable b_k also appears through the expression for $P(y_\mu | \boldsymbol{b})$ as in Eq. (10).

C. Weakly interacting nodes

One of the main differences with the previous case is that we exploit the dense character of the weak interactions and expand the contribution around the mean by excluding the contribution of a single variable as in [9]. We now consider $\Delta_{\mu k} = \frac{1-\gamma}{\sqrt{K}} \sum_{l=1, l \neq k}^{K} s_{\mu l} b_l + \frac{\gamma}{\sqrt{J}} \sum_{j=1}^{J} s_{\mu i_j} b_{i_j}$. Using the central limit theorem $\Delta_{\mu k}$ obeys a Gaussian distribution $\mathcal{N}(\langle \Delta_{\mu k}^t \rangle, (1 - Q_{\mu k}^t)_\gamma)$, at each iteration t, where

$$\langle \Delta_{\mu k}^t \rangle = \frac{1-\gamma}{\sqrt{K}} \sum_{l, l \neq k}^{K} s_{\mu l} m_{\mu l}^{t} + \frac{\gamma}{\sqrt{J}} \sum_{j=1}^{J} s_{\mu i_j} b_{i_j} \tag{12}$$

and

$$Q_{\mu k}^t = (1/K) \sum_{l=1}^{K} (m_{\mu l}^t)^2. \tag{13}$$

Following a similar derivation as in [9], the conditional probability can be written as

$$P(y_\mu | \boldsymbol{b}) = \left[1 + \frac{[y_\mu - \langle \Delta_{\mu k}^t \rangle] s_{\mu k} b_k (1-\gamma)}{[\sigma_0^2 + (1 - Q^t)_\gamma] \sqrt{K}} \right] \times \frac{1}{\sqrt{2\pi [\sigma_0^2 + (1 - Q^t)_\gamma]}} \exp\left[-\frac{(y_\mu - \langle \Delta_{\mu k}^t \rangle)^2}{2[\sigma_0^2 + (1 - Q^t)_\gamma]} \right].$$

Using the notation

$$A = \frac{[y_\mu - \langle \Delta_{\mu k}^t \rangle] s_{\mu k} (1-\gamma)}{[\sigma_0^2 + (1 - Q^t)_\gamma] \sqrt{K}} \quad \text{and}$$

$$B = \frac{1}{\sqrt{2\pi [\sigma_0^2 + (1 - Q^t)_\gamma]}} \exp\left[-\frac{(y_\mu - \langle \Delta_{\mu k}^t \rangle)^2}{2[\sigma_0^2 + (1 - Q^t)_\gamma]} \right],$$

one can rewrite the message as

$$\hat{m}_{\mu k}^{t+1} = \frac{\sum_{\boldsymbol{b}_s} AB \prod_{l \in S(\mu)} [(1 + m_{\mu l}^{t} b_l)/2]}{\sum_{\boldsymbol{b}_s} B \prod_{l \in S(\mu)} [(1 + m_{\mu l}^{t} b_l)/2]}. \tag{14}$$

Notice that there is still a sum over $\boldsymbol{b}_s$, as in the sparse case, which has implications for the complexity of the suggested

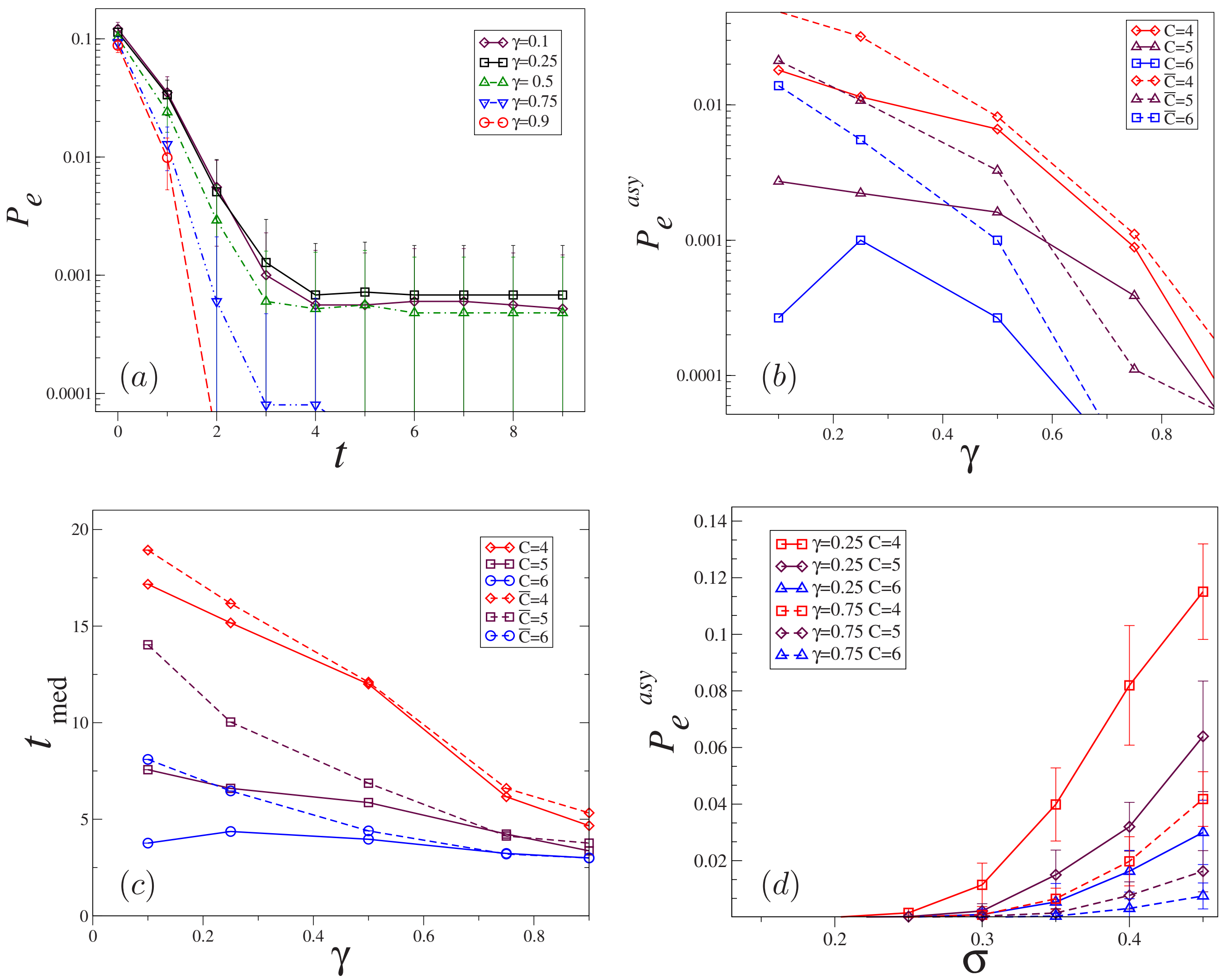

FIG. 2. (Color online) Performance of the BP-based algorithm for composite systems. (a) Error probability as a function of the number of iterations t for different mixing parameter values γ and fixed connectivity systems of $C=3$. Rapid convergence to low error-probability values is observed for high γ values while residual asymptotic error probability remains for dense connectivity dominated systems of low γ values. (b) Asymptotic error probability P_e^{asy} as a function of γ, for fixed (solid lines) and random connectivity (of mean $\bar{C}$—dashed lines). (c) Median number of time steps required for convergence $t_{\rm med}$ as a function of γ, both for the fixed (solid lines) and random (dashed lines) connectivities. (d) Asymptotic error probability as a function of the noise variance for fixed connectivity C, and mixing parameters γ $=0.25$ (solid lines), and 0.75 (dashed lines).

inference algorithm. Contributions from the densely connected nodes dominate the complexity of the algorithm and require $O(K^2)$ operations while the sparse components scale linearly in N but are exponential in the (small) number of sparse connections J.

IV. EXPERIMENT

To examine the efficacy of the algorithm for inference in composite systems and gain insight into the behavior of such systems in particular cases, we carried out a number of experiments for a system of size $N=1000$ and varying $K=NJ/\bar{C}$, as determined by the connectivity degree of strongly interacting neighbors per variable (average $\bar{C}$) and measurement (J). We kept the measurement node connectivity, of degree $J=3$, fixed and varied the value of $K=500$, 600, 750, in correspondence with the fixed (average) variable connectivities $C(\bar{C})=6,5,4$. The systems vary in the type of sparse variable connectivity applied (fixed or random), the strengths of the sparse (dense) interactions $\gamma(1-\gamma)$, and the noise level σ_0. Random connectivity graphs, representing the sparse interactions, and random interaction values have been generated in each of the experiments.

In each case we carried out 500 experiments by iterating Eqs. (7), (11), and (14) from randomly chosen initial conditions until convergence (more precisely, 25 graphs and 20 input and output tests for each point); we then inferred the values of the various variables on the basis of the pseudoposterior (6). The convergence and halting criterion was defined

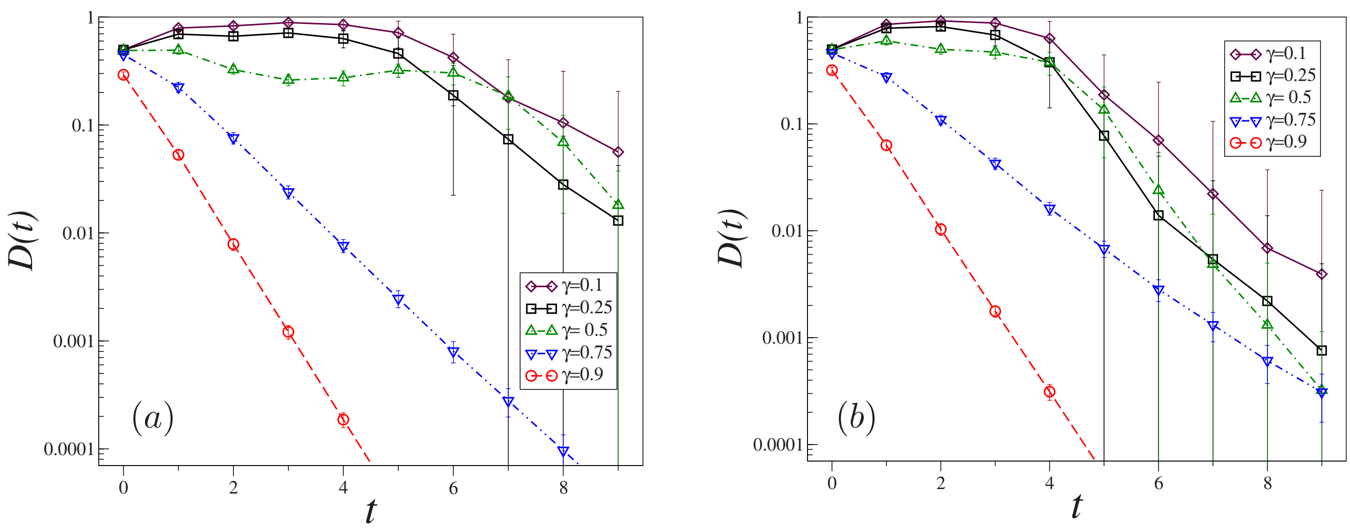

FIG. 3. (Color online) The convergence measure $D(t)$ for various γ and connectivity values. (a) Fixed connectivity $C=5$; (b) Fixed connectivity $C=6$.

as an unchanged solution over four iterations, or reaching a maximum of 20 iterations.

Two main sparse connectivity patterns have been used, fixed connectivity C and random connectivity, resulting in a Poissonian distribution of mean $\bar{C}$. Results obtained for the various cases are shown in Fig. 2 for fixed connectivity sparse couplings and randomly connected sparse graphs.

Figure 2(a) shows the average error probability of the inference algorithm as a function of the number of iterations for different γ values in the case of fixed connectivity $C=3$. While the algorithm results in rapid convergence to a very low error probability for high γ values, corresponding to a strong sparse component, it approaches a residual asymptotic error probability at low γ values, characteristic of densely connected and weakly interacting systems. The quality of the solutions obtained is shown in Fig. 2(b), where the asymptotic error-probability values P_e^{asy} are plotted as a function of γ, for different fixed connectivity values C (solid line) and Poissonian distributions of mean $\bar{C}$ (dashed line); error bars have been removed for brevity. We see that fixed connectivity systems typically show lower asymptotic values than the randomly connected systems for low γ values, a difference that disappears for high γ, where the dense couplings dominate; the asymptotic error probability typically decrease with the increase in γ. It is interesting to note that asymptotic results typically improve with the increase in connectivity as the ratio of variable to evidential nodes decreases.

To study the dependence of the rate of convergence on both γ and the degree of connectivity, we plotted in Fig. 2(c) the median number of time steps required for the system to converge as a function of γ, for both the fixed (solid lines) and random (dashed lines) connectivities; error bars have been removed for brevity. We see that the number of iterations required for convergence generally decreases as γ and the connectivity increase with little difference between the fixed and random sparse connectivity profiles, mainly in the low γ values. Also here, the dependence on the connectivity values may be explained by the varying ratio of variable to evidential nodes.

In Fig. 2(d) we examine the dependence of the asymptotic error probability on the noise variance for several fixed and random connectivity values $\gamma=0.25$ and 0.75 (solid and dashed lines, respectively). We see that, unsurprisingly, the asymptotic error rate increases with the noise variance with systems of lower γ values and higher connectivity values exhibiting higher robustness to noise as the ratio of variable to evidential nodes decreases.

Finally, we studied the convergence rate of the algorithm for the various systems studied. While the convergence criterion used in our simulations was sufficient for evaluating the asymptotic performance of the suggested algorithm it cannot provide reliable information on the rate of convergence.

To quantify the convergence rate for various γ and connectivity values C, we plotted the evolution of the convergence measure,

$$D(t) = \max_{\forall \mu,k}[P^{t+1}(y_\mu|b_k) - P^t(y_\mu|b_k)].$$

Simulations for fixed connectivity graphs have been carried out using the same experimental setup as before. The results of Fig. 3 show that inference in sparse-dominated systems converges rapidly while dense-dominated systems tend to converge much slower. Naturally, the effect is more emphasized for low sparse connectivity values C. It is interesting to note that the existence of stronger sparse inputs *delays* the convergence for low γ values, presumably since stronger average messages from the densely connected nodes are required to counterbalance messages from the more volatile dense couplings (e.g., for $\gamma=0.1, 0.25$).

V. CONCLUSION

The study of complex systems poses significant new challenges as they are, by their very nature, large, nonuniform, and their components interact in a nontrivial manner. While a significant effort is dedicated to numerical studies of specific complex systems we see the main way forward through a principled distributive Bayesian approach; this enables one to carry out the calculations in times that scale linearly or quadratically with the systems size and provides both the approximate inferred values and the related confidence level.

In the current study we have combined message passing based algorithms that were developed for both sparsely and densely connected systems to study exemplar composite systems. These comprise a densely connected system of weakly interacting components and an overlaid sparse but strong interaction.

We demonstrated the efficacy of the suggested algorithm by studying the error rate obtained in the composite system case for various mixtures of dense and sparse interactions, governed by the mixing parameter γ, and different sparse connectivity characteristics (fixed and random of various mean values).

Unsurprisingly, the system examined exhibits a sparse system behavior as long as γ values are high, and gradually exhibits a behavior characteristic of a densely connected system for low γ values. It shows fast convergence to low error rates for high γ values (systems dominated by the sparse couplings), but performance improves as the connectivity value of the sparse couplings increases (but is still small) due to the decreasing ratio of variable to evidential nodes. While sparse-dominated systems tend in general to converge faster, low connectivity values tend to slow down convergence in the remainder of the system for low γ values, presumably due to the higher volatility of dominating messages from the sparsely connected components. The current algorithm can be used to derive general properties of such systems by applying density evolution, as well as inference in specific instances.

Future work in this area is practically unlimited. First, one may consider extending the work to study multistate composite systems where variables are not limited to the binary representation [3,16], or to other connectivity profiles that require the use of generalized BP [17] or cluster variation techniques [18–20]. Second, one may seek a principled approximation for systems that have an intermediate range of interactions that is not readily accommodated within the current two-level system. Finally, one may want to apply the algorithm and its derivatives to real systems that exhibit a two- (or multi-) level behavior, such as recently introduced stochastic weather forecasting models that accommodate both nearest cell and global long range interactions; or of densely connected sensor (or mechanical) arrays implemented in integrated circuits that exhibit both strong interactions with neighboring sensors and a weak coupling with all other sensors.

ACKNOWLEDGMENTS

Support from EPSRC Grant No. EP/E049516/1 is acknowledged.

[1] J. Reichardt and S. Bornholdt, Phys. Rev. Lett. **93**, 218701 (2004).
[2] J. Reichardt and S. Bornholdt, Phys. Rev. E **74**, 016110 (2006).
[3] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific Publishing Co., Singapore, 1987).
[4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann Publishers, Inc., San Francisco, 1988).
[5] F. Jensen, *An Introduction to Bayesian Networks* (UCL Press, London, 1996).
[6] R. Gallager, IRE Trans. Inf. Theory **IT-8**, 21 (1962).
[7] M. Opper and O. Winther, Phys. Rev. Lett. **76**, 1964 (1996).
[8] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice* (MIT Press, Cambridge, MA, 2001).
[9] Y. Kabashima, J. Phys. A **36**, 11111 (2003).
[10] J. Neirotti and D. Saad, Europhys. Lett. **71**, 866 (2005).
[11] J. Neirotti and D. Saad, Phys. Rev. E **76**, 046121 (2007).
[12] A. Braunstein and R. Zecchina, Phys. Rev. Lett. **96**, 030201 (2006).
[13] J. Raymond and D. Saad (unpublished).
[14] M. O. Hase and J. F. F. Mendes, J. Phys. A **41**, 145002 (2008).
[15] J. Raymond and D. Saad, J. Phys. A (to be published).
[16] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, Oxford, UK, 2001).
[17] J. S. Yedidia, W. T. Freeman, and Y. Weiss, IEEE Trans. Inf. Theory **51**, 2282 (2005).
[18] R. Kikuchi, Phys. Rev. **81**, 988 (1951).
[19] T. Morita, J. Phys. Soc. Jpn. **12**, 753 (1957).
[20] G. An, J. Stat. Phys. **52**, 727 (1988).