EUROPHYSICS LETTERS 1 September 2005

Europhys. Lett., **71** (5), pp. 866–872 (2005)
DOI: 10.1209/epl/i2005-10148-5

Improved message passing for inference in densely connected systems

J. P. Neirotti and D. Saad

The Neural Computing Research Group, Aston University - Birmingham B4 7ET, UK

received 1 April 2005; accepted in final form 28 June 2005
published online 29 July 2005

PACS. 89.70.+c – Information theory and communication theory.
PACS. 75.10.Nr – Spin-glass and other random models.
PACS. 64.60.Cn – Order-disorder transformations; statistical mechanics of model systems.

Abstract. – An improved inference method for densely connected systems is presented. The approach is based on passing condensed messages between variables, representing macroscopic averages of microscopic messages. We extend previous work that showed promising results in cases where the solution space is contiguous to cases where fragmentation occurs. We apply the method to the signal detection problem of Code Division Multiple Access (CDMA) for demonstrating its potential. A highly efficient practical algorithm is also derived on the basis of insight gained from the analysis.

Graphical models (Bayes belief networks) provide a powerful framework for modelling statistical dependencies between variables [1–3]. They play an essential role in devising a principled probabilistic framework for inference in a broad range of applications from medical expert systems to decoders in telecommunication systems.

Message passing techniques are typically used for inference in graphical models that can be represented by a sparse graph. They are aimed at obtaining posterior-based estimates of the system's variables by iteratively passing messages (locally calculated conditional probabilities) between variables. Iterative message passing of this type is guaranteed to converge to the globally correct estimate when the system is tree-like; there are no such guarantees for systems with loops, although message passing techniques have been successfully used in loopy systems. A clear link has been established between certain message passing algorithms and well-known methods of statistical mechanics [4] such as the Bethe approximation [5, 6].

Two inherent limitations seem to prevent the use of message passing techniques in densely connected systems: 1) Their high connectivity implies an exponentially growing computational cost. 2) The existence of an exponential number of loops that render the method inconsistent. However, an exciting new approach was recently suggested [7] for extending Belief Propagation (BP) techniques [1–3] to densely connected systems. In this approach, messages are grouped together, giving rise to macroscopic random variables drawn from a different Gaussian distribution of varying mean and variance for each of the nodes. The technique has been successfully applied to signal detection in Code Division Multiple Access (CDMA) problems

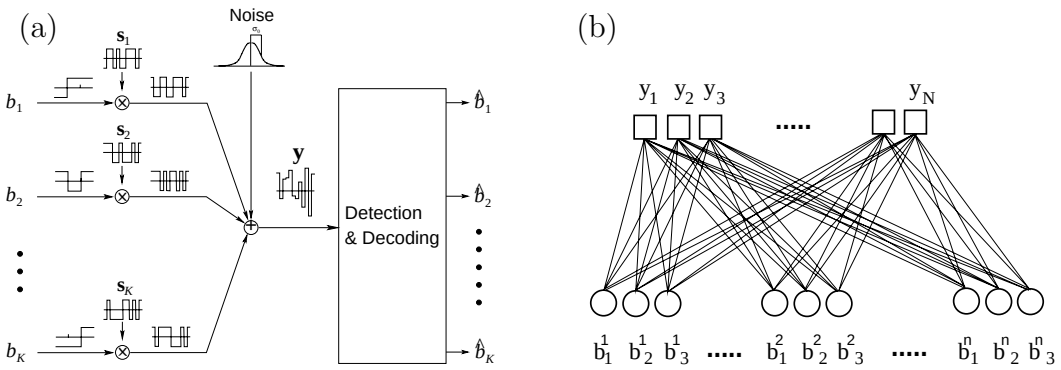

Fig. 1 – (a) Signal detection in CDMA. (b) Replicated solutions $\mathbb{B}=(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_K)$ for given data.

and the results reported are competitive with those of other state-of-the-art techniques. However, the current approach shows some inherent limitations [7], presumably of a similar nature to those of the replica-symmetric solution in equivalent Ising spin models [8, 9].

In a separate recent development [10], the replica-symmetric-equivalent BP was extended to Survey Propagation (SP), which corresponds to one-step replica symmetry breaking in diluted systems. This new algorithm, motivated by a theoretical physics interpretation of the space of solutions, has been highly successful in solving hard computational problems [10], far beyond other existing approaches. In addition, it has facilitated theoretical studies of the corresponding physical system and has contributed to our understanding of it [11].

Inspired by the extension of BP to SP we have extended the approach of [7], designed for inference in densely connected systems, in a similar manner by including an average over multiple pure states. In this letter we outline the derivation of this extension, which is general and can be applied to a broad range of inference problems. However, for giving a specific example and for highlighting the advantages with respect to the original method [7], we apply it to the problem of signal detection in CDMA and devise a *practical* algorithm based on insight gained from the analysis. Other applications will be presented elsewhere.

Multiple access communication refers to the transmission of multiple messages to a single receiver. The scenario we study here is that of K users transmitting independent messages over an additive white Gaussian noise (AWGN) channel of zero mean and variance σ_0^2. Various methods are in place for separating the messages, in particular Time, Frequency and Code Division Multiple Access [12]. The latter, is based on spreading the signal by using K individual random binary spreading codes of spreading factor N. We consider the large-system limit, in which the number of users K tends to infinity while the system load $\beta \equiv K/N$ is kept to be $\mathcal{O}(1)$. We focus on a CDMA system using binary phase shift keying (BPSK) symbols, shown schematically in fig. 1(a), where signals are modulated (spread) using K random binary modulation sequences, and will assume the power to be completely controlled to unit energy. The received aggregated, modulated and corrupted signal is of the form

$$y_\mu = \frac{1}{\sqrt{N}} \sum_{k=1}^{K} s_{\mu k} b_k + \sigma_0 n_\mu \,,$$

where b_k is the bit transmitted by user k, $s_{\mu k}$ is the binary spreading chip value, n_μ is the Gaussian noise variable drawn from $\mathcal{N}(0,1)$, and y_μ the received message. The goal is to get an accurate estimate of the vector $\boldsymbol{b}$ for all users given the received message vector $\boldsymbol{y}$ by

approximating the posterior $P(\boldsymbol{b}|\boldsymbol{y})$. A method for obtaining a good estimate of the posterior probability in the case where the noise level is accurately known was presented in [7]. However, the derivation is based on inferring a single set of system variables in the presence of multiple solutions and is therefore bound to fail, as has been observed, when the solution space becomes fragmented. This occurs, for instance, when the noise level is unknown.

The reason for the failure in this case can be qualitatively understood by the same arguments as in the case of sparse graphs. The existence of competing solutions gives rise to conflicting messages that prevent the algorithm from converging to an accurate estimate. An improved solution can be obtained by averaging over the various solutions, inferred from the same data, in a similar manner to the SP approach. The main difference is that the messages in the current case are more complex.

Figure 1(b) shows the CDMA signal detection problem we aim to solve as a bipartite graph where $\mathbb{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_K)$ the set of binary vectors, $\boldsymbol{b}_k = (b_k^1, b_k^2, \ldots, b_k^n)$, where n is the solution (replica) index. Using Bayes rule one obtains the BP equations:

$$\begin{aligned} P^{t+1}(y_\mu|\boldsymbol{b}_k, \{y_{\nu\neq\mu}\}) &= \widehat{a}_{\mu k}^{t+1} \sum_{\boldsymbol{b}_{l\neq k}} P\left(y_\mu|\mathbb{B}\right) \prod_{l\neq k} P^t\left(\boldsymbol{b}_l|\left\{y_{\nu\neq\mu}\right\}\right), \\ P^t\left(\boldsymbol{b}_l|\left\{y_{\nu\neq\mu}\right\}\right) &= a_{\mu k}^t \prod_{\nu\neq\mu} P^t\left(y_\nu|\boldsymbol{b}_l\,, \left\{y_{\sigma\neq\nu}\right\}\right), \end{aligned} \tag{1}$$

where $\widehat{a}_{\mu k}^{t+1}$ and $a_{\mu k}^t$ are normalization constants. An explicit expression for the likelihood is required for deriving the posterior

$$P\left(\mathbb{B}\,|\,\boldsymbol{y}\right) = \frac{\prod_{\mu=1}^N P\left(y_\mu\,|\,\mathbb{B}\right)}{\mathsf{Tr}_{\{\mathbb{B}\}} \prod_{\mu=1}^N P\left(y_\mu\,|\,\mathbb{B}\right)}\,. \tag{2}$$

The latter is derived from the noise model (assuming zero mean and variance σ^2)

$$P\left(y_\mu\,|\,\mathbb{B}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(\boldsymbol{y}_\mu - \boldsymbol{\Delta}_\mu\right)^{\mathsf{T}} \mathbb{I}\left(\boldsymbol{y}_\mu - \boldsymbol{\Delta}_\mu\right)}{2\sigma^2}\right], \tag{3}$$

where $\boldsymbol{y}_\mu = y_\mu \boldsymbol{u}$, $\boldsymbol{u}^{\mathsf{T}} \equiv \overbrace{(1,\,1,\,\cdots,\,1)}^{n}$ and $\boldsymbol{\Delta}_\mu \equiv \frac{1}{\sqrt{N}} \sum_{k=1}^K s_{\mu k} \boldsymbol{b}_k$. Understanding the correlation between the replicated solutions is at the heart of the new approach. An explicit expression for the statistical dependence between solutions is required for obtaining a closed set of update equations. We assume a dependence of the form

$$P^t\left(\boldsymbol{b}_k\,|\,\left\{y_{\nu\neq\mu}\right\}\right) \propto \exp\left[\boldsymbol{h}_{\mu k}^{t\mathsf{T}}\,\boldsymbol{b}_k + \frac{1}{2}\boldsymbol{b}_k^{\mathsf{T}} \mathbb{Q}_{\mu k}^t\,\boldsymbol{b}_k\right], \tag{4}$$

where $\boldsymbol{h}_{\mu k}^t$ is a vector representing an external field and $\mathbb{Q}_{\mu k}^t$ the matrix of cross-replica interaction. Furthermore, we assume the following symmetry between replica:

$$\begin{aligned} \left(\mathbb{Q}_{\mu k}^t\right)^{\mathrm{ab}} &= \delta^{\,\mathrm{ab}}\, q_{\mu k}^t + \left(1 - \delta^{\,\mathrm{ab}}\right) p_{\mu k}^t\,, \\ \boldsymbol{h}_{\mu k}^t &= h_{\mu k}^t \boldsymbol{u}. \end{aligned} \tag{5}$$

An explicit expression for eq. (4) immediately follows.

We expect the free energy to be self-averaging and obtain the scaling behavior of the various parameters: $h, q_0 \sim \mathcal{O}(1)$, $p_0, \sigma_q^2 \sim \mathcal{O}(n^{-1})$, and $\sigma_p^2 \sim \mathcal{O}(n^{-3})$. These will be instrumental later

on for expanding expressions to leading order. More specifically, in the remainder of the paper we will rescale the off-diagonal elements of $\mathbb{Q}^t_{\mu k}$ to $g^t_{\mu k}/n$, where $g^t_{\mu k} \sim \mathcal{O}(1)$. The marginalized posterior at time (iteration step) t takes the form

$$P^t(\boldsymbol{b}_k|\,\{y_{\nu\neq\mu}\}) = \frac{\int_{-\infty}^{\infty} \mathrm{d}x\, \exp\left[-n\frac{\left(x-h^t_{\mu k}\right)^2}{2g^t_{\mu k}} + x\sum_{\mathrm{a}=1}^{n} b^{\mathrm{a}}_k\right]}{2^n \int_{-\infty}^{\infty} \mathrm{d}x\, \exp\left[-n\Phi\left(x;\, h^t_{\mu k},\, g^t_{\mu k}\right)\right]},$$

$$\Phi\left(x;\, h^t_{\mu k},\, g^t_{\mu k}\right) = -\frac{\left(x-h^t_{\mu k}\right)^2}{2g^t_{\mu k}} + \ln\left(\cosh(x)\right). \tag{6}$$

Equation (6) is similar to expressions obtained using the cavity approach (*e.g.*, for the SK model) in the case of 1-step replica symmetry breaking [8]; but differ both in the value assumed for the variable n and its interpretation.

To find the dominant solutions in the case of large n, one studies the maxima of $\Phi(x;\, h,\, g)$. The main contribution comes from a regime where $g^t_{\mu k} > 1$ and $0 < h^t_{\mu k}/g^t_{\mu k} \ll 1$, in which $\Phi(x;\, h,\, g)$ takes the form of an almost symmetric pair of Gaussians located at

$$x^t_{\pm,\mu k} \simeq \pm x^t_{0,\mu k} + \frac{g^t_{\mu k}}{g^t_{\mu k} + \left(x^t_{0,\mu k}\right)^2 - \left(g^t_{\mu k}\right)^2} h^t_{\mu k}\,, \tag{7}$$

where $\pm x_0$ are the positions of the peaks at zero field. To calculate the correlations between replicas we expand $P(y_\mu\,|\,\mathbb{B})$ in the large-N limit (eq. (3)), as in [7], to obtain

$$P\left(y_\mu\,|\,\mathbb{B}\right) \simeq \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\left(\boldsymbol{y}_\mu - \boldsymbol{\Delta}_{\mu k}\right)^{\mathsf{T}}\left(\boldsymbol{y}_\mu - \boldsymbol{\Delta}_{\mu k}\right)}{2\sigma^2}\right]\left[1 + \frac{s_{\mu k}}{\sqrt{N}\sigma^2}\left(\boldsymbol{y}_\mu - \boldsymbol{\Delta}_{\mu k}\right)^{\mathsf{T}}\boldsymbol{b}_k\right], \tag{8}$$

where $\boldsymbol{\Delta}_{\mu k} = \frac{1}{\sqrt{N}}\sum_{l\neq k} s_{\mu l}\boldsymbol{b}_l$. However, to use eq. (8) for deriving an explicit update rule in eq. (1) we need to obtain the distribution $P(\boldsymbol{\Delta}_{\mu k})$.

For large n, and using the marginalized distribution (6), the mean values of $\langle b^{\mathrm{a}}_k\rangle$, $\langle b^{\mathrm{a}}_k b^{\mathrm{b}}_k\rangle$ and the corresponding covariance matrix can be obtained explicitly as functions of $a^t_{\pm,\mu k}$,

$$a^t_{\pm,\mu k} \simeq \exp\left[\pm n m^t_{\mu k} h^t_{\mu k}\right] / \left[\exp\left[n m^t_{\mu k} h^t_{\mu k}\right] + \exp\left[-n m^t_{\mu k} h^t_{\mu k}\right]\right], \tag{9}$$

where $m^t_{\mu k} \equiv \tanh(x^t_{0,\mu k}) = x^t_{0,\mu k}/g^t_{\mu k}$. From the previously obtained scaling behavior we define $a^t_{\pm,\mu k}$ *up to a free parameter* that will prove to be essential for deriving the new algorithm.

To calculate the Gaussian distribution $P(\boldsymbol{\Delta}_{\mu k}) = \mathcal{N}(\langle \Delta^{\mathrm{a}}_{\mu k}\rangle, \boldsymbol{\chi}^{t,\mathrm{ab}}_{\mu k})$, one uses the mean values of $\langle b^{\mathrm{a}}_k\rangle$, $\langle b^{\mathrm{a}}_k b^{\mathrm{b}}_k\rangle$ and the corresponding covariance matrix to obtain:

$$\begin{aligned}
\left\langle \Delta^{\mathrm{a}}_{\mu k}\right\rangle &= \frac{1}{\sqrt{N}}\sum_{l\neq k} s_{\mu l} m^t_{\mu l}\,,\\
\left(\boldsymbol{\chi}^t_{\mu k}\right)^{\mathrm{ab}} &\equiv \left\langle \Delta^{\mathrm{a}}_{\mu k}\Delta^{\mathrm{b}}_{\mu k}\right\rangle - \left\langle \Delta^{\mathrm{a}}_{\mu k}\right\rangle\left\langle \Delta^{\mathrm{b}}_{\mu k}\right\rangle = \delta^{\mathrm{ab}}\beta\left(1 - Q^t_{\mu k}\right) + \left(1-\delta^{\mathrm{ab}}\right)\beta\, R^t_{\mu k}\,,
\end{aligned} \tag{10}$$

where $Q^t_{\mu k}$ and $R^t_{\mu k}$ can be approximated using the law of large numbers as

$$\begin{aligned}
Q^t_{\mu k} &\equiv \frac{1}{K}\sum_{l\neq k}\left(a^t_{+,\mu k}\tanh\left(x^t_{+,\mu k}\right) + a^t_{-,\mu k}\tanh\left(x^t_{-,\mu k}\right)\right)^2 \simeq \frac{1}{K}\sum_{l\neq k}\left(m^t_{\mu k}\right)^2,\\
R^t_{\mu k} &\equiv \frac{2}{K}\sum_{l\neq k} a^t_{+,\mu k}a^t_{-,\mu k}\tanh\left(x^t_{+,\mu k}\right)\tanh\left(x^t_{-,\mu k}\right) \simeq \frac{2}{K}\sum_{l\neq k} a^t_{+,\mu k}a^t_{-,\mu k}\left(m^t_{\mu k}\right)^2 \equiv \frac{1}{n}\Upsilon^t_{\mu k}\,.
\end{aligned}$$

Having obtained the distribution $P(\mathbf{\Delta}_{\mu k})$, and using eqs. (1) and (8), one can then calculate the expected value of b_k^{a} at time $t+1$:

$$\widehat{m}_{\mu k}^{t+1} = \left(\sigma^2 + \beta\left(1 - Q_{\mu k}^t\right) + \beta \Upsilon_{\mu k}^t\right)^{-1} \left(\frac{y_\mu \boldsymbol{s}_\mu}{\sqrt{N}} - \beta\left(\mathbb{P}_\mu - \mathbb{I}/K\right) \boldsymbol{m}_\mu^t\right)_k , \tag{11}$$

where $\mathbb{P}_\mu \equiv (1/K) s_{\mu k} s_{\mu l}$ and $\mathbb{I} \equiv \delta_{kl}$, respectively. We assume that the macroscopic variables are self-averaging and omit the μ, k indices.

The main difference between eq. (11) and the equivalent equation in [7] is the emergence of an extra term in the prefactor, $\beta\Upsilon^t$, reflecting correlations between different solutions groups (replica). More importantly, there is a remaining degree of freedom in the choice of the cross-replica covariance matrix Υ^t, which one can exploit to minimize the bit error probability at each time step. To calculate the bit error probability, one follows ref. [7] to obtain

$$P_b^t \equiv \frac{1}{2K} \sum_{k=1}^{K} \left(b_k - \operatorname{sgn}\left(m_k^t\right)\right) = \int_{-\infty}^{-E^t/\sqrt{F^t}} \mathcal{D}z \tag{12}$$

$$\text{with } m_k^t \simeq \tanh\left(\sum_{\mu=1}^{N} \widehat{m}_{\mu k}^t\right), \tag{13}$$

where $\mathcal{D}z \equiv \mathrm{d}z \, \exp[-z^2/2]/\sqrt{2\pi}$. We also define several macroscopic correlation measures:

$$\begin{aligned} M^t &\equiv \frac{1}{NK} \sum_{\mu=1}^{N} \sum_{k=1}^{K} b_k m_{\mu k}^t = \int \mathcal{D}z \, \tanh\left(\sqrt{F^t} z + E^t\right), \\ Q^t &\equiv \frac{1}{NK} \sum_{\mu=1}^{N} \sum_{k=1}^{K} \left(b_k m_{\mu k}^t\right)^2 = \int \mathcal{D}z \, \tanh^2\left(\sqrt{F^t} z + E^t\right) \end{aligned} \tag{14}$$

and

$$\begin{aligned} E^{t+1} &\equiv \frac{1}{K} \sum_{\mu=1}^{N} \sum_{k=1}^{K} b_k \widehat{m}_{\mu k}^{t+1} = \frac{1}{\sigma^2 + \beta\left(1 - Q^t + \Upsilon^t\right)}, \\ F^{t+1} &\equiv \sum_{\mu=1}^{N} \left[\frac{1}{K} \sum_{k=1}^{K} \left(b_k \widehat{m}_{\mu k}^{t+1}\right)^2 - \frac{1}{K^2} \left(\sum_{k=1}^{K} b_k \widehat{m}_{\mu k}^{t+1}\right)^2\right] \\ &\approx \left[\beta\left(1 - 2M^t + Q^t\right) + \sigma_0^2\right] \left(E^{t+1}\right)^2 . \end{aligned} \tag{15}$$

Optimizing P_b^t with respect to Υ^t, one obtains straightforwardly that $E^t = F^t$ and $Q^t = M^t$. In principle, the optimization can be done globally [13] but is of a limited practical value.

This implies that $\Upsilon^t = (\sigma_0^2 - \sigma^2)/\beta$ is just a constant. However, it holds the key to obtaining improved inference results and an efficient inference algorithm of significant practical value. If the noise estimate is identical to the true noise, the term vanishes and one retrieves the expression of [7]; otherwise, an estimate of the difference between the two noise values is required for computing E^t. Exploiting the result obtained from the optimization of (12), one

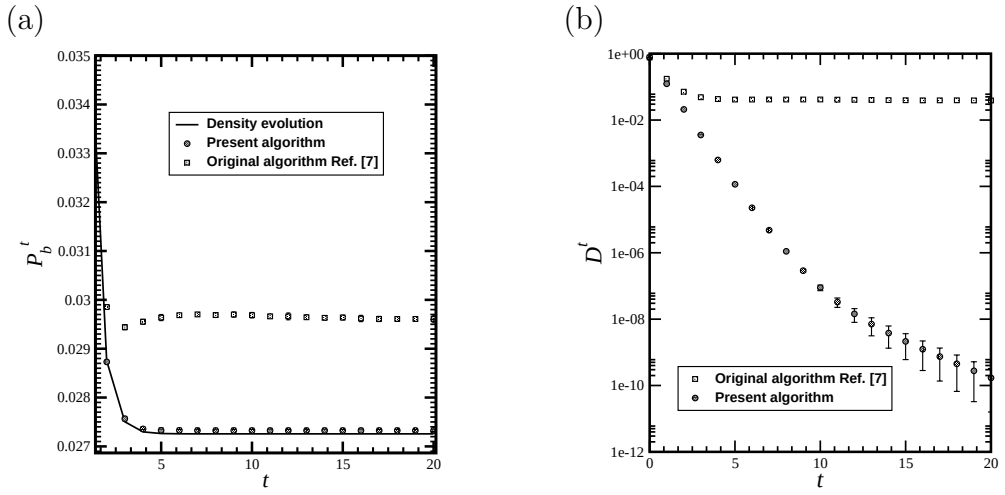

Fig. 2 – (a) Error probability of the inferred solution as a function of time. The system load $\beta = 0.25$, true and estimated noise levels $\sigma_0^2 = 0.25$ and $\sigma^2 = 0.01$, respectively. Squares represent results obtained by the original algorithm [7], the solid line the dynamics obtained from our equations; circles represent results obtained from the suggested *practical* algorithm. Variances are smaller than the symbol size. (b) The measure of convergence D of the obtained solutions, as a function of time; symbols are as in (a).

obtains an explicit expression for E^{t+1} that does not depend on σ^2:

$$E^{t+1} \simeq \frac{1}{K}\sum_{\mu=1}^{N} \widehat{\boldsymbol{m}}_\mu^{t+1}\cdot\widehat{\boldsymbol{m}}_\mu^{t+1} = \left[\frac{1}{\sigma^2+\beta\,(1-Q^t+\Upsilon^t)}\right]^2\left[\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2-\beta\left(2M^t-Q^t\right)\right]$$
$$= \left(E^{t+1}\right)^2\left[\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2-\beta Q^t\right] = \left[\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2-\beta Q^t\right]^{-1}. \tag{16}$$

This enables us to rewrite the update equation for $\widehat{m}_{\mu k}$, eq. (11), as

$$\widehat{m}_{\mu k}^{t+1} = \left\{\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2 - \beta Q^t\right\}^{-1}\left(\frac{y_\mu \boldsymbol{s}_\mu}{\sqrt{N}} - \beta\,\left(\mathbb{P}_\mu - K^{-1}\mathbb{I}\right)\,\boldsymbol{m}_\mu^t\right)_k \tag{17}$$

where no estimate on σ_0 is required.

This transforms the inference algorithm into a highly practical technique as it obviates the need for a prior belief of the noise level. The inference algorithm merely requires an iterative update of eqs. (16), (17), (13) and converges to a reliable estimate of the signal. The computational complexity of the algorithm is of $\mathcal{O}(NK^2)$ (reducing back to $\mathcal{O}(K^2)$ once the noise has been estimated).

To test the performance of our algorithm we studied the CDMA signal detection problem under typical conditions. The error probability of the inferred signals has been calculated for a system load $\beta = 0.25$, where the true noise level is $\sigma_0^2 = 0.25$ and the estimated value is $\sigma^2 = 0.01$, as shown in fig. 2(a). In this scenario we expect the original algorithm [7] to fail due to the discrepancy between the two noise levels. The solid line represents the expected theoretical results (density evolution), knowing the exact values of σ_0^2 and σ^2, while circles represent simulation results obtained via the suggested *practical* algorithm, where no such

knowledge is assumed. The results presented are based on 10^5 trials per point and a system size $N=2000$ and are superior to those obtained using the original algorithm [7].

Another performance measure one should consider is $D^t \equiv \frac{1}{K}|\boldsymbol{m}^t - \boldsymbol{m}^{t-1}|^2$. It provides an indication to the stability of the solutions obtained; when the algorithm converges to a single stable solution one would expect a vanishing D value, while fluctuating solutions will maintain a finite D value. In fig. 2(b) we see that results obtained using our algorithm show convergence to a reliable single solution in stark contrast to the results obtained by the original algorithm [7]. The physical interpretation of the difference between the two results is related to a replica-symmetry-breaking phenomenon.

In summary, we present a new algorithm for using message passing in densely connected systems that enables one to obtain reliable solutions even when the solution space is fragmented. It represents an extension of an existing algorithm similar to the extension of BP to SP. In addition, a method for estimating the true noise level emerges naturally, making the algorithm highly relevant for practitioners. The algorithm has been tested on the signal detection problem in CDMA and has provided superior results to other existing algorithms [7]([1]).

Further research is required to fully determine the potential of the new approach and its applicability for a variety of problems. Applications to other densely connected problems, such as the Ising perceptron parameter estimation and lossy compression, are underway.

* * *

Support from EVERGROW, IP No. 1935 in FET, EU FP6 is gratefully acknowledged.

REFERENCES

[1] Pearl J., *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann Publishers, San Francisco) 1988.
[2] Jensen F. V., *An Introduction to Bayesian Networks* (UCL Press, London) 1996.
[3] MacKay D. J. C., *Information Theory, Inference and Learning Algorithms* (Cambridge University Press) 2003.
[4] Opper M. and Saad D., *Advanced Mean Field Methods: Theory and Practice* (MIT Press, Cambridge, Mass.) 2001.
[5] Kabashima Y. and Saad D., *Europhys. Lett.*, **44** (1998) 668.
[6] Yedidia J. S., Freeman W. T. and Weiss Y., *Adv. Neural Inf. Process. Syst.*, **13** (2000) 698.
[7] Kabashima Y., *J. Phys. A*, **36** (2003) 11111.
[8] Mézard M., Parisi G. and Virasoro M. A., *Spin Glass Theory and Beyond* (World Scientific, Singapore) 1987.
[9] Nishimori H., *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, UK) 2001.
[10] Mézard M., Parisi G. and Zecchina R., *Science*, **297** (2002) 812.
[11] Mézard M. and Zecchina R., *Phys. Rev. E*, **66** (2002) 056126.
[12] Verdú S., *Multiuser Detection* (Cambridge University Press, UK) 1998.
[13] Saad D. and Rattray M., *Phys. Rev. Lett.*, **79** (1997) 2578.

([1])After we have submitted this paper for publication, two related postings have appeared by Yoshiyuki Kabashima of TITECH, cond-mat/0506311 and cs.IT/0506062. However, in the absence of details we cannot compare our results to those obtained by his approach.