

Efficient Bayesian Inference for Learning in the Ising Linear Perceptron and Signal Detection in CDMA

Juan P. Neirotti and David Saad

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

Abstract

Efficient new Bayesian inference technique is employed for studying critical properties of the Ising linear perceptron and for signal detection in Code Division Multiple Access (CDMA). The approach is based on a recently introduced message passing technique for densely connected systems. Here we study both critical and non-critical regimes. Results obtained in the non-critical regime give rise to a highly efficient signal detection algorithm in the context of CDMA; while in the critical regime one observes a first order transition line that ends in a continuous phase transition point. Finite size effects are also studied.

PACS numbers: 89.70.+c, 75.10.Nr, 64.60.Cn

I. INTRODUCTION

Efficient inference in large complex systems is a major challenge with significant implications in science, engineering and computing. Exact inference is computationally hard in complex systems and a range of approximation methods have been devised over the years, many of which have been originated in the physics literature. A recent review [1] highlights the links between the various approximation methods and their applications.

In the current paper, we extend a method that was introduced only recently [2] for inference in dense graphs using message passing techniques. The method has been employed previously only in the non-critical regime [3], and is used here for studying both critical and non-critical regimes. We apply the method to two different but related problems: signal detection in CDMA and learning in the Ising linear perceptron (ILP).

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 channel of zero mean and variance σ_0^2 . In the scenario of a Code Division Multiple Access (CDMA) system [4], the signal from each user is modulated by a randomly chosen spreading code of length N ; these signals are added up and sent through a noisy channel to the receiving station, which extracts the original message from the received signal using knowledge of the user's spreading codes.

We consider the large-system limit, in which the number of users K tends to infinity while the system load $\beta \equiv K/N \sim \mathcal{O}(1)$. We focus on a CDMA system using binary phase shift keying symbols and will assume the power is completely controlled to unit energy. The received aggregated, modulated and corrupted signal is then of the form:

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

where b_k is the bit transmitted by user k , $s_{\mu k}$ the spreading chip value, n_μ the Gaussian noise variable drawn from $\mathcal{N}(0, 1)$, and y_μ the received message. This process is reminiscent of the learning task performed by a perceptron with binary weights and linear output.

The perceptron is a network which sums a single layer of inputs $s_{\mu j}$, each weighed by a corresponding synaptic weight b_j ; the cumulative contribution is an argument of some

transfer function $g(\cdot)$ that gives rise to the output y_μ

$$y_\mu = g\left(\frac{1}{\sqrt{K}} \sum_{j=1}^K s_{\mu j} b_j\right). \quad (2)$$

A normalisation factor $K^{-1/2}$ is included to make the argument of the transfer function of $\mathcal{O}(1)$. If the entries of \mathbf{b} are ± 1 the perceptron is termed Ising perceptron. When the transfer function is the identity, the perceptron is referred to as linear [5].

The similarity between the linear perceptron of Eq. (2) and the CDMA detection problem of Eq. (1) allows for a direct relation between the two problems to be established. The main difference between the problems is the regime of interest. While CDMA detection applications are of interest mainly for non-critical low load values, ILP studies focused on the critical regime. We consider both regimes in this paper, but to unify the treatment we will use the notation and scaling conventions of the CDMA system.

II. MESSAGE PASSING

Graphical models (Bayes belief networks) provide a powerful framework for modelling statistical dependencies between variables [6–8]. They play an essential role in devising a principled probabilistic framework for inference in a broad range of applications.

Message passing techniques are guaranteed to converge to the globally correct estimate in graphical models that can be represented by a sparse graph with a few (typically long) loops. There are no such guarantees for systems with loops even in the case of large loops and a local tree-like structure (although see [9]). A clear link has been established between certain message passing algorithms and methods of statistical mechanics [1, 10, 11].

In a recent development, we presented a new approach [3], which was inspired by both, the extension of Belief Propagation (BP) to tackle densely connected graphs [2] and that of the replica-symmetric-equivalent BP to Survey Propagation (SP) [12].

The systems we consider here are characterised by multiplicity of pure states and a possible fragmentation of the space of solutions. To address the inference problem in such cases we consider an ensemble of replicated systems where averages are taken over the ensemble of potential solutions. This amounts to the presentation of a new graph, where the observables y_μ are linked to variables in all replicated systems, namely $\mathbf{B} = (\mathbf{b}^1, \mathbf{b}^2, \dots, \mathbf{b}^n)$; where $\mathbf{b}^a = (b_1^a, b_2^a, \dots, b_K^a)^\top$. To estimate the parameters \mathbf{B} given the data $\mathbf{y} = (y_1, y_2, \dots, y_N)^\top$, in

a Bayesian framework, we have to maximise the posterior $P(\mathbf{B}|\mathbf{y}) \propto \prod_{\mu=1}^N P(y_\mu|\mathbf{B}) P(\mathbf{B})$, where we have considered independent data, and thus $P(\mathbf{y}|\mathbf{B}) = \prod_{\mu=1}^N P(y_\mu|\mathbf{B})$.

The likelihood so defined is of a general form; the explicit expression depends on the particular problem studied. Here, we are interested in cases where $\mathbf{b} \in \{\pm 1\}^K$ is an unbiased vector and $P(\mathbf{B}) = 2^{-Kn}$. The estimate we would like to obtain is the maximiser of the posterior marginal (MPM) $\hat{\mathbf{b}}_k = \text{argmax}_{\mathbf{b}_k \in \{\pm 1\}^n} \text{Tr}_{\{\mathbf{b}_{l \neq k}\}} P(\mathbf{B}|\mathbf{y})$, which is expected to be a vector with equal entries for all replica $\hat{b}_k^1 = \hat{b}_k^2 = \dots = \hat{b}_k^n$. The number of operations required to obtain the full MPM estimator is of $\mathcal{O}(2^K)$ which is infeasible for large K values.

For calculating the posterior in the case of both CDMA and the ILP, we use the explicit dependency of y_μ on \mathbf{b}_k from Eqs. (1) and (2) $y_\mu = \sum_{l=1}^K \varepsilon_{\mu l} b_l + \sigma n_\mu$, where σ is a free parameter of the model, to be optimised later in the process; it reflects our ignorance of the true noise parameter σ_0 . The variable n_μ is drawn from $\mathcal{N}(0, 1)$ and the $\varepsilon_{\mu l}$ are small enough to ensure that $\sum_{l=1}^K \varepsilon_{\mu l} b_l \sim \mathcal{O}(1)$. For facilitating the derivation that follows we also define the variable $\Delta_\mu \equiv \sum_{l=1}^K \varepsilon_{\mu l} \mathbf{b}_l = \sum_{l \neq k} \varepsilon_{\mu l} \mathbf{b}_l + \varepsilon_{\mu k} \mathbf{b}_k = \Delta_{\mu k} + \varepsilon_{\mu k} \mathbf{b}_k$, representing the uncorrupted signal. Subsequently, the likelihood can be expanded to take the form

$$\begin{aligned} P(y_\mu|\mathbf{B}) &\propto \prod_{a=1}^n \exp \left[-\frac{1}{2\sigma^2} (y_\mu - \Delta_\mu^a)^2 \right] \\ &\simeq \int \prod_{a=1}^n \left(d\Delta_{\mu k}^a \exp \left[-\frac{1}{2\sigma^2} (y_\mu - \Delta_{\mu k}^a)^2 \right] \left[1 + \varepsilon_{\mu k} \frac{(y_\mu - \Delta_{\mu k}^a)}{\sigma^2} b_k^a \right] \right) P(\Delta_{\mu k}) . \end{aligned} \quad (3)$$

Using Bayes rule one obtains the BP equations

$$P^{t+1}(y_\mu|\mathbf{b}_k, \{y_{\nu \neq \mu}\}) = \text{Tr}_{\{\mathbf{b}_{l \neq k}\}} P(y_\mu|\mathbf{B}) \prod_{l \neq k} P^t(\mathbf{b}_l | \{y_{\nu \neq \mu}\}) \quad (4)$$

$$P^t(\mathbf{b}_l | \{y_{\nu \neq \mu}\}) \propto \prod_{\nu \neq \mu} P^t(y_\nu | \mathbf{b}_l, \{y_{\sigma \neq \nu}\}) . \quad (5)$$

An explicit expression for the inter-dependency between solutions is required for obtaining a closed set of update equations. We assume a dependence of the form $P^t(\mathbf{b}_k | \{y_{\nu \neq \mu}\}) \propto \exp \{ \mathbf{h}_{\mu k}^t \mathbf{b}_k + \frac{1}{2} \mathbf{b}_k^T \mathbf{Q}_{\mu k}^t \mathbf{b}_k \}$, where $\mathbf{h}_{\mu k}^t$ is a vector representing an external field and $\mathbf{Q}_{\mu k}^t$ the matrix of cross-replica interaction. We expect the free energy obtained from the well behaved distribution P^t to be self-averaging, thus we assume the following symmetry between replica $(\mathbf{Q}_{\mu k}^t)^{ab} = (1 - \delta^{ab}) g_{\mu k}^t/n$ and $(\mathbf{h}_{\mu k}^t)^a = h_{\mu k}^t$, where both $h_{\mu k}^t$ and $g_{\mu k}^t$ are of $\mathcal{O}(1)$. An

expression for P^t immediately follows

$$P^t(\mathbf{b}_k | \{y_{\nu \neq \mu}\}) = \frac{\int_{-\infty}^{\infty} dx \exp \left\{ -n \frac{(x - h_{\mu k}^t)^2}{2g_{\mu k}^t} + x \sum_{a=1}^n b_k^a \right\}}{\int_{-\infty}^{\infty} dx \exp \{ -n \Phi(x; h_{\mu k}^t, g_{\mu k}^t) \}}, \quad (6)$$

where $\Phi(x; h_{\mu k}^t, g_{\mu k}^t) = (x - h_{\mu k}^t)^2 / 2g_{\mu k}^t - \ln(2 \cosh(x))$. We exploit the assumption that the number of replica n is large and employ Laplace's method to find dominant contributions to the integral. The function $\Phi(x; h_{\mu k}^t, g_{\mu k}^t)$ exhibits two minima if $h_{\mu k}^t \rightarrow 0$ and $g_{\mu k}^t > 1$; these will provide the only contributions in that limit. Other regimes will provide trivial solutions. If the field $h_{\mu k}^t$ goes to zero as $m_{\mu k}^t h_{\mu k}^t \sim \ln(4n(n_{\mu k}^t)^{-2})/2n$, where $m_{\mu k}^t$ is the spontaneous magnetisation and $n_{\mu k}^t$ a constant, the first two moments of b_k^a are, up to $\mathcal{O}(n^{-1})$,

$$\begin{aligned} \langle b_k^a \rangle &= \text{Tr}_{\{\mathbf{b}_k\}} P^t(\mathbf{b}_k | \{y_{\nu \neq \mu}\}) b_k^a \simeq \left[1 - \frac{(n_{\mu k}^t)^2}{2n} \right] m_{\mu k}^t \\ \langle b_k^a b_l^b \rangle - \langle b_k^a \rangle \langle b_l^b \rangle &\simeq \delta_{kl} \left\{ \delta^{ab} \left[1 - (m_{\mu k}^t)^2 \right] + (1 - \delta^{ab}) \frac{(n_{\mu k}^t m_{\mu k}^t)^2}{n} \right\}. \end{aligned}$$

However, for calculating the posterior we need the distribution on the variable $\Delta_{\mu k}^a$, which is a sum of a large number of unbiased and uncorrelated random variables $\varepsilon_{\mu k}$ and b_k^a . Therefore, by virtue of the central limit theorem, the variable $\Delta_{\mu k}^a = \sum_{l \neq k} \varepsilon_{\mu l} b_l^a$ obeys a normal distribution, whose mean value and covariance matrix are given by

$$(\mathbf{u}_{\mu k}^t)^a \equiv \langle \Delta_{\mu k}^a \rangle = \text{Tr}_{\{\mathbf{b}_{l \neq k}\}} \prod_{l \neq k} P^t(\mathbf{b}_l | \{y_{\nu \neq \mu}\}) \sum_{l \neq k} \varepsilon_{\mu l} b_l^a = \sum_{l \neq k} \varepsilon_{\mu l} m_{\mu l}^t \quad (7)$$

$$\begin{aligned} (\mathbf{r}_{\mu k}^t)^{ab} &\equiv \langle \Delta_{\mu k}^a \Delta_{\mu k}^b \rangle - \langle \Delta_{\mu k}^a \rangle \langle \Delta_{\mu k}^b \rangle = \text{Tr}_{\{\mathbf{b}_{l \neq k}\}} \prod_{l \neq k} P^t(\mathbf{b}_l | \{y_{\nu \neq \mu}\}) \sum_{\substack{l \neq k \\ j \neq k}} \varepsilon_{\mu l} \varepsilon_{\mu j} b_l^a b_j^b - \left(\sum_{l \neq k} \varepsilon_{\mu l} m_{\mu l}^t \right)^2 \\ &= \sum_{l \neq k} \varepsilon_{\mu l}^2 \{ \langle b_l^a b_j^b \rangle - \langle b_l^a \rangle \langle b_j^b \rangle \} = \delta^{ab} (X_{\mu k} - Q_{\mu k}^t) - (1 - \delta^{ab}) \frac{1}{n} R_{\mu k}^t, \end{aligned} \quad (8)$$

where $X_{\mu k} \equiv \sum_{l \neq k} \varepsilon_{\mu l}^2$, $Q_{\mu k}^t \equiv \sum_{l \neq k} (\varepsilon_{\mu l} m_{\mu l}^t)^2$ and $R_{\mu k}^t \equiv \sum_{l \neq k} (\varepsilon_{\mu l} n_{\mu l}^t m_{\mu l}^t)^2$ are macroscopic variables of $\mathcal{O}(1)$. In particular, $R_{\mu k}^t$ is a free variable that can be used to optimise with respect to a given performance measure. This corresponds to fine tuning of the variational model considered. All three quantities $X_{\mu k}$, $Q_{\mu k}^t$ and $R_{\mu k}^t$ are self averaging so we can drop both indices μ and k . The probability of $\Delta_{\mu k}$ can be expressed as $P(\Delta_{\mu k}) \propto \exp \left\{ -\frac{1}{2} (\Delta_{\mu k} - \mathbf{u}_{\mu k}^t)^T (\mathbf{r}_{\mu k}^t)^{-1} (\Delta_{\mu k} - \mathbf{u}_{\mu k}^t) \right\}$.

Having the probability distribution of $\Delta_{\mu k}$ we can express the message from nodes y_μ to nodes b_k^a at time $t + 1$ explicitly, using Eqs. (3), (4) and (5)

$$\hat{m}_{\mu k}^{t+1} = \frac{\text{Tr}_{\{\mathbf{b}_k\}} b_k^a P^{t+1}(y_\mu | \mathbf{b}_k, \{y_{\nu \neq \mu}\})}{\text{Tr}_{\{\mathbf{b}_k\}} P^{t+1}(y_\mu | \mathbf{b}_k, \{y_{\nu \neq \mu}\})} = \frac{\varepsilon_{\mu k}}{\sigma^2 + \beta - Q^t + R^t} (y_\mu - u_{\mu k}^t). \quad (9)$$

It is then straightforward to prove the following equation and its approximation, the latter due to the fact that $\hat{m}_{\nu k}^t \sim \mathcal{O}(\varepsilon_{\nu k})$

$$m_{\mu k}^t = \tanh \left(\sum_{\nu \neq \mu} \text{arctanh}(\hat{m}_{\nu k}^t) \right) \simeq \tanh \left(\sum_{\nu \neq \mu} \hat{m}_{\nu k}^t \right). \quad (10)$$

To study the quality of the inferred vectors one considers the gauged field with respect to the true message $b_k h_{\mu k}^t$ where $h_{\mu k}^t \equiv \text{artanh}(m_{\mu k}^t) = \sum_{\nu \neq \mu} \text{artanh}(\hat{m}_{\nu k}^t) \simeq \sum_{\nu \neq \mu} \hat{m}_{\nu k}^t$. The distribution of this field is likely to be well approximated by a Gaussian, as a result of the central limit theorem, whose mean and variance are E^t and F^t respectively

$$E^t = \frac{1}{K} \sum_{k=1}^K \sum_{\mu=1}^N b_k \hat{m}_{\mu k}^t, \quad F^t \simeq \frac{1}{K} \sum_{k=1}^K \sum_{\mu=1}^N (\hat{m}_{\mu k}^t)^2; \quad (11)$$

both are assumed to be independent of the index μ due to self-averaging. For the same reason we expect the macroscopic variables, representing the overlap between the vectors \mathbf{m}_μ and \mathbf{b} at any time t and the squared length of \mathbf{m}_μ , defined as $M_\mu^t \equiv \sum_{k=1}^K b_k m_{\mu k}^t / K \simeq \sum_{k=1}^K b_k \hat{m}_{\mu k}^t / K = M^t$ and $N_\mu^t \equiv \sum_{k=1}^K (m_{\mu k}^t)^2 / K \simeq \sum_{k=1}^K (\hat{m}_{\mu k}^t)^2 / K = N^t$, to be μ independent. Using the distribution we obtained for the gauged field $b_k h_{\mu k}^t$, both variables can be evaluated by $M^t = \int \mathcal{D}u \tanh(\sqrt{F^t}u + E^t)$, $N^t = \int \mathcal{D}u \tanh^2(\sqrt{F^t}u + E^t)$, where $\mathcal{D}u = \exp(-u^2/2)/\sqrt{2\pi}$. Applying a method equivalent to the EM algorithm [13] for the independent parameter of the model σ^2 we have that the optimal selection of the parameter is given by the condition $E^t = F^t$, which also implies that $N^t = M^t$. Notice that this result is not surprising as it maximises the normalised overlap between the vectors \mathbf{m}_μ and \mathbf{b} .

It is important to notice at this point the different scaling factors used in the two models we examine. For CDMA one uses $\varepsilon_{\mu k} = s_{\mu k}/\sqrt{N}$ while $\varepsilon_{\mu k} = s_{\mu k}/\sqrt{K}$ is used for the (ILP). Imposing the condition $E^t = F^t$ leads to a relation between the structure of the space of solutions, represented by R^t , and the free parameter of the model σ^2 . From Eqs. (11) one obtains for the two models $E^{t+1} = e_1^{-1} [\sigma^2 + R^t + e_2 (1 - N^t)]^{-1}$, $F^{t+1} = e_1 [\sigma_0^2 + e_2 (1 - N^t)] (E^{t+1})^2$, where $e_1 = 1$ (β) and $e_2 = \beta$ (1) for the CDMA (ILP) system. This implies, after simplification, that for both cases $R^t = \sigma_0^2 - \sigma^2$.

Despite the simplicity of this result, the process from which we obtained it provides a mechanism for estimating the true noise variance. In deriving E^t and F^t we used the fact that $K, N \rightarrow \infty$ with $K/N = \beta$. So that the true noise variance σ_0^2 that appears in the expression for F^t has been obtained from a signal vector with an infinite number of entries y_μ . Thus $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\mu=1}^N (y_\mu)^2 = e_2 + \sigma_0^2$. Using this we can express the message as

$$\hat{m}_{\mu k}^{t+1} \simeq \varepsilon_{\mu k} \left[\frac{1}{N} \sum_{\mu=1}^N (y_\mu)^2 - e_2 N^t \right]^{-1} (y_\mu - u_{\mu k}^t), \quad (12)$$

where no prior belief of the noise level σ_0 is required.

The steady state equation for the macroscopic variable E^t is obtained in the limit $t \rightarrow \infty$, leading to the definition of $\overline{E} \equiv \lim_{t \rightarrow \infty} E^t$. In this regime the following relation holds

$$\overline{E}(\sigma_0^2, \beta) = e_1^{-1} \left\{ \sigma_0^2 + e_2 \left[1 - \int \mathcal{D}u \tanh^2 \left(\sqrt{\overline{E}(\sigma_0^2, \beta)} u + \overline{E}(\sigma_0^2, \beta) \right) \right] \right\}^{-1}. \quad (13)$$

From these expressions one can calculate directly the error per bit rate

$$\overline{P}_b(\sigma_0^2, \beta) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\sqrt{\frac{\overline{E}(\sigma_0^2, \beta)}{2}} \right) \right). \quad (14)$$

III. NUMERICAL RESULTS

The inference algorithm requires an iterative update of Eqs. (10) and (12) until they converge to a reliable estimate of the signal. We emphasise again that there is no need for prior information on the noise level. The computational complexity of the algorithm, as it has been presented here, is of $\mathcal{O}(NK^2)$ but can be reduced to be $\mathcal{O}(K^2)$ in a similar way to the approach taken in [2].

To test the performance of our algorithm we carried out a set of experiments of CDMA signal detection under typical conditions. Error probability of the inferred signals has been calculated for a system load of $\beta = 0.25$, where the true noise level is $\sigma_0^2 = 0.25$ and the estimated noise is $\sigma^2 = 0.01$, as shown in Figure 1(a). 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 of Ref. [2].

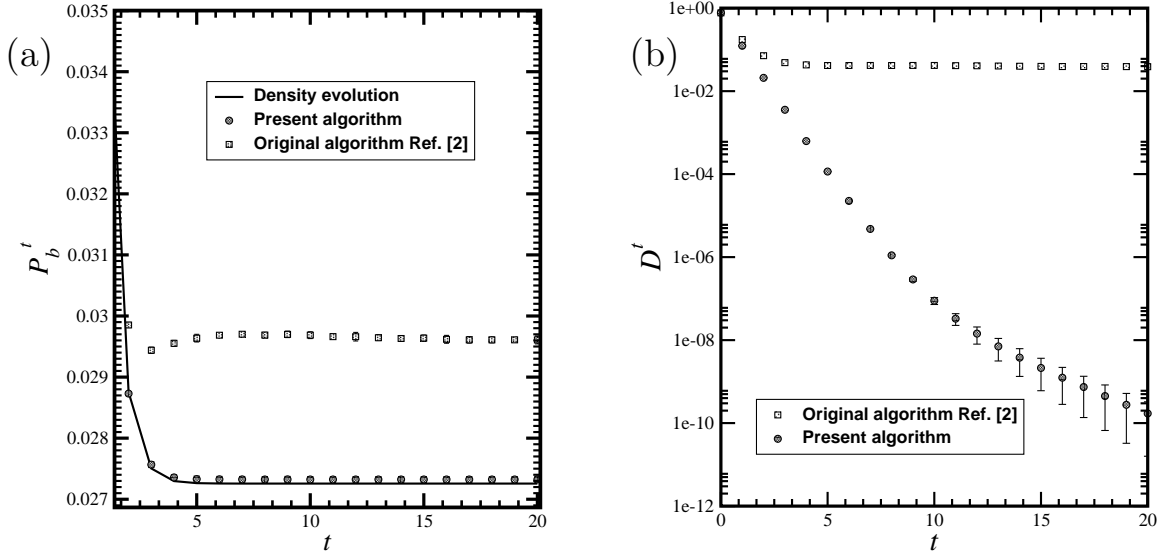


FIG. 1: (a) Error probability of the inferred solution as a function of time. The system load $\beta=0.25$, true and estimated noise levels $\sigma_0^2=0.25$ and $\sigma^2=0.01$, respectively. Squares represent results obtained by the algorithm of [2], solid line the dynamics obtained from our equations; 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).

Another performance measure to be consider is $D^t \equiv K^{-1}(\mathbf{m}^t - \mathbf{m}^{t-1}) \cdot (\mathbf{m}^t - \mathbf{m}^{t-1})$, that provides an indication to the stability of the solutions obtained. In Fig. 1(b) we compare results obtained from our algorithm, that exhibit fast convergence to a reliable solution, in stark contrast to the original algorithm [2] which does not converge.

For the ILP, the $K > N$ regime is highly interesting as the system develops a critical behaviour for a range of (σ_0^2) values. We carried out a set of experiments for this system (the CDMA scaling was kept for consistency) based on density evolution. In Fig. 2(a) we present curves of \bar{P}_b , defined in Eq. (14), as a function of the inverse load β^{-1} for different values of σ_0^2 . Three different regimes have been observed: For $\sigma_0^2 < 0.15$ the curves exhibit a discontinuity at a value of β that varies with σ_0^2 (first order phase transition-like behaviour). At $\sigma_0^2=0.15$ the curve becomes continuous but its slope diverges (second order phase transition-like behaviour). The \bar{P}_b curves show analytical behaviour for noise values above 0.15. In Fig. 2b we present a phase diagram of the CDMA system. It shows the dependency of the critical load β_C^{-1} as a function of the noise parameter. The first order line

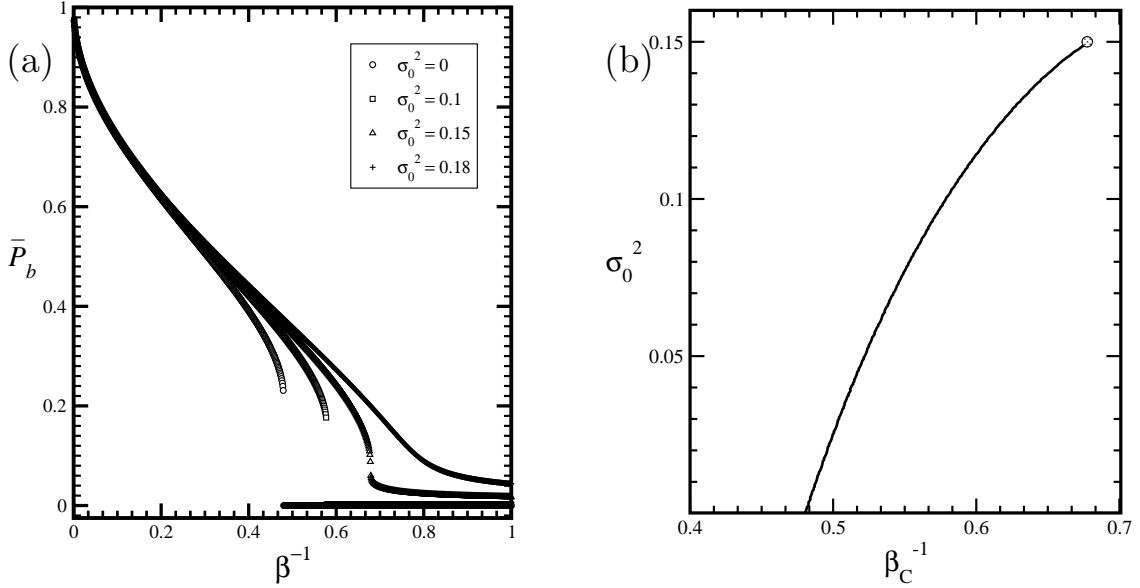


FIG. 2: (a) \bar{P}_b at the steady state, Eq. (14), as a function of β^{-1} for different values of the noise parameter. For values of σ_0^2 below 0.15 the curves show discontinuity at certain β values, which becomes continuous but non-analytic at $\sigma_0^2 = 0.15$ around $\beta^{-1} \simeq 0.68$. For noise variance values above $\sigma_0^2 = 0.15$ the curves become analytical. (b) Position of the non analyticity of the error rate curve β_C^{-1} as a function of the noise parameter σ_0^2 . This first order phase transition-like curve ends in a second order phase transition-like point marked by \circ .

ends in a second order transition point marked by a circle.

Another indication for the critical behaviour is the number of steps required for the recursive update of Eq. (13) to convergence. In Fig. 3(a) we present the number of iterations needed to reach a steady state as a function of β^{-1} when the noise parameter is set to $\sigma_0^2 = 0.10$. The number of iterations diverge when the critical value of β is reached.

Finally, we wish to explore the efficiency of the algorithm as a function of the system size. In Fig. 3(b) we present the result of iterating Eqs. (10) and (12) for system sizes of $K = 200, 400, 800, 1600$ and 3200 . The curves represent mean values over 1000 experiments. There is a strong dependency of the error per bit rate on the size of the system, which is expected to converge to the asymptotic limit (infinite system size) represented by the solid line.

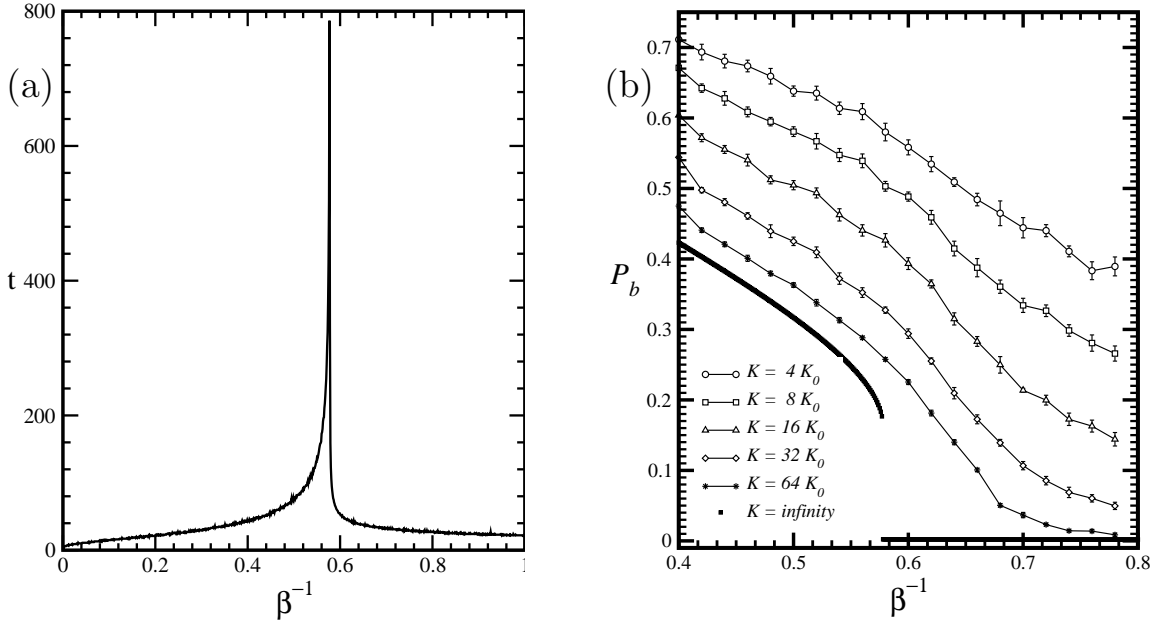


FIG. 3: (a) Number of iterations of Eq. (13) required for convergence as a function of β , for $\sigma_0^2 = 0.10$. the error rate curve exhibits a discontinuity. (b) Finite size effects observed in the error rate curve when the Eqs. (10) and (12) are iterated over the number of steps needed to reach the steady state. The noise level used is $\sigma_0^2 = 0.10$ with $K_0 = 50$. The curves are mean values over 1000 experiments. The curve obtained from the iteration of the steady state equations is presented as a reference.

IV. CONCLUSIONS

In summary, we employed a new variational algorithm based on replicated variable systems to investigate two related problems: signal detection in CDMA and learning in the ILP. The new algorithm facilitates the use of message passing techniques in densely connected systems, even in systems that show a fragmented solution space and represents an extension of existing algorithms similar to the extension of BP to SP.

Results on the CDMA signal detection problem are superior than other existing algorithms [2, 16], without using any prior for the expected noise level.

Results have also been obtained for low and intermediate load levels under various noise conditions, which are of higher relevance to ILP learning than to CDMA. These exhibit a first-order like transition for critical load levels and below a certain noise level ($\sigma_0^2 < 0.15$), that become second order as the noise level increases (at $\sigma_0^2 = 0.15$). No transition points have been identified above this noise level. Finally, we also examined finite size effects in the system, which are clearly present even at a system size of 3200 nodes.

We are in the process of examining the suitability of the method for other applications [15]. While the approach seems promising, there is clearly a need for further research to fully determine the potential of the new algorithm.

Acknowledgments

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

-
- [1] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice*, MIT Press, Cambridge, MA (2001).
 - [2] Y. Kabashima, J. Phys. A **36**, 11111 (2003).
 - [3] J.P. Neirotti and D. Saad, Europhys. Lett. **71**, 866 (2005).
 - [4] S. Verdú, *Multiuser Detection*, Cambridge University Press UK (1998).
 - [5] H. S. Seung, H. Sompolinsky and N. Tishby, Phys. Rev. A **45**, 6056 (1992).
 - [6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publishers, San Francisco, CA (1988).
 - [7] F.V. Jensen, *An Introduction to Bayesian Networks*, UCL Press, London (1996).
 - [8] D.J.C. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press (2003).
 - [9] Y. Weiss, *Neural Computation* **12**, 1 (2000).
 - [10] Y. Kabashima, D. Saad, Europhys. Lett. **44**, 668 (1998).
 - [11] J.S. Yedidia, W.T. Freeman and Y. Weiss, in *Advances in Neural Information Processing Systems* **13**, 698 (2000).
 - [12] M. Mézard, G. Parisi and R. Zecchina, Science **297**, 812 (2002).
 - [13] A.P. Dempster, N. M. Laird and D.B. Rubin, J. Roy. Stat. Soc. B **39**, 1 (1977).
 - [14] Y. Kabashima, cs.IT/0506062 (2005).
 - [15] J.P. Neirotti and D. Saad, in preparation (2005).
 - [16] Comparison to the recently presented signal detection algorithm of Ref. [14] cannot be carried out due to absence of numerical data.