Improved message passing for inference in densely connected systems

Juan P. Neirotti and David Saad

The Neural Computing Research Group, Aston University, Birmingham B4 7ET, UK
E-mail: {j.p.neirotti}{d.saad}aston.ac.uk, tel:+44 121 2043648, fax:+44 121 2043685

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, by considering average messages from a large number of replicated systems. We present an application of the problem 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.

PACS No: 89.70.+c, 75.10.Nr, 64.60.Cn
Short title: Survey propagation for densely connected systems
Key words: Statistical physics, inference, message passing

1 Introduction

Graphical models (Bayes belief networks) provide a powerful framework for modelling statistical dependencies between variables [1, 2, 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 with a few (typically long) loops. They are aimed at obtaining (pseudo) posterior estimates for 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 even in the case of large loops and a local tree-like structure (although message passing techniques have been used successfully in loopy systems, supported by some limited theory [4]). A clear link has been established between certain message passing algorithms and well known methods of statistical mechanics [5] such as the Bethe approximation [6, 7].

These inherent limitations seem to prevent the use of message passing techniques in densely connected systems due to their high connectivity, implying an exponentially growing cost (with the connectivity), and an exponential number of loops that render the method inconsistent. However, an exciting new approach has been recently suggested [8] for extending Belief Propagation (BP) techniques [1, 2, 3] to densely connected systems. In this approach, messages are grouped together, giving rise to a macroscopic random variable, drawn from a Gaussian distribution of varying mean and variance for each of the nodes. This enables one to approximate local probability values in feasible time scales, by iteratively updating a set of newly-defined macroscopic variables. The technique has been successfully applied to signal detection in Code Division Multiple Access (CDMA) problems and the results reported are competitive with those of other state of the art techniques. However, the current approach has some inherent limitations [8], presumably due to its similarity to the replica symmetric solution in equivalent Ising spin models [9, 10] and the existence of multiple competing solutions.

In a separate recent development [11], the replica-symmetric-equivalent BP has been extended to Survey Propagation (SP), which corresponds to one-step replica symmetry breaking in diluted systems.

This new algorithm, motivated by the theoretical physics interpretation of such problems, has been highly successful in solving hard computational problem [11], far beyond other existing approaches. This approach relies on considering message averages over a large number of replicated systems and a decimation process that guides the process towards commonly-preferred specific solutions. The SP algorithm also facilitated theoretical studies of the corresponding physical system and contributed to our understanding of it [12] as well as of related systems [13].

Inspired by the extension of BP to SP we have extended the approach of [8], designed for inference in densely connected systems, in a similar manner to include an average over multiple pure states. The approach relies on averaging over messages in a large number of replicated systems given common observables.

The paper is organised as follows, in section 2 we present the general formalism, followed by a specific application (CDMA) in section 3. We will conclude the paper with a brief discussion on the applicability of the method to a range of inference problems in a densely connected systems.

2 General formalism

The aim of this work is to develop an efficient algorithm for obtaining solutions in a general inference problem that can be mapped onto a dense graph. This refers to obtaining variable estimates that maximise the posterior distribution of K variables $\mathbf{b} = (b_1, b_2, \ldots, b_K)$ given N data (observables) $\mathbf{y} = (y_1, y_2, \ldots, y_N)$. Using Bayes rule we can rewrite the posterior to be maximised as

$$P(\mathbf{b}|\mathbf{y}) = \frac{\prod_{\mu=1}^{N} P(y_\mu|\mathbf{b})\, P(\mathbf{b})}{\underset{\{\mathbf{b}_{l\neq k}\}}{\mathrm{Tr}} \prod_{\mu=1}^{N} P(y_\mu|\mathbf{b})\, P(\mathbf{b})} , \tag{1}$$

where data is assumed to be independent, so that $P(\mathbf{y}|\mathbf{b}) = \prod_{\mu=1}^{N} P(y_\mu|\mathbf{b})$. Clearly the explicit expression for the likelihood depends on the particular problem studied. Here we will look at cases where $\mathbf{b}$ is an unbiased vector in $\{\pm 1\}^K$ and $P(\mathbf{b}) = 2^{-K}$. The estimates one would like to obtain are based on the marginal posterior maximiser (MPM)

$$\hat{b}_k = \underset{b_k \in \{\pm\}}{\mathrm{argmax}} \underset{\{\mathbf{b}_{l\neq k}\}}{\mathrm{Tr}} P(\mathbf{b}|\mathbf{y}) \ . \tag{2}$$

The number of operations required to obtain the MPM estimator is of order $\mathcal{O}\left(2^{K-1}\right)$ which is computationally infeasible where K is large.

To approximate the MPM estimator in problems that can be mapped onto a densely connected graph one may employ a message passing technique such as BP [1, 3] that works effectively and efficiently, with a computational complexity that grows linearly with the system size, in cases characterised by a contiguous space of solutions. In these cases, the approximate estimates slowly converge towards a single solution. However, BP is based on local updates and ignores long range correlations that emerge with the increase in the number of constraints in the system, or in other cases when there is a mismatch between the prior used for the variables and the true values. Typically in these cases, multiple solutions emerge and conflicting messages are being passed between variables leading to non-convergence.

To overcome the problem, it has been suggested [11] to extend the method such that estimates are based on averages over a large number of replicated systems, each with its own messages, given a common set of observables (data). These averaged messages, in conjunction with a decimation process that directs the search towards solutions that are favoured by the majority of the replicated systems, form the powerful SP algorithm [11].

However, we are interested in the application of BP to problems that can be mapped onto densely connected graphs, similar to the one suggested in [8]. Similarly to BP, this method works very well when the space of solutions is contiguous and the algorithm converges to a single solution and is bound to fail, as has been observed, when the solution space becomes fragmented; for instance, when there is a mismatch between the assumed and true noise levels.

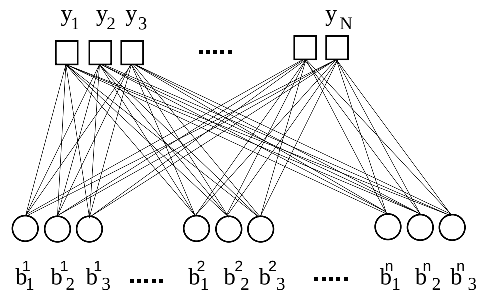

Figure 1: Replicated variable systems $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, .., \mathbf{b}_K)$ given data $\mathbf{y} = (y_1, y_2, \ldots, y_N)$.

We adopt a similar approach to that of SP by considering messages from n replicated systems in the presence of common data. Figure 1 shows the detection problem we aim to solve as a bipartite graph where $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_K)$ the set of replicated (binary) variable vectors, $\mathbf{b}_k = \left(b_k^1, b_k^2, \ldots, b_k^n\right)$, where n is the solution (replica) index.

Using Bayes rule one straightforwardly obtains the extended BP equations

$$P^{t+1}\left(y_\mu|\mathbf{b}_k, \{y_{\nu\neq\mu}\}\right) = \widehat{a}_{\mu k}^{t+1} \underset{\{\mathbf{b}_{l\neq k}\}}{\mathrm{Tr}} P\left(y_\mu|\mathbf{B}\right) \prod_{l\neq k} P^t\left(\mathbf{b}_l|\{y_{\nu\neq\mu}\}\right) \tag{3}$$

$$P^t\left(\mathbf{b}_l|\{y_{\nu\neq\mu}\}\right) = a_{\mu k}^t \prod_{\nu\neq\mu} P^t\left(y_\nu|\mathbf{b}_l, \{y_{\sigma\neq\nu}\}\right) \tag{4}$$

where $\widehat{a}_{\mu k}^{t+1}$ and $a_{\mu k}^t$ are normalisation constants. For calculating the posterior of Eq.(1) we have to assume a certain form of data dependence on the variables. The dependence used here is still quite general,

$$y_\mu = \mathcal{F}\left(\sum_{l=1}^{K} \varepsilon_{\mu l} b_l^{\mathrm{a}}; \boldsymbol{\gamma}\right) , \tag{5}$$

where $\boldsymbol{\gamma}$ are some parameters that characterise the probabilistic dependency (e.g., the noise model) and $\varepsilon_{\mu l} \ll \mathcal{O}(1)$ are small variables as is typically assumed in densely connected systems (the exact scaling of these variables will be discussed later).

Thus, we can expand for the likelihood with respect to $\varepsilon_{\mu l}$ using the new variables

$$\Delta_{\mu k}^{\mathrm{a}} - \sum_{l\neq k} \varepsilon_{\mu l} b_l^{\mathrm{a}}$$

that correspond to the first argument of $\mathcal{F}()$ of equation (5) after omitting contribution of variables of index k,

$$\begin{aligned}
\prod_{\mathrm{a}=1}^{n} P\left(y_\mu|\mathbf{b}^{\mathrm{a}}\right) P\left(\mathbf{b}^{\mathrm{a}}\right) &= \int \left[\prod_{\mathrm{a}=1}^{n} \mathrm{d}\Delta_{\mu k}^{\mathrm{a}} \delta\left(\Delta_{\mu k}^{\mathrm{a}} - \sum_{l\neq k} \varepsilon_{\mu l} b_l^{\mathrm{a}}\right)\right] P\left(y_\mu|\boldsymbol{\Delta}_{\mu k}, \mathbf{b}_k; \boldsymbol{\gamma}\right) P\left(\boldsymbol{\Delta}_{\mu k}\right) P\left(\mathbf{b}_k\right) \\
&\simeq \int \left[\prod_{\mathrm{a}=1}^{n} \mathrm{d}\Delta_{\mu k}^{\mathrm{a}}\ \delta\left(\Delta_{\mu k}^{\mathrm{a}} - \sum_{l\neq k} \varepsilon_{\mu l} b_l^{\mathrm{a}}\right)\right] \\
&\qquad \left[P\left(y_\mu|\boldsymbol{\Delta}_{\mu k}; \boldsymbol{\gamma}\right) + \varepsilon_{\mu k} \nabla_{\boldsymbol{\Delta}_{\mu k}} P\left(y_\mu|\boldsymbol{\Delta}_{\mu k}; \boldsymbol{\gamma}\right) \cdot \mathbf{b}_k\right] P\left(\boldsymbol{\Delta}_{\mu k}\right) P\left(\mathbf{b}_k\right) \ .
\end{aligned} \tag{6}$$

2.1 Structure of the solutions

An explicit expression for inter-dependence between solutions is required for obtaining a closed set of update equations. We assume a dependence of the form

$$P^t\left(\mathbf{b}_k|\{y_{\nu\neq\mu}\}\right) \propto \exp\left\{\mathbf{h}_{\mu k}^{t\mathsf{T}}\, \mathbf{b}_k + \frac{1}{2} \mathbf{b}_k^{\mathsf{T}} \mathbf{Q}_{\mu k}^t\, \mathbf{b}_k\right\} , \tag{7}$$

where $\mathbf{h}^t_{\mu k}$ is a vector representing an external field and $\mathbf{Q}^t_{\mu k}$ the matrix of cross-replica interaction. Furthermore, we assume the following replica symmetry ansatz

$$\begin{aligned}\left(\mathbf{Q}^t_{\mu k}\right)^{\mathrm{ab}} &= \delta^{\mathrm{ab}}\, q^t_{\mu k} + \left(1-\delta^{\mathrm{ab}}\right)\, p^t_{\mu k} \qquad (8)\\ \mathbf{h}^t_{\mu k} &= h^t_{\mu k}\mathbf{u}. \qquad (9)\end{aligned}$$

where $\mathbf{u}^\mathsf{T} := \overbrace{(1,\, 1,\, \cdots,\, 1)}^{n}$. An expression for equation (7) immediately follows

$$P^t\left(\mathbf{b}_k|\left\{y_{\nu\neq\mu}\right\}\right) = [\mathcal{Z}^t_{\mu k}]^{-1}\exp\left\{h^t_{\mu k}\sum_{\mathrm{a}=1}^{n} b^{\mathrm{a}}_k + \frac{1}{2}p^t_{\mu k}\left(\sum_{\mathrm{a}=1}^{n} b^{\mathrm{a}}_k\right)^2\right\}\ , \qquad (10)$$

where $\mathcal{Z}^t_{\mu k}$ is a normalisation constant.

We expect the free energy obtained from the well behaved distribution P^t to be self-averaging, such that

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\overline{\mathcal{Z}^t_{\mu k}}\right) = \lim_{n\to\infty}\frac{1}{n}\log\left(\mathcal{Z}^t_{\mu k}\left(h^t_{\mu k}, q_0, p_0\right)\right)\ ,$$

where the sub-index 0 represents the mean value of the parameters for the corresponding distributions and the over-line represents the mean value of the partition function over such distributions.

To obtain the scaling behaviour of the various parameters we calculate $\mathcal{Z}\left(h,q,p\right)$ explicitly, assuming the parameters q and p are taken from normal distributions $\mathcal{N}_q\left(q_0,\sigma^2_q\right)$ and $\mathcal{N}_p\left(p_0,\sigma^2_p\right)$. The partition function takes the form (dropping the super-index t and sub-indices μ and k)

$$\mathcal{Z}\left(h,q,p\right) = \exp\left\{\frac{n}{2}\left(q-p\right) + n\ln(2)\right\}\frac{1}{\sqrt{2\pi p}}\int_{-\infty}^{\infty}\mathrm{d}x\,\exp\left(-\frac{\left(x-h\right)^2}{2p} + n\ln\left(\cosh(x)\right)\right). \qquad (11)$$

The mean value of the partition function over the set of parameters becomes

$$\overline{\mathcal{Z}\left(h,q,p\right)} = \int\mathcal{D}_q\int\mathcal{D}_p\,\mathcal{Z}\left(h,\, q,\, p\right)\ ,$$

where $\mathcal{D}_{q(p)} = \mathrm{d}q(p)\,\mathcal{N}_{q(p)}\left(q_0(p_0),\, \sigma^2_{q(p)}\right)$ denotes a Gaussian integration measure with the respective mean and variance. After a straightforward but tedious calculation one obtains

$$\begin{aligned}\overline{\mathcal{Z}\left(h,q,p\right)} &= \sum_{\mathrm{a}=1}^{n}\binom{n}{\mathrm{a}}\exp\left\{n\left[-h\left(1-\frac{2\mathrm{a}}{n}\right) + \frac{q_0}{2} + p_0\left[\left(1-\frac{2\mathrm{a}}{n}\right)^2 - \frac{1}{2n}\right]n\right.\right.\\ &+ \left.\left.\frac{\sigma^2_q}{8}n + \frac{\sigma^2_p}{2}\left[\left(1-\frac{2\mathrm{a}}{n}\right)^2 - \frac{1}{2n}\right]^2 n^3\right]\right\}\\ &\simeq \sqrt{\frac{2}{\pi}}\mathcal{A}(n)\exp\left\{n\left[\ln(2) + |h| + \frac{q_0}{2} + p_0 n + \frac{\sigma^2_q}{8}n + \frac{\sigma^2_p}{2}n^3\right]\right\}, \qquad (12)\end{aligned}$$

where $\mathcal{A}(n)\sim\mathcal{O}(1)$. The scaling properties of the various variables immediately follow: $h\sim\mathcal{O}\left(1\right)$, $q_0\sim\mathcal{O}\left(1\right)$, $p_0\sim\mathcal{O}\left(n^{-1}\right)$, $\sigma^2_q\sim\mathcal{O}\left(n^{-1}\right)$, and $\sigma^2_p\sim\mathcal{O}\left(n^{-3}\right)$. For brevity and transparency we will change notation for the off-diagonal elements $\left(\mathbf{Q}^t_{\mu k}\right)^{a\neq b}\equiv g^t_{\mu k}n^{-1}$, where $g^t_{\mu k}\sim\mathcal{O}\left(1\right)$.

The diagonal elements define the zero of the energy, so we can take them equal to zero with no loss of generality.

The marginalised posterior (10) at time t then reduces to

$$P^t\left(\mathbf{b}_k|\left\{y_{\nu\neq\mu}\right\}\right) = \frac{\displaystyle\int_{-\infty}^{\infty}\mathrm{d}x\,\exp\left\{-n\frac{\left(x-h^t_{\mu k}\right)^2}{2g^t_{\mu k}} + x\sum_{\mathrm{a}=1}^{n} b^{\mathrm{a}}_k\right\}}{\displaystyle\int_{-\infty}^{\infty}\mathrm{d}x\,\exp\left\{-n\mathit{\Phi}\left(x;h^t_{\mu k},g^t_{\mu k}\right)\right\}}, \qquad (13)$$

where

$$\Phi\left(x; h_{\mu k}^t, g_{\mu k}^t\right) = -\frac{\left(x - h_{\mu k}^t\right)^2}{2 g_{\mu k}^t} + \ln\left(\cosh(x)\right). \tag{14}$$

As we plan on taking the limit $n \to \infty$, equation (13) can be solved using the saddle point method. More specifically, we study the extrema points of the function $\Phi(x; h, g)$ to identify the dominant values as described in Appendix A. From this study we find that the only non trivial solution, of two maxima, is obtained when $g_{\mu k}^t > 1$ and $\left|h_{\mu k}^t\right| / g_{\mu k}^t \ll 1$.

Using the marginalised distribution (13) one can then obtain mean values of solutions in each of the replicated systems as well as cross-replica correlations

$$\begin{aligned} \langle b_k^{\mathrm{a}} \rangle &= \operatorname*{Tr}_{\{\mathbf{b}_k\}} P^t\left(\mathbf{b}_k \middle| \{y_{\nu \neq \mu}\}\right) b_k^{\mathrm{a}} \\ &= \frac{\displaystyle\int_{-\infty}^{\infty} \mathrm{d}x \, \exp\left\{-n\Phi\left(x; h_{\mu k}^t, g_{\mu k}^t\right)\right\} \tanh(x)}{\displaystyle\int_{-\infty}^{\infty} \mathrm{d}x \, \exp\left\{-n\Phi\left(x; h_{\mu k}^t, g_{\mu k}^t\right)\right\}} \end{aligned} \tag{15}$$

and

$$\begin{aligned} \left\langle b_k^{\mathrm{a}} b_k^{\mathrm{b}} \right\rangle &= \operatorname*{Tr}_{\{\mathbf{b}_k\}} P^t\left(\mathbf{b}_k \middle| \{y_{\nu \neq \mu}\}\right) b_k^{\mathrm{a}} b_k^{\mathrm{b}} \\ &= \delta^{\mathrm{ab}} + \left(1 - \delta^{\mathrm{ab}}\right) \frac{\int_{-\infty}^{\infty} \mathrm{d}x \, \exp\left\{-n\Phi\left(x; h_{\mu k}^t, g_{\mu k}^t\right)\right\} \tanh(x)^2}{\int_{-\infty}^{\infty} \mathrm{d}x \, \exp\left\{-n\Phi\left(x; h_{\mu k}^t, g_{\mu k}^t\right)\right\}} . \end{aligned} \tag{16}$$

Cross replica and cross site averages factorise such that

$$\left\langle b_k^{\mathrm{a}} b_l^{\mathrm{b}} \right\rangle = \langle b_k^{\mathrm{a}} \rangle \left\langle b_l^{\mathrm{b}} \right\rangle . \tag{17}$$

It would also be useful to define the magnetisation

$$\left|m_{\mu k}^t\right| \equiv \tanh\left(\left|x_{0,\mu k}^t\right|\right) = \frac{\left|x_{0,\mu k}^t\right|}{g_{\mu k}^t} ,$$

to simplify some of the following equations.

Assuming that $g_{\mu k}^t > 1$ and $\left|h_{\mu k}^t\right| / g_{\mu k}^t \ll 1$ (see Appendix A for details), and exploiting the relation $\operatorname{sgn}\left(m_{\mu k}^t\right) = \operatorname{sgn}\left(h_{\mu k}^t\right)$, we find that the positions of the two peaks of $\Phi\left(x; h_{\mu k}^t, g_{\mu k}^t\right)$ are nearly symmetric, located at

$$x_{\pm,\mu k}^t \simeq \pm g_{\mu k}^t \left|m_{\mu k}^t\right| + \frac{1}{1 - g_{\mu k}^t \left[1 - \left(m_{\mu k}^t\right)^2\right]} h_{\mu k}^t. \tag{18}$$

If n is large and the field is small, the mean values in equations (15) and (16) can be expressed as:

$$\langle b_k^{\mathrm{a}} \rangle \simeq a_{+,\mu k}^t \tanh\left(x_{+,\mu k}^t\right) + a_{-,\mu k}^t \tanh\left(x_{-,\mu k}^t\right) \tag{19}$$

$$\left\langle b_k^{\mathrm{a}} b_k^{\mathrm{b}} \right\rangle \simeq a_{+,\mu k}^t \tanh\left(x_{+,\mu k}^t\right)^2 + a_{-,\mu k}^t \tanh\left(x_{-,\mu k}^t\right)^2 , \tag{20}$$

where

$$a_{\pm,\mu k}^t \simeq \frac{\exp\left\{\pm n \left|m_{\mu k}^t\right| h_{\mu k}^t\right\}}{\exp\left\{n m_{\mu k}^t h_{\mu k}^t\right\} + \exp\left\{-n m_{\mu k}^t h_{\mu k}^t\right\}} . \tag{21}$$

Thus, if the covariance matrix

$$\begin{aligned} \left(\mathbf{\Psi}_{\mu k}^t\right)^{\mathrm{ab}} &\equiv \left\langle b_k^{\mathrm{a}} b_k^{\mathrm{b}} \right\rangle - \langle b_k^{\mathrm{a}} \rangle \left\langle b_k^{\mathrm{b}} \right\rangle \\ &\simeq 4 a_{+,\mu k}^t a_{-,\mu k}^t \left(m_{\mu k}^t\right)^2 \end{aligned} \tag{22}$$

has nonzero entries at small field, we can choose the behaviour of the small field to be

$$m_{\mu k}^{t} h_{\mu k}^{t} \simeq \frac{1}{2n} \ln\left(4n/\left(n_{\mu k}^{t}\right)^{2}\right), \tag{23}$$

where the $n_{\mu k}^{t}$ are free parameters. The dominant peak is located at the position that shares a common sign with the field. So, in the large n limit, $a_{\mathrm{sgn}\left(h_{\mu k}^{t}\right),\mu k}^{t} \to 1 - \frac{\left(n_{\mu k}^{t}\right)^{2}}{4n}$ and the product $a_{+,\mu k}^{t} a_{-,\mu k}^{t} \to \frac{\left(n_{\mu k}^{t}\right)^{2}}{4n}$ as desired. We also obtain the relation

$$\langle b_{k}^{\mathrm{a}} \rangle = m_{\mu k}^{t}. \tag{24}$$

If the $\varepsilon_{\mu k}$ and b_{k}^{a} are unbiased variables, the variable $\Delta_{\mu k}^{\mathrm{a}} = \sum_{l \neq k} \varepsilon_{\mu l} b_{l}^{\mathrm{a}}$, by virtue of the central limit theorem, obeys a normal distribution, with mean value and covariance matrix given by

$$\begin{aligned}
\left(\mathbf{u}_{\mu k}^{t}\right)^{\mathrm{a}} &\equiv \left\langle \Delta_{\mu k}^{\mathrm{a}} \right\rangle \\
&= \underset{\{\mathbf{b}_{l \neq k}\}}{\mathrm{Tr}} \prod_{l \neq k} P^{t}\left(\mathbf{b}_{l} \middle| \{y_{\nu \neq \mu}\}\right) \sum_{l \neq k} \varepsilon_{\mu l} b_{l}^{\mathrm{a}} \\
&= \sum_{l \neq k} \varepsilon_{\mu l} m_{\mu l}^{t}
\end{aligned} \tag{25}$$

$$\begin{aligned}
\left(\boldsymbol{\Upsilon}_{\mu k}^{t}\right)^{\mathrm{ab}} \equiv \left\langle \Delta_{\mu k}^{\mathrm{a}} \Delta_{\mu k}^{\mathrm{b}} \right\rangle - \left\langle \Delta_{k}^{\mathrm{a}} \right\rangle \left\langle \Delta_{k}^{\mathrm{b}} \right\rangle &= \underset{\{\mathbf{b}_{l \neq k}\}}{\mathrm{Tr}} \prod_{l \neq k} P^{t}\left(\mathbf{b}_{l} \middle| \{y_{\nu \neq \mu}\}\right) \sum_{\substack{l \neq k \\ j \neq k}} \varepsilon_{\mu l} \varepsilon_{\mu j} b_{l}^{\mathrm{a}} b_{j}^{\mathrm{b}} - \left(\sum_{l \neq k} \varepsilon_{\mu l} m_{\mu l}^{t}\right)^{2} \\
&= \delta^{\mathrm{ab}} \sum_{l \neq k} \varepsilon_{\mu l}^{2} \left(1 - \left(m_{\mu l}^{t}\right)^{2}\right) - \left(1 - \delta^{\mathrm{ab}}\right) \frac{1}{n} \sum_{l \neq k} \left(\varepsilon_{\mu l} n_{\mu l}^{t} m_{\mu l}^{t}\right)^{2} \\
&= \delta^{\mathrm{ab}} \left(X_{\mu k} - Q_{\mu k}^{t}\right) - \left(1 - \delta^{\mathrm{ab}}\right) \frac{1}{n} R_{\mu k}^{t},
\end{aligned} \tag{26}$$

where $X_{\mu k} \equiv \sum_{l \neq k} \varepsilon_{\mu l}^{2}$, $Q_{\mu k}^{t} \equiv \sum_{l \neq k} \left(\varepsilon_{\mu l} m_{\mu l}^{t}\right)^{2}$ and $R_{\mu k}^{t} \equiv \sum_{l \neq k} \left(\varepsilon_{\mu l} n_{\mu l}^{t} m_{\mu l}^{t}\right)^{2}$ are macroscopic variables. It is important to notice that $R_{\mu k}^{t}$ are free variables that can be used to optimise the inference with respect to a given performance measure.

Given the covariance matrix values one can write down an explicit expression of the probability $P\left(\boldsymbol{\Delta}_{\mu k}\right)$

$$\begin{aligned}
P\left(\boldsymbol{\Delta}_{\mu k}\right) &= \frac{1}{\sqrt{(2\pi)^{n} \left|\boldsymbol{\Upsilon}_{\mu k}^{t}\right|}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\Delta}_{\mu k} - \mathbf{u}_{\mu k}^{t}\right)^{\mathsf{T}} \left(\boldsymbol{\Upsilon}_{\mu k}^{t}\right)^{-1} \left(\boldsymbol{\Delta}_{\mu k} - \mathbf{u}_{\mu k}^{t}\right)\right\} \\
&= \sqrt{\frac{n}{(2\pi)^{n+1}} \left[R_{\mu k}^{t} \left(X_{\mu k} - Q_{\mu k}^{t}\right)^{n-2}\right]^{-1}} \exp\left\{-\frac{n}{2} \frac{\left(u_{\mu k}^{t}\right)^{2}}{X_{\mu k} - Q_{\mu k}^{t} + R_{\mu k}^{t}}\right\} \\
&\times \int \mathrm{d}\omega \exp\left\{-\frac{n}{2} \frac{X_{\mu k} - Q_{\mu k}^{t}}{R_{\mu k}^{t}} \left(X_{\mu k} - Q_{\mu k}^{t} + R_{\mu k}^{t}\right) \omega^{2}\right\} \\
&\times \prod_{\mathrm{a}=1}^{n} \exp\left\{-\frac{\left(\Delta_{\mu k}^{\mathrm{a}}\right)^{2}}{2\left(X_{\mu k} - Q_{\mu k}^{t}\right)} + \left(\omega + \frac{u_{\mu k}^{t}}{X_{\mu k} - Q_{\mu k}^{t} + R_{\mu k}^{t}}\right) \Delta_{\mu k}^{\mathrm{a}}\right\}.
\end{aligned} \tag{27}$$

2.2 Messages

From the original BP equations (3) and (4) and with using the explicit expression of the probability $P(\mathbf{\Delta}_{\mu k})$ Eq.(27) in Eq.(7) we can express the message from nodes y_μ to nodes b_k^{a} at time $t+1$

$$\begin{aligned} \widehat{m}_{\mu k}^{t+1} &= \mathop{\mathrm{Tr}}_{\{\mathbf{b}_k\}} P^{t+1}\left(y_\mu|\mathbf{b}_k, \{y_{\nu\neq\mu}\}\right) b_k^{\tilde{\mathrm{a}}} \\ &= \frac{\mathop{\mathrm{Tr}}_{\{\mathbf{B}\}} \prod_{\mathrm{a}=1}^{n} P\left(y_\mu|\mathbf{b}^{\mathrm{a}}\right) P\left(\mathbf{b}^{\mathrm{a}}\right) \prod_{l\neq k} P\left(\mathbf{b}_l|\{y_{\nu\neq\mu}\}\right) b_k^{\tilde{\mathrm{a}}}}{\mathop{\mathrm{Tr}}_{\{\mathbf{B}\}} \prod_{\mathrm{a}=1}^{n} P\left(y_\mu|\mathbf{b}^{\mathrm{a}}\right) P\left(\mathbf{b}^{\mathrm{a}}\right) \prod_{l\neq k} P\left(\mathbf{b}_l|\{y_{\nu\neq\mu}\}\right)} . \end{aligned} \tag{28}$$

After somewhat lengthy derivation, described in Appendix B, one obtains the following expression for the message update $\widehat{m}_{\mu k}^{t+1}$

$$\widehat{m}_{\mu k}^{t+1} = \varepsilon_{\mu k} \frac{\int \left(\prod_{l\neq k} \mathrm{d}x_l\right) \exp\left\{-n\mathcal{H}_{\mu k}^t\left(\boldsymbol{x}_{l\neq k}\right)\right\} \frac{\partial}{\partial z} \ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=v_{\mu k}}}{\int \left(\prod_{l\neq k} \mathrm{d}x_l\right) \exp\left\{-n\mathcal{H}_{\mu k}^t\left(\boldsymbol{x}_{l\neq k}\right)\right\}} \tag{29}$$

where $v_{\mu k}\left(\boldsymbol{x}_{l\neq k}\right) \equiv \sum_{l\neq k} \varepsilon_{\mu l} \tanh\left(x_l\right)$, and

$$\mathcal{H}_{\mu k}^t\left(\boldsymbol{x}_{l\neq k}\right) = \sum_{l\neq k} \left[\frac{x_l^2}{2g_{\mu l}^t} - \ln\cosh\left(x_l\right)\right] + \frac{\left(u_{\mu k}^t - v_{\mu k}\right)^2}{2\left(X_{\mu k} - Q_{\mu k}^t + R_{\mu k}^t\right)} - \ln P\left(y_\mu|v_{\mu k};\boldsymbol{\gamma}\right) .$$

In the large n limit, only the solutions $\tilde{\boldsymbol{x}}_{l\neq k}$ of $\nabla_{\boldsymbol{x}_{l\neq k}}\mathcal{H}_{\mu k}^t\left(\boldsymbol{x}_{l\neq k}\right) = \mathbf{0}$, that correspond to the lowest minimum, contribute to the integral. The expression for the message is therefore reduced to

$$\widehat{m}_{\mu k}^{t+1} = \varepsilon_{\mu k} \frac{\partial}{\partial z} \ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=v_{\mu k}(\tilde{\boldsymbol{x}}_{l\neq k})}. \tag{30}$$

The expression for the messages from $\mathbf{b}$-nodes to $\mathbf{y}$-nodes is derived in a similar manner (details again in Appendix C) to obtain

$$m_{\mu k}^t = \mathop{\mathrm{Tr}}_{\{\mathbf{b}_k\}} b_k^{\tilde{\mathrm{a}}} P^t\left(\mathbf{b}_k|\{y_{\nu\neq\mu}\}\right) = \tanh\left(\sum_{\nu\neq\mu} \mathrm{arctanh}\left(\widehat{m}_{\nu k}^t\right)\right) . \tag{31}$$

2.3 Obtaining solutions

To solve Eq.(29) we employ again the saddle point method, as $n\to\infty$, obtaining a set of equations to be solved

$$\begin{aligned} \frac{\partial}{\partial x_l}\mathcal{H}_{\mu k}^t\left(\boldsymbol{x}_{l\neq k}\right) &= \frac{\tilde{x}_l}{g_{\mu l}^t} - \tanh\left(\tilde{x}_l\right) - \varepsilon_{\mu l}\left[1-\tanh^2\left(\tilde{x}_l\right)\right] \\ &\times \left[\frac{\left(u_{\mu k}^t - \tilde{v}_{\mu k}\right)}{\left(X_{\mu k} - Q_{\mu k}^t + R_{\mu k}^t\right)} + \frac{\partial}{\partial z} \ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=\tilde{v}_{\mu k}}\right] = 0 , \end{aligned} \tag{32}$$

where $\tilde{v}_{\mu k} = v_{\mu k}\left(\tilde{\boldsymbol{x}}_{l\neq k}\right)$. To find the solution of Eq.(32) we will define the value $\overline{v}_{\mu k}$ such that

$$0 = \frac{\left(u_{\mu k}^t - \overline{v}_{\mu k}\right)}{\left(X_{\mu k} - Q_{\mu k}^t + R_{\mu k}^t\right)} + \frac{\partial}{\partial z} \ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=\overline{v}_{\mu k}} . \tag{33}$$

After some calculations, detailed in Appendix D, this leads to a solution of the form

$$\begin{aligned}
\tilde{v}_{\mu k} &\simeq u^t_{\mu k} - \frac{(\tilde{v}_{\mu k} - \overline{v}_{\mu k})}{\left(X_{\mu k} - Q^t_{\mu k} + R^t_{\mu k}\right)} W^t_{\mu k} \\
&= \frac{X_{\mu k} - Q^t_{\mu k} + R^t_{\mu k}}{X_{\mu k} - Q^t_{\mu k} + R^t_{\mu k} + W^t_{\mu k}} u^t_{\mu k} + \frac{W^t_{\mu k}}{X_{\mu k} - Q^t_{\mu k} + R^t_{\mu k} + W^t_{\mu k}} \overline{v}_{\mu k} \ ,
\end{aligned} \tag{34}$$

where

$$W^t_{\mu k} = \sum_{l \neq k} \varepsilon^2_{\mu l} \frac{g^t_{\mu l}\left[1 - \left(m^t_{\mu l}\right)^2\right]}{1 - g^t_{\mu l}\left[1 - \left(m^t_{\mu l}\right)^2\right]} \left[1 - \left(X_{\mu k} - Q^t_{\mu k} + R^t_{\mu k}\right) \frac{\partial^2}{\partial z^2} \ln P\left(y_\mu | z; \boldsymbol{\gamma}\right)|_{z=\lambda \overline{v}_{\mu k} + (1-\lambda)\tilde{v}_{\mu k}}\right] \ .$$

A recursion is needed to determine both λ (from the first derivative of Eq.(61)) and $\tilde{v}_{\mu k}$. Equation (34) represents the value of v at which the expression for the message Eq.(30) has to be evaluated.

3 Application: CDMA

A number of inference problems in densely connected systems can be tackled by this approach. We focus here on the CDMA detection problem as it is the only densely connected system we are aware of that had been studied previously using a message passing algorithm [8]. This will enable us to demonstrate the potential of our method and its relation to the algorithm of [8], mirroring the extension of BP to SP.

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 [14]. 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 and will assume the power is 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 spreading chip value, n_μ is the Gaussian noise variable drawn from $\mathcal{N}(0,1)$, and y_μ the received message. According to previous notation, it holds that $\varepsilon_{\mu k} = s_{\mu k}/\sqrt{N}$. The goal is to get an accurate estimate of the vector $\mathbf{b}$ for all users given the received message vector $\mathbf{y}$ by approximating the posterior $P(\mathbf{b}|\mathbf{y})$. An expression representing the likelihood is required and is easily derived from the noise model (assuming zero mean and variance σ^2)

$$P\left(y_\mu \,|\, \mathbf{B}\right) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left\{ -\frac{\left(\mathbf{y}_\mu - \boldsymbol{\Delta}_\mu\right)^{\mathsf{T}} \left(\mathbf{y}_\mu - \boldsymbol{\Delta}_\mu\right)}{2\sigma^2} \right\}, \tag{35}$$

where $\mathbf{y}_\mu = y_\mu \mathbf{u}$ and $\mathbf{u}^{\mathsf{T}} \equiv \overbrace{(1, 1, \cdots, 1)}^{n}$.

To calculate correlation between replica we expand $P\left(y_\mu \,|\, \mathbf{B}\right)$ in the large N limit (Eq. 35), as shown in Eq.(7). Following the model described in the previous chapters, one can map the CDMA macroscopic variables onto the parameters used previously relationships:

$$u_{\mu k}^t = \frac{1}{\sqrt{N}} \sum_{l \neq k} s_{\mu l} m_{\mu l}^t \tag{36}$$

$$X_{\mu k} \simeq \frac{K}{N} \equiv \beta \tag{37}$$

$$Q_{\mu k}^t = \frac{\beta}{K} \sum_{l \neq k} \left(m_{\mu l}^t\right)^2 \tag{38}$$

$$R_{\mu k}^t = \frac{\beta}{K} \sum_{l \neq k} \left(n_{\mu l}^t m_{\mu l}^t\right)^2 \tag{39}$$

$$D_{\mu k}^t \equiv \frac{\beta}{K} \sum_{l \neq k} \frac{g_{\mu l}^t \left[1 - \left(m_{\mu l}^t\right)^2\right]}{1 - g_{\mu l}^t \left[1 - \left(m_{\mu l}^t\right)^2\right]} . \tag{40}$$

The mean value of b_k^{a} at time $t+1$ is then given by

$$\widehat{m}_{\mu k}^{t+1} = A^t \left(R^t, Q^t, \beta, \sigma^2\right) \left(\frac{y_\mu \mathbf{s}_\mu}{\sqrt{N}} - \beta \left(\mathbf{P}_\mu - \mathbf{I}/K\right) \mathbf{m}_\mu^t \right)_k \tag{41}$$

where $\mathbf{P}_\mu \equiv (1/K) s_{\mu k} s_{\mu l})$ and $\mathbf{I} \equiv \delta_{kl}$, respectively, and

$$A^t \left(R^t, Q^t, \beta, \sigma^2\right) = \frac{\beta - Q^t + R^t + D^t}{\sigma^2 \left(\beta - Q^t + R^t\right) + \left(\sigma^2 + \beta - Q^t + R^t\right) D^t} .$$

We assume that the macroscopic variables are self averaging and omit the μ, k indices.

The main difference between Eq.(41) and the equivalent equation in [8] is the dependency of the pre-factor A^t on R^t, reflecting correlations between different solutions groups (replica). To determine this term we optimise the choice of R^t by minimising the bit error at each time step. To find the optimal choice of the $n_{\mu k}^t$ appears to be difficult. Instead we will choose an appropriate R^t that minimises the error in the iterative calculation of the macroscopic variables. Following [8] one defines M^t and writes the following expressions for both M^t and Q^t

$$M^t = \frac{1}{NK} \sum_{\mu=1}^{N} \sum_{k=1}^{K} b_k m_{\mu k}^t = \int \mathcal{D}z \, \tanh\left(\sqrt{F^t} z + E^t\right) \tag{42}$$

$$\frac{Q^t}{\beta} = \frac{1}{NK} \sum_{\mu=1}^{N} \sum_{k=1}^{K} \left(b_k m_{\mu k}^t\right)^2 = \int \mathcal{D}z \, \tanh^2\left(\sqrt{F^t} z + E^t\right), \tag{43}$$

where $\mathcal{D}z \equiv \mathrm{d}z \, \exp\left[-z^2/2\right]/\sqrt{2\pi}$ and

$$E^{t+1} \equiv \frac{1}{K} \sum_{\mu=1}^{N} \sum_{k=1}^{K} b_k \widehat{m}_{\mu k}^{t+1} = A^t \left(R^t, Q^t, \beta, \sigma^2\right) \tag{44}$$

$$F^{t+1} \equiv \sum_{\mu=1}^{N} \left[\frac{1}{K} \sum_{k=1}^{K} \left(b_k \widehat{m}_{\mu k}^{t+1}\right)^2 - \frac{1}{K^2} \left(\sum_{k=1}^{K} b_k \widehat{m}_{\mu k}^{t+1} \right)^2 \right] \tag{45}$$

$$\simeq \frac{1}{K} \sum_{\mu=1}^{N} \widehat{\mathbf{m}}_\mu^{t+1} \cdot \widehat{\mathbf{m}}_\mu^{t+1} = \left[\beta - 2\beta M^t + Q^t + \sigma_0^2\right] \left(E^{t+1}\right)^2 . \tag{46}$$

The function to be optimised is the bit error rate

$$P_b^t \equiv \frac{1}{2K} \sum_{k=1}^{K} \left(b_k - \mathrm{sgn}\left(m_k^t\right)\right) = \int_{-\infty}^{-E^t/\sqrt{F^t}} \mathcal{D}z \tag{47}$$

and

$$m_k^t \simeq \tanh\left(\sum_{\mu=1}^{N} \widehat{m}_{\mu k}^t\right) \ . \tag{48}$$

To determine R^t we proceeded as follows. First we have to consider that the quotient

$$E^t/\sqrt{F^t} = \left[\beta - 2\beta M^{t-1} + Q^{t-1} + \sigma_0^2\right]^{-1}$$

is a function of R^{t-2} through M^{t-1} and Q^{t-1}. To optimise it we have to find the roots of

$$\frac{\partial P_b^t}{\partial\left(R^{t-2}\right)} = -\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{\beta - 2\beta M^{t-1} + Q^{t-1} + \sigma_0^2}\right\} \frac{\partial}{\partial\left(R^{t-2}\right)} \frac{E^t}{\sqrt{F^t}} = 0,$$

which implies that

$$\begin{aligned} 2\frac{\partial M^{t-1}}{\partial\left(R^{t-2}\right)} - \frac{1}{\beta}\frac{\partial Q^{t-1}}{\partial\left(R^{t-2}\right)} &= 2\frac{\partial E^{t-1}}{\partial R^{t-2}} \int_{-\infty}^{+\infty} \frac{\mathrm{d}z}{\sqrt{2\pi F^{t-1}}} \exp\left\{-\frac{\left(z-E^{t-1}\right)^2}{2F^{t-1}}\right\} z \\ &\times \left[1-\tanh^2(z)\right]\left[1-\tanh(z)\right] = 0 \ . \end{aligned} \tag{49}$$

If the argument of the integral in Eq.(49) is an odd function, the integral is zero. The argument of the integral at any time t, $\Psi\left(z; E^t, F^t\right)$ can be expressed as

$$\Psi\left(z; E^t, F^t\right) = \mathcal{E}\left(z; E^t, F^t\right) z \left[1-\tanh^2(z)\right]\left[1-\tanh(z)\right] \ .$$

Assuming that $\Psi\left(z; E^t, F^t\right) = -\Psi\left(-z; E^t, F^t\right)$ then

$$\mathcal{E}\left(z; E^t, F^t\right) z \left[1-\tanh^2(z)\right]\left[1-\tanh(z)\right] = -\mathcal{E}\left(-z; E^t, F^t\right) (-z) \left[1-\tanh^2(z)\right]\left[1+\tanh(z)\right] \tag{50}$$

$$\text{leading to} \quad \tanh(z) = \frac{\mathcal{E}\left(z; E^t, F^t\right) - \mathcal{E}\left(-z; E^t, F^t\right)}{\mathcal{E}\left(z; E^t, F^t\right) + \mathcal{E}\left(-z; E^t, F^t\right)}, \tag{51}$$

which holds for all functions $\mathcal{E}\left(z; E^t, F^t\right)$ of the form

$$\mathcal{E}\left(z; E^t, F^t\right) = \exp(z)\, \mathcal{G}\left(z; E^t, F^t\right) \ , \tag{52}$$

where $\mathcal{G}\left(z; E^t, F^t\right)$ is an even function of z. In particular, for the Gaussian $\mathcal{N}\left(z; E^t, \sqrt{F^t}\right)$ we have

$$\mathcal{N}\left(z; E^t, \sqrt{F^t}\right) = \frac{1}{\sqrt{2\pi F^t}} \exp\left(\frac{zE^t}{F^t}\right) \exp\left(-\frac{z^2 + E^t}{2F^t}\right)$$

which satisfies Eq.(52) if and only if

$$E^t = F^t \ . \tag{53}$$

In the same way one can find the condition at which $Q^t = \beta M^t$

$$Q^t - \beta M^t = \beta \int_{-\infty}^{+\infty} \frac{\mathrm{d}z}{\sqrt{2\pi F^t}} \exp\left\{-\frac{\left(z-E^t\right)^2}{2F^t}\right\} \tanh(z) \left[1-\tanh(z)\right] \ ,$$

the argument of this integral can be expressed as

$$\Omega\left(z; E^t, F^t\right) = \mathcal{F}\left(z; E^t, F^t\right) \tanh(z) \left[1-\tanh(z)\right] \ .$$

For $\Omega\left(z; E^t, F^t\right)$ be odd in z, $\mathcal{F}\left(z; E^t, F^t\right)$ has to satisfied Eq.(51), so it is of the form

$$\mathcal{F}\left(z; E^t, F^t\right) = \exp(z)\, \mathcal{G}\left(z; E^t, F^t\right) \ , \tag{54}$$

where again $\mathcal{G}\left(z; E^t, F^t\right)$ is an even function of z. For the Gaussian function we obtain again the same requirement (ie Eq.(53)). Setting $E^t = F^t$ we then obtain $M^t = Q^t$.

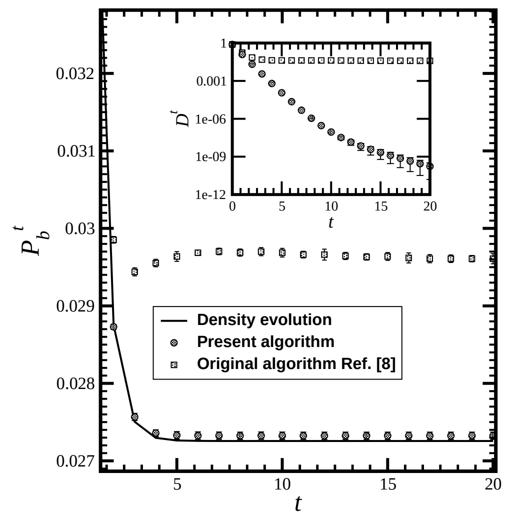

Figure 2: Error probability of the inferred solution evolving in time. The system load $\beta = 0.25$, true noise level $\sigma_0^2 = 0.25$ and estimated noise $\sigma^2 = 0.01$. Squares represent results of the original algorithm [8], solid line the dynamics obtained from our equations; circles represent results obtained from the suggested *practical* algorithm. Variances are smaller than the symbol size. In the inset, D^t, a measure of convergence in the obtained solutions, as a function of time; symbols are as in the main figure.

For the calculation of E^t one has to provide an estimate of the noise (σ_0), but by condition (53) one can compute E^t using Eq.(46). From this equation one obtains:

$$E^{t+1} \simeq \frac{1}{K}\sum_{\mu=1}^{N} \widehat{\mathbf{m}}_\mu^{t+1}\cdot\widehat{\mathbf{m}}_\mu^{t+1} = \left(E^{t+1}\right)^2\left\{\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2 - 2\beta M^t + Q^t\right\} \tag{55}$$

$$= \left(E^{t+1}\right)^2\left\{\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2 - Q^t\right\} = \left\{\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2 - Q^t\right\}^{-1} . \tag{56}$$

Using Eq.(44) we conclude that E^{t+1} is equal to the pre-factor of the RHS of Eq.(41) and obtain a new expression for $\widehat{m}_{\mu k}^{t+1}$

$$\widehat{m}_{\mu k}^{t+1} = A^t\left(\frac{y_\mu \mathbf{s}_\mu}{\sqrt{N}} - \beta\left(\mathbf{P}_\mu - K^{-1}\mathbf{I}\right)\mathbf{m}_\mu^t\right)_k$$
$$A^t \simeq \left\{\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2 - Q^t\right\}^{-1} \tag{57}$$

where no estimate on σ_0 is required at all. This is clearly of great significance for practical CDMA signal detection as no prior knowledge of the channel characteristics is required and there is no risk of a mismatch between the assumed and true noise levels that may lead to errors.

3.1 Results

The inference algorithm requires an iterative update of Eqs.(56,57,48) and converges to a reliable estimate of the signal, with no need for an accurate prior information of the noise level. The computational complexity of the algorithm is of $\mathcal{O}(K^2)$.

To test the performance of our algorithm we carried out a set of experiments of CDMA signal detection problem 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 2. 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 [8].

Another performance measure one should consider is

$$D^t \equiv \frac{1}{K}\left(\mathbf{m}^t - \mathbf{m}^{t-1}\right)\cdot\left(\mathbf{m}^t - \mathbf{m}^{t-1}\right) \ ,$$

that provides an indication to the stability of the solutions obtained. In the inset of Figure 2 we see that results obtained from our algorithm show convergence to a reliable solution in stark contrast to the original algorithm [8]. The physical interpretation of the difference between the two results is assumed to be related to a replica symmetry breaking phenomena.

4 Conclusions

In summary, we present a new approach for using belief propagation in densely connected systems. that enables one to obtain reliable solutions even when the solution space is fragmented. It represents an extension to existing algorithms of that type which is reminiscent to the extension of BP to SP.

The approach is presented in general terms and can potentially be used in a range of problems that can be mapped onto a dense graph. For demonstrating the performance of the algorithm and to compare it with the BP-equivalent algorithm of [8], we have derived explicit expressions for the CDMA signal detection problem.

The algorithm we have obtained for this particular problem on the basis of the general formulation, does not require any prior knowledge of the channel characteristics and is highly applicable. It has been tested on the signal detection problem has showed superior results to other existing algorithms [8, 16].

Further research is required to fully determine the potential of the new approach and its applicability for a variety of problems. Its application to problems with a real noise model (or likelihood term) is rather straightforward although it would depend on the specific type of noise considered. Application to cases with discrete noise models are likely to be more difficult. Specific applications of the approach to other densely connected problems, such as lossy compression are underway.

Acknowledgement

Support from the EU FP-6 EVERGROW IP is gratefully acknowledged.

References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publishers, San Francisco, CA (1988)

[2] F.V. Jensen, *An Introduction to Bayesian Networks*, UCL Press, London (1996)

[3] D.J.C. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press (2003)

[4] Y. Weiss *Neural Computation* **12** 1 (2000)

[5] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice*, MIT Press, Cambridge, MA 2001

[6] Y. Kabashima, D. Saad, Europhys. Lett. **44** 668 (1998)

[7] J.S. Yedidia, W.T. Freeman and Y. Weiss, in *Advances in Neural Information Processing Systems* **13** 698 (2000)

[8] Y. Kabashima, J. Phys. A **36** 11111 (2003)

[9] M. Mézard, G. Parisi and M.A Virasoro, *Spin Glass Theory and Beyond*, World Scientific, Singapore (1987)

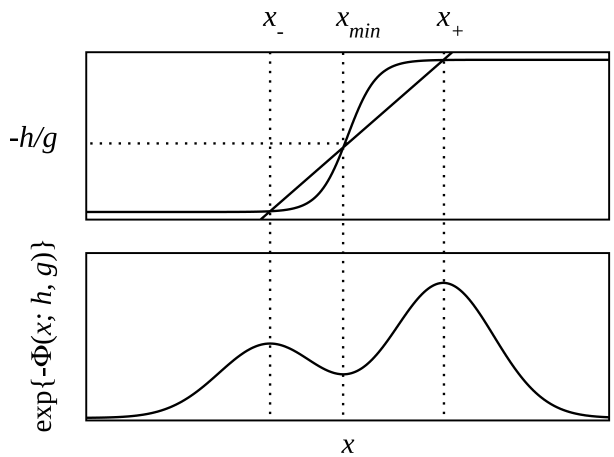

Figure 3: Solutions for the mean field Equation with two maxima and one minimum.

[10] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*, Oxford University Press UK (2001)

[11] M. Mézard, G. Parisi and R. Zecchina, Science **297** 812 (2002)

[12] M. Mézard and R. Zecchina Phys. Rev. E **66** 056126 (2002)

[13] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. Lett. **89** 268701 (2002)

[14] S. Verdú, *Multiuser Detection*, Cambridge University Press UK (1998)

[15] D. Saad and M. Rattray, Phys. Rev. Lett. **79** 2578 (1997)

[16] In preparing the manuscript we have found that a similar activity is being carried out by Yoshiyuki Kabashima at TITECH. However, in the absence of any details we cannot comment on his approach.

A Extrema points of the function $\Phi\,(x; h, g)$

Analysis of the equation

$$\frac{\partial \Phi\,(x; h, g)}{\partial x} = \frac{x-h}{g} - \tanh(x) = 0\ ,$$

shows that the function $\Phi\,(x; h, g)$ admits one or two maxima according to the following table

h	g	Number of max.
$h \in \mathbb{R}$	$0 < g \leq 1$	one max.
$\lvert h\rvert = h_c$	$g > 1$	one max. and one hump
$\lvert h\rvert < h_c$	$g > 1$	two max.

where $h_c = \sqrt{g(g-1)} - \cosh^{-1}\left(\sqrt{g}\right)$. The case of 2 maxima is presented in Fig. 3. Notice also that the variable g plays a similar role to that of the inverse temperature and a spontaneous magnetisation appears below a critical value $g_c = 1$.

B Deriving expressions for the messages - $\widehat{m}_{\mu k}$

We derive expressions for the denominator and numerator of Eq.(29) separately. The denominator $\mathcal{D}$ can be expressed as

$$\begin{aligned}\mathcal{D} \quad = \quad & \int \left(\prod_{l\neq k} \mathrm{d}x_l\right) \exp\left\{-\frac{n}{2}\sum_{l\neq k}\frac{x_l^2}{g_{\mu l}^t}\right\} \int \mathrm{d}\omega \exp\left\{-\frac{n}{2}\frac{X_{\mu k}-Q_{\mu k}^t}{R_{\mu k}^t}\left(X_{\mu k}-Q_{\mu k}^t+R_{\mu k}^t\right)\omega^2\right\} \\ & \int \left(\prod_{\mathrm{a}=1}^{n} \mathrm{d}\Delta_{\mu k}^{\mathrm{a}}\ \mathrm{d}z^{\mathrm{a}}\right) \exp\left\{-\frac{\sum_{\mathrm{a}=1}^{n}\left(\Delta_{\mu k}^{\mathrm{a}}\right)^2}{2\left(X_{\mu k}-Q_{\mu k}^t\right)} + \sum_{\mathrm{a}=1}^{n}\left(\omega + \frac{u_{\mu k}^t}{X_{\mu k}-Q_{\mu k}^t+R_{\mu k}^t} - iz^{\mathrm{a}}\right)\Delta_{\mu k}^{\mathrm{a}}\right\} \\ & \underset{\{\mathbf{b}_{l\neq k}\}}{\mathsf{Tr}} \exp\left\{\sum_{l\neq k}\sum_{\mathrm{a}=1}^{n}\left(i\varepsilon_{\mu l}z^{\mathrm{a}}+x_l\right)b_l^{\mathrm{a}}\right\} \underset{\{\mathbf{b}_k\}}{\mathsf{Tr}} \left[P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right) + \varepsilon_{\mu k}\nabla_{\mathbf{\Delta}_{\mu k}}P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right)\cdot\mathbf{b}_k\right] \ ,\end{aligned}$$

and the numerator $\mathcal{N}$

$$\begin{aligned}\mathcal{N} \quad = \quad & \int \left(\prod_{l\neq k} \mathrm{d}x_l\right) \exp\left\{-\frac{n}{2}\sum_{l\neq k}\frac{x_l^2}{g_{\mu l}^t}\right\} \int \mathrm{d}\omega \exp\left\{-\frac{n}{2}\frac{X_{\mu k}-Q_{\mu k}^t}{R_{\mu k}^t}\left(X_{\mu k}-Q_{\mu k}^t+R_{\mu k}^t\right)\omega^2\right\} \\ & \int \left(\prod_{\mathrm{a}=1}^{n} \mathrm{d}\Delta_{\mu k}^{\mathrm{a}}\ \mathrm{d}z^{\mathrm{a}}\right) \exp\left\{-\frac{\sum_{\mathrm{a}=1}^{n}\left(\Delta_{\mu k}^{\mathrm{a}}\right)^2}{2\left(X_{\mu k}-Q_{\mu k}^t\right)} + \sum_{\mathrm{a}=1}^{n}\left(\omega + \frac{u_{\mu k}^t}{X_{\mu k}-Q_{\mu k}^t+R_{\mu k}^t} - iz^{\mathrm{a}}\right)\Delta_{\mu k}^{\mathrm{a}}\right\} \\ & \underset{\{\mathbf{b}_{l\neq k}\}}{\mathsf{Tr}} \exp\left\{\sum_{l\neq k}\sum_{\mathrm{a}=1}^{n}\left(i\varepsilon_{\mu l}z^{\mathrm{a}}+x_l\right)b_l^{\mathrm{a}}\right\} \underset{\{\mathbf{b}_k\}}{\mathsf{Tr}}\, b_k^{\tilde{\mathrm{a}}} \left[P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right) + \varepsilon_{\mu k}\nabla_{\mathbf{\Delta}_{\mu k}}P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right)\cdot\mathbf{b}_k\right] \ .\end{aligned}$$

Using the small $\varepsilon_{\mu l}$ approximation as in Eq.(7) we can write

$$\begin{aligned}\underset{\{\mathbf{b}_{l\neq k}\}}{\mathsf{Tr}} \exp\left\{\sum_{l\neq k}\sum_{\mathrm{a}=1}^{n}\left(i\varepsilon_{\mu l}z^{\mathrm{a}}+x_l\right)b_l^{\mathrm{a}}\right\} \quad &= \quad 2^{n(K-1)}\prod_{\mathrm{a}=1}^{n}\prod_{l\neq k}\cosh\left(x_l\right)\cos\left(\varepsilon_{\mu l}z^{\mathrm{a}}\right)\left[1+i\tanh\left(x_l\right)\tan\left(\varepsilon_{\mu l}z^{\mathrm{a}}\right)\right] \\ &\simeq \quad 2^{n(K-1)}\prod_{\mathrm{a}=1}^{n}\prod_{l\neq k}\cosh\left(x_l\right)\left[1+i\varepsilon_{\mu l}z^{\mathrm{a}}\tanh\left(x_l\right)\right] \\ &\simeq \quad 2^{n(K-1)}\prod_{l\neq k}\cosh^n\left(x_l\right)\prod_{\mathrm{a}=1}^{n}\exp\left\{iz^{\mathrm{a}}\sum_{l\neq k}\varepsilon_{\mu l}\tanh\left(x_l\right)\right\} \ . \qquad (58)\end{aligned}$$

Using $P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right)=\prod_{\mathrm{a}=1}^{n}P\left(y_\mu|\Delta_{\mu k}^{\mathrm{a}};\boldsymbol{\gamma}\right)$, the traces on $\mathbf{b}_k$ can be written as

$$\underset{\{\mathbf{b}_k\}}{\mathsf{Tr}} \left[P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right) + \varepsilon_{\mu k}\nabla_{\mathbf{\Delta}_{\mu k}}P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right)\cdot\mathbf{b}_k\right] \quad = \quad 2^n\prod_{\mathrm{a}=1}^{n}P\left(y_\mu|\Delta_{\mu k}^{\mathrm{a}};\boldsymbol{\gamma}\right)$$

$$\begin{aligned}&\underset{\{\mathbf{b}_k\}}{\mathsf{Tr}}\, b_k^{\tilde{\mathrm{a}}} \left[P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right) + \varepsilon_{\mu k}\nabla_{\mathbf{\Delta}_{\mu k}}P\left(y_\mu|\mathbf{\Delta}_{\mu k};\boldsymbol{\gamma}\right)\cdot\mathbf{b}_k\right] \\ &\qquad = 2^n\varepsilon_{\mu k}\left(\prod_{\mathrm{a}=1}^{n}P\left(y_\mu|\Delta_{\mu k}^{\mathrm{a}};\boldsymbol{\gamma}\right)\right)\qquad \frac{\partial}{\partial\Delta_{\mu k}^{\tilde{\mathrm{a}}}}\ln P\left(y_\mu|\Delta_{\mu k}^{\tilde{\mathrm{a}}};\boldsymbol{\gamma}\right) \ .\end{aligned}$$

Putting all together we find that the integrals over z^{a} generate the delta function

$$\prod_{\mathrm{a}=1}^{n}\delta\left(\Delta_{\mu k}^{\mathrm{a}}-\sum_{l\neq k}\varepsilon_{\mu l}\tanh\left(x_l\right)\right) \ ,$$

the integrals over $\Delta^{\mathrm{a}}_{\mu k}$ just replace all the $\Delta^{\mathrm{a}}_{\mu k}$ by $v_{\mu k}\left(\boldsymbol{x}_{l\neq k}\right) \equiv \sum_{l\neq k} \varepsilon_{\mu l} \tanh\left(x_l\right)$, and the integral over ω gives an expression that is proportional to $\exp\left\{-\frac{n}{2}\frac{\left(u^t_{\mu k}-v_{\mu k}\right)^2}{X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}}\right\}$. Thus the final expression for the message is

$$\widehat{m}^{t+1}_{\mu k} = \varepsilon_{\mu k} \frac{\int \left(\prod_{l\neq k} \mathrm{d}x_l\right) \exp\left\{-n\mathcal{H}^t_{\mu k}\left(\boldsymbol{x}_{l\neq k}\right)\right\} \frac{\partial}{\partial z} \ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=v_{\mu k}}}{\int \left(\prod_{l\neq k} \mathrm{d}x_l\right) \exp\left\{-n\mathcal{H}^t_{\mu k}\left(\boldsymbol{x}_{l\neq k}\right)\right\}}$$

where

$$\mathcal{H}^t_{\mu k}\left(\boldsymbol{x}_{l\neq k}\right) = \sum_{l\neq k}\left[\frac{x_l^2}{2g^t_{\mu l}} - \ln\cosh\left(x_l\right)\right] + \frac{\left(u^t_{\mu k}-v_{\mu k}\right)^2}{2\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)} - \ln P\left(y_\mu|v_{\mu k};\boldsymbol{\gamma}\right) \ ,$$

as in Eq.(29).

C Deriving expressions for the messages - $m_{\mu k}$

The expression for the messages from **b**-nodes to **y**-nodes is derived in a similar manner

$$\begin{aligned}
m^t_{\mu k} &= \mathop{\mathsf{Tr}}_{\{\mathbf{b}_k\}} b^{\tilde{\mathrm{a}}}_k P^t\left(\mathbf{b}_k|\left\{y_{\nu\neq\mu}\right\}\right) \\
&= \frac{\mathop{\mathsf{Tr}}_{\{\mathbf{b}_k\}} b^{\tilde{\mathrm{a}}}_k \prod_{\nu\neq\mu} \mathop{\mathsf{Tr}}_{\{\mathbf{b}_{l\neq k}\}} P\left(y_\nu|\mathbf{B}\right)\mathbf{b}_l|\left\{y_{\sigma\neq\nu}\right\} \prod_{l\neq k} P^{t-1}\left(\mathbf{b}_l|\left\{y_{\sigma\neq\nu}\right\}\right)}{\mathop{\mathsf{Tr}}_{\{\mathbf{b}_k\}} \prod_{\nu\neq\mu} \mathop{\mathsf{Tr}}_{\{\mathbf{b}_{l\neq k}\}} P\left(y_\nu|\mathbf{B}\right) \prod_{l\neq k} P^{t-1}\left(\mathbf{b}_l|\left\{y_{\sigma\neq\nu}\right\}\right)} &(59)\\
&\simeq \frac{\mathop{\mathsf{Tr}}_{\{\mathbf{b}_k\}} b^{\tilde{\mathrm{a}}}_k \prod_{\nu\neq\mu} \int \left(\prod_{l\neq k} \mathrm{d}x_{\nu l}\right) \exp\left\{-n\mathcal{H}^{t-1}_{\nu k}\left(\boldsymbol{x}_{\nu,l\neq k}\right)\right\} \left[1+\varepsilon_{\nu k}\frac{\partial}{\partial z}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=v_{\nu k}} \sum_{\mathrm{a}=1}^{n} b^{\mathrm{a}}_k\right]}{\mathop{\mathsf{Tr}}_{\{\mathbf{b}_k\}} \prod_{\nu\neq\mu} \int \left(\prod_{l\neq k} \mathrm{d}x_{\nu l}\right) \exp\left\{-n\mathcal{H}^{t-1}_{\nu k}\left(\boldsymbol{x}_{\nu,l\neq k}\right)\right\} \left[1+\varepsilon_{\nu k}\frac{\partial}{\partial z}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=v_{\nu k}} \sum_{\mathrm{a}=1}^{n} b^{\mathrm{a}}_k\right]} &(60)\\
&= \frac{\int \left(\prod_{\substack{\nu\neq\mu\\ l\neq k}} \mathrm{d}x_{\nu l}\right) \exp\left\{-n\sum_{\nu\neq\mu}\mathcal{H}^{t-1}_{\nu k}\left(\boldsymbol{x}_{\nu,l\neq k}\right)\right\} \sum_{b^{\tilde{\mathrm{a}}}_k=\pm1} \left[b^{\tilde{\mathrm{a}}}_k \prod_{\nu\neq\mu}\left(1+\varepsilon_{\nu k}\frac{\partial}{\partial z}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right) b^{\tilde{\mathrm{a}}}_k\right)\right]}{\int \left(\prod_{\substack{\nu\neq\mu\\ l\neq k}} \mathrm{d}x_{\nu l}\right) \exp\left\{-n\sum_{\nu\neq\mu}\mathcal{H}^{t-1}_{\nu k}\left(\boldsymbol{x}_{\nu,l\neq k}\right)\right\} \sum_{b^{\tilde{\mathrm{a}}}_k=\pm1} \left[\prod_{\nu\neq\mu}\left(1+\varepsilon_{\nu k}\frac{\partial}{\partial z}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right) b^{\tilde{\mathrm{a}}}_k\right)\right]} \\
&= \frac{\prod_{\nu\neq\mu}\left(1+\widehat{m}^t_{\nu k}\right) - \prod_{\nu\neq\mu}\left(1-\widehat{m}^t_{\nu k}\right)}{\prod_{\nu\neq\mu}\left(1+\widehat{m}^t_{\nu k}\right) + \prod_{\nu\neq\mu}\left(1-\widehat{m}^t_{\nu k}\right)} \\
&= \tanh\left(\sum_{\nu\neq\mu} \operatorname{arctanh}\left(\widehat{m}^t_{\nu k}\right)\right) \ ,
\end{aligned}$$

to obtain Eq.(31), where we have used the approximation Eq.(58) to go from Eq.(59) to Eq.(60).

D Obtaining the roots of $\tilde{v}_{\mu k}$

Having defined $\overline{v}_{\mu k}$,as in Eq.(33), one can then rewrite Eq.(32) as

$$
\begin{aligned}
0 &= \frac{\tilde{x}_l}{g^t_{\mu l}} - \tanh\left(\tilde{x}_l\right) - \varepsilon_{\mu l}\left[1-\tanh^2\left(\tilde{x}_l\right)\right]\frac{\left(\overline{v}_{\mu k}-\tilde{v}_{\mu k}\right)}{\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)}\times \\
&\times\left[1-\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)\frac{\frac{\partial}{\partial z}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=\overline{v}_{\mu k}}-\frac{\partial}{\partial z}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=\tilde{v}_{\mu k}}}{\overline{v}_{\mu k}-\tilde{v}_{\mu k}}\right], \qquad (61)
\end{aligned}
$$

where, by continuity of the derivative of P we can write

$$
\begin{aligned}
0 &= \frac{\tilde{x}_l}{g^t_{\mu l}} - \tanh\left(\tilde{x}_l\right) - \varepsilon_{\mu l}\left[1-\tanh^2\left(\tilde{x}_l\right)\right]\frac{\left(\overline{v}_{\mu k}-\tilde{v}_{\mu k}\right)}{\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)}\times \\
&\times\left[1-\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)\frac{\partial^2}{\partial z^2}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=\lambda\overline{v}_{\mu k}+(1-\lambda)\tilde{v}_{\mu k}}\right],
\end{aligned}
$$

for some $\lambda \in (0,1)$. The last term can be identified with a small field

$$
\begin{aligned}
\tilde{h}^t_{\mu l} &\equiv \varepsilon_{\mu l} g^t_{\mu l}\left[1-\tanh^2\left(\tilde{x}_l\right)\right]\frac{\left(\overline{v}_{\mu k}-\tilde{v}_{\mu k}\right)}{\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)} \qquad (62)\\
&\times\left[1-\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)\frac{\partial^2}{\partial z^2}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=\lambda\overline{v}_{\mu k}+(1-\lambda)\tilde{v}_{\mu k}}\right].
\end{aligned}
$$

The field is small $\mathcal{O}\left(\varepsilon_{\mu l}\right)$, so the lower minimum should be located near by $m^t_{\mu l}$. According to Eq.(18) the minimum is located at

$$
\tilde{x}_l \simeq m^t_{\mu l} + \frac{1}{1-g^t_{\mu l}\left[1-\left(m^t_{\mu l}\right)^2\right]}\tilde{h}^t_{\mu l}.
$$

Thus the position of the minimum is shifted by the action of the field. Equation (61) can be expressed as

$$
\tanh\left(\tilde{x}_l\right) \simeq m^t_{\mu l} - \varepsilon_{\mu l}\mathcal{A}_{\mu l,k}\left[1-\tanh^2\left(\tilde{x}_l\right)\right], \qquad (63)
$$

where

$$
\begin{aligned}
\mathcal{A}_{\mu l,k} &\equiv \frac{g^t_{\mu l}\left[1-\left(m^t_{\mu l}\right)^2\right]}{1-g^t_{\mu l}\left[1-\left(m^t_{\mu l}\right)^2\right]}\frac{\left(\tilde{v}_{\mu k}-\overline{v}_{\mu k}\right)}{\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)} \qquad (64)\\
&\times\left[1-\left(X_{\mu k}-Q^t_{\mu k}+R^t_{\mu k}\right)\frac{\partial^2}{\partial z^2}\ln P\left(y_\mu|z;\boldsymbol{\gamma}\right)|_{z=\lambda\overline{v}_{\mu k}+(1-\lambda)\tilde{v}_{\mu k}}\right].
\end{aligned}
$$

The solution of Eq.(63) is:

$$
\tanh\left(\tilde{x}_l\right) \simeq m^t_{\mu l} - \varepsilon_{\mu l}\mathcal{A}_{\mu l,k}\left[1-\left(m^t_{\mu l}\right)^2\right] + \mathcal{O}\left(\varepsilon^2_{\mu l}\right),
$$

so, multiplying both members by $\varepsilon_{\mu l}$ and adding over all $l \neq k$ one obtains Eq.(34).