Inference by replication in densely connected systems

Juan P. Neirotti and David Saad

The Neural Computing Research Group, Aston University, Birmingham B4 7ET, United Kingdom
(Received 4 July 2007; published 30 October 2007)

An efficient Bayesian inference method for problems that can be mapped onto dense graphs is presented. The approach is based on message passing where messages are averaged over a large number of replicated variable systems exposed to the same evidential nodes. An assumption about the symmetry of the solutions is required for carrying out the averages; here we extend the previous derivation based on a replica-symmetric-(RS)-like structure to include a more complex one-step replica-symmetry-breaking-like (1RSB-like) ansatz. To demonstrate the potential of the approach it is employed for studying critical properties of the Ising linear perceptron and for multiuser detection in code division multiple access (CDMA) under different noise models. Results obtained under the RS assumption in the noncritical 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 observed. While the 1RSB ansatz is not required for the original problems, it was applied to the CDMA signal detection problem with a more complex noise model that exhibits RSB behavior, resulting in an improvement in performance.

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 [1]. A recent review [2] highlights the links between the various approximation methods and their applications.

Approximative Bayesian inference techniques arguably offer the most principled approach to information extraction, by combining a rigorous statistical approach with a feasible but systematic approximation. Although message passing techniques have existed for some time in the computer science community [3,4] they have enjoyed growing popularity in recent years [5], mainly within the context of Bayesian networks and the use of belief propagation (BP) for a range of inference applications, from signal extraction in telecommunication to machine learning.

The main advantage of these techniques is their moderate growth in computational cost, with respect to the systems size, due to the local nature of the calculation when applied to sparse graphs. Until recently, message passing techniques were deemed unsuitable for inference in densely connected systems due to the inherently high number of short loops in the corresponding graphical representation, and the large number of connections per node, which results in a high computational cost. Both properties are considered prohibitive to the use of conventional message passing techniques in such problems.

A recently suggested method for message passing in densely connected systems [6] relies on replacing individual messages by averages sampled from a Gaussian distribution of some mean and variance that are modified iteratively. The method has been applied for the code division multiple access (CDMA) signal detection inference problem; it successfully finds optimal solutions where the space of solutions is

contiguous but breaks down when the solution space becomes fragmented, for instance, when there is a mismatch between the true and assumed noise levels in the CDMA detection problem. The emergence of competing solutions gives rise to conflicting messages that result in bungled average messages and suboptimal performance. In statistical physics terms, it corresponds to the replica symmetric solution in dense systems [7] and gives poor estimates when more complex solution structures are required.

In the current paper, we methodologically extend the approach of Kabashima [6] for inference in dense graphs by considering a large (infinite) number of replicated variable systems, exposed to the same evidential data (received signals). Each one of the systems represents a pure state and a possible solution. The pseudoposteriors, that form the basis for our estimates, are based on averages over the replicated systems. The method has been employed previously only in the noncritical regime [8], using the most basic replica-symmetric-like (RS-like) ansatz for the solution structure. In this paper we study both critical and noncritical regimes and extend the solution structure considered to include one-step-replica-symmetry-breaking-like (1RSB-like) structures [9]. To demonstrate the potential of this approach and the performance obtained using the resulting algorithm we apply the method to two different but related problems: signal detection in code division multiple access and learning in the Ising linear perceptron (ILP).

CDMA is a signal modulation (or spreading) method used when multiple transmitters communicate with a single receiver simultaneously [10]. It is mainly used in wireless communication for efficiently separating the multiple messages (e.g., from mobile phones) at the receiver end (base station) given the aggregated and corrupted received signal. Messages are modulated by the various users prior to transmission (uplink) and by the single transmitter on the way back (downlink). The shared knowledge of the modulation sequences by the receiver facilitates the signal detection; the aim is to optimally retrieve the multiple messages with mini-

046121-1

mal corruption. The method has been thoroughly studied in the information theory literature.

Perceptrons [11] serve as rudimentary models of interacting biological neurons; they have been studied using a range of analytical methods to gain insight into the behavior, limitations, and capabilities of groups of interacting nonlinear elements (neurons). The basic perceptron consists of a product between a binary input vector and a weight vector of equal dimensionality; a nonlinear function of that product constitutes the perceptron's output. Learning in such systems corresponds to an estimation of the interaction coefficients (weights) between elements, such that a certain input-output relation is realized. The latter is instantiated by a sample of input vectors and the corresponding output values and the goal is to optimally infer the values of the weights. In addition to their role as simple biological models, perceptrons serve as basic building blocks in many practical pattern analysis techniques and algorithms [12]. The variant studied here, ILP, is a special case where the weight vector comprises binary values and the output function is linear.

We investigate both RS and 1RSB-like structures. The former is applied to both CDMA and ILP problems and seems to be sufficient for obtaining optimal performances; the latter is applied to a variant of the CDMA signal detection problem with a more complex noise model that exhibits RSB-like behavior, to demonstrate its efficacy for particularly difficult inference tasks.

In Sec. II we will introduce the general models studied, followed by a brief review of message passing techniques for dense systems in Sec. III. The general derivation of our approach, for both RS and RSB-like solution structures, will be presented in Sec. IV; numerical studies of both CDMA signal detection and ILP learning will be reported in Sec. V. To demonstrate the method based on the more complex 1RSB solution structure, and to examine its efficacy to problems that require such structures, we will introduce a variant of the CDMA signal detection problem and study it numerically in Sec. VI. We will conclude the presentation with a summary and point to future research directions. Details of the derivation will be provided in Appendixes A–E.

II. MODELS STUDIED

Before describing the inference method, the approach taken and the algorithms derived from it, it would be helpful to briefly describe the exemplar inference problems tackled in this paper.

We apply the method to two different but related inference problems: signal detection in CDMA and learning in the Ising linear perceptron. Both correspond to inference problems where data points are noisy representations of sums of binary variables modulated by random binary values.

Multiple access communication refers to the transmission of multiple messages to a single receiver. The scenario we study here, described schematically in Fig. 1(a), 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
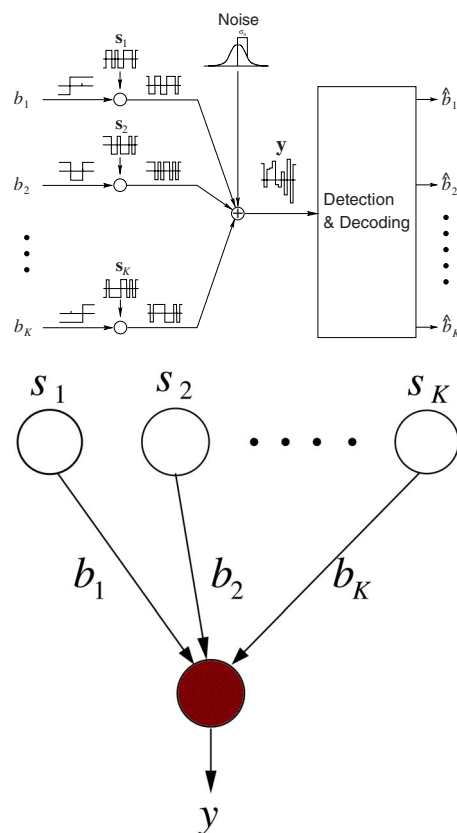

FIG. 1. (Color online) Schematic representation of (a) the CDMA system, (b) the ILP.

messages, in particular time, frequency, and code division multiple access [10]. 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 $O(1)$. 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

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

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 is the received message. The task is to infer the original transmission from the set of received messages. This process is reminiscent of the learning task performed by a perceptron with binary weights and linear output, which is the next example considered in this paper.

Learning in neural networks has attracted considerable theoretical interest. In particular we focus on supervised learning from examples, which relies on a training set consisting of examples of the target task [13]. We consider a perceptron, described schematically in Fig. 1(b), which is a network that sums a single layer of inputs $s_{\mu k}$ with synaptic

weights b_k and passes the result through a transfer function y_μ,

$$y_\mu = g\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} s_{\mu k}b_k\right),\qquad(2)$$

where g is typically a nonlinear sigmoidal function. If $g(x)=x$ the network is termed *linear output perceptron*. If the weights $b_k \in \{\pm1\}$ the network is called *Ising perceptron*. Learning is a search through the weight space for the perceptron that best approximates a target rule.

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 noncritical low load values, ILP studies focused on the critical regime. We consider both regimes in this paper.

III. MESSAGE PASSING FOR INFERENCE IN DENSELY CONNECTED SYSTEMS

Graphical models (Bayes belief networks) provide a powerful framework for modelling statistical dependencies between variables [3–5]. They play an essential role in devising a principled probabilistic framework for inference in a broad range of applications.

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 treelike; there are no such guarantees for systems with loops even in the case of large loops and a local treelike structure (although message passing techniques have been used successfully in loopy systems, supported by some limited theory [14]). A clear link has been established between certain message passing algorithms and well-known methods of statistical mechanics [2] such as the Bethe approximation [15,16].

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, and an exponential number of loops. However, an exciting new approach has been recently suggested [6] for extending BP techniques [3–5] 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. The technique has been successfully applied to CDMA signal detection problems and the results reported are competitive with those of other state-of-the-art techniques. However, the current approach has some inherent limitations [6], presumably due to its similarity to the replica symmetric solution in the equivalent Ising spin models [1,7].

In a separate recent development [17], the replica-symmetric-equivalent BP has been extended to survey propa-
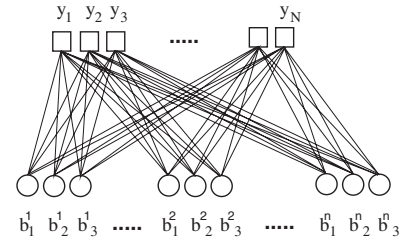

FIG. 2. Replicated solutions $\mathbf{B}=(\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_K)$ given data.

gation (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 problems [17], far beyond other existing approaches. In addition, the algorithm facilitated theoretical studies of the corresponding physical system and contributed to our understanding of it [18]. The SP algorithm has recently been modified to handle Ising and multilayer perceptrons [19].

IV. GENERAL FORMALISM

We recently presented an approach [8] for inference in densely connected systems, which was inspired by both the extension of BP to densely connected graphs and the introduction of SP. The systems we consider here are characterized 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 graph, where the observables y_μ are linked to variables in all replicated systems, namely $\mathbf{B}=(\mathbf{b}^1,\mathbf{b}^2,\ldots,\mathbf{b}^n)$; where $\mathbf{b}^a=(b_1^a,b_2^a,\ldots,b_K^a)^\mathsf{T}$, as shown in Fig. 2. To estimate the variables $\mathbf{B}$ given the data $\mathbf{y}^\mathsf{T}=(y_1,y_2,\ldots,y_N)$, in a Bayesian framework, we must maximize the posterior $P(\mathbf{B}|\mathbf{y}) \propto \Pi_{\mu=1}^N P(y_\mu|\mathbf{B})P(\mathbf{B})$, where we have considered independent data, and thus $P(\mathbf{y}|\mathbf{B})=\Pi_{\mu=1}^N P(y_\mu|\mathbf{B})$.

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 $\mathbf{b} \in \{\pm1\}^K$ is an unbiased vector and $P(\mathbf{B})=2^{-Kn}$. The estimate we would like to obtain is the maximizer of the posterior marginal (MPM) $\hat{\mathbf{b}}_k =\text{argmax}_{\mathbf{b}_k \in \{\pm1\}^n}\Sigma_{\{\mathbf{b}_{l\neq k}\}}P(\mathbf{B}|\mathbf{y})$, which is expected to be a vector with equal entries for all replica $\hat{b}_k^1=\hat{b}_k^2=\cdots=\hat{b}_k^n$. The number of operations required to obtain the full MPM estimator is of $O(2^K)$ which is infeasible for large K values.

To obtain an approximate MPM estimate we apply BP message passing technique [3–5]. In particular we are interested here in the application of BP to densely connected graphs, similar to the one presented in [6]. The latter is based on estimating a single solution and therefore does not converge, as has been observed, when the solution space becomes fragmented and multiple solutions emerge. This arguably corresponds to the replica symmetry breaking phenomena and occurs, for instance, when the noise level is unknown in the CDMA signal detection case.

A potential algorithmic improvement is achieved by the introduction of an SP-like approach, based on replicated variable systems, similar to the approach taken in problems that can be mapped onto sparsely connected graphs.

Using Bayes rule one straightforwardly obtains the BP equations

$$P^{t+1}(y_\mu|\mathbf{b}_k,\{y_{\nu\neq\mu}\}) = \sum_{\{\mathbf{b}_{l\neq k}\}} P(y_\mu|\mathbf{B}) \prod_{l\neq k} P^t(\mathbf{b}_l|\{y_{\nu\neq\mu}\}), \quad (3)$$

$$P^t(\mathbf{b}_l|\{y_{\nu\neq\mu}\}) \propto \prod_{\nu\neq\mu} P^t(y_\nu|\mathbf{b}_l,\{y_{\sigma\neq\nu}\}). \quad (4)$$

For calculating the posterior $P(\mathbf{y}|\mathbf{B})$, we assume a dependency of the data on the parameters of the form $y_\mu = \mathcal{F}(\Sigma_{l=1}^K \varepsilon_{\mu l}\mathbf{b}_l; \gamma)$, where $\mathcal{F}$ is some general smooth function, γ are model parameters and $\varepsilon_{\mu l}$ are small enough to ensure that $\Sigma_{l=1}^K \varepsilon_{\mu l} b_l^a \sim O(1)$. We define the vector $\mathbf{\Delta}_\mu \equiv \Sigma_{l=1}^K \varepsilon_{\mu l}\mathbf{b}_l = \Sigma_{l\neq k}\varepsilon_{\mu l}\mathbf{b}_l + \varepsilon_{\mu k}\mathbf{b}_k = \mathbf{\Delta}_{\mu k} + \varepsilon_{\mu k}\mathbf{b}_k$, where the index k indicates that the contribution of the $\mathbf{b}_k$ component has been taken out. Thus, using $y_\mu = \mathcal{F}(\mathbf{\Delta}_{\mu k} + \varepsilon_{\mu k}\mathbf{b}_k; \gamma)$ and separating the contribution of $\mathbf{b}_k$, we can express the likelihood with the help the k-dependent integration variable $\mathbf{\Delta}_{\mu k}$ (for any k) such that

$$P(y_\mu|\mathbf{B}) = \int d\mathbf{\Delta}_{\mu k} P(y_\mu,\mathbf{\Delta}_{\mu k}|\mathbf{B};\gamma)$$

$$= \int d\mathbf{\Delta}_{\mu k} P(y_\mu|\mathbf{\Delta}_{\mu k},\mathbf{B};\gamma) P(\mathbf{\Delta}_{\mu k}|\mathbf{B})$$

$$= \int d\mathbf{\Delta}_{\mu k} P(y_\mu|\mathbf{\Delta}_{\mu k} + \varepsilon_{\mu k}\mathbf{b}_k;\gamma) P(\mathbf{\Delta}_{\mu k}|\mathbf{B})$$

$$\simeq \int d\mathbf{\Delta}_{\mu k}[1 + \varepsilon_{\mu k}\mathbf{b}_k^\mathsf{T}\nabla_{\mathbf{\Delta}_{\mu k}} \ln P(y_\mu|\mathbf{\Delta}_{\mu k};\gamma)]$$

$$\times P(y_\mu|\mathbf{\Delta}_{\mu k};\gamma) P(\mathbf{\Delta}_{\mu k}|\mathbf{B}), \quad (5)$$

where we have assumed that $P(y_\mu|\mathbf{\Delta}_{\mu k},\mathbf{B};\gamma) \approx P(y_\mu|\mathbf{\Delta}_{\mu k} + \varepsilon_{\mu k}\mathbf{b}_k;\gamma)$, due to the assumed dependence of the observed values y_μ on $\mathbf{\Delta}_{\mu k}$ and $\mathbf{b}_k$.

A. Interreplica correlations

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

$$P^t(\mathbf{b}_k|\{y_{\nu\neq\mu}\}) \propto \exp\{\mathbf{h}_{\mu k}^{t\mathsf{T}}\mathbf{b}_k + \tfrac{1}{2}\mathbf{b}_k^\mathsf{T}\mathbf{Q}_{\mu k}^t\mathbf{b}_k\}, \quad (6)$$

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. The form of $\mathbf{Q}_{\mu k}^t$ depends upon the particular case considered. We assume one of the following symmetry relation between the replicated solutions:

$$(\mathbf{h}_{\mu k}^t)^{\ell a} = h_{\mu k}^t,$$

$$(^{(\mathrm{RS})}\mathbf{Q}_{\mu k}^t)^{aa'} = \delta^{aa'} q_{0\mu k}^t + (1 - \delta^{aa'}) q_{1\mu k}^t,$$

$$(^{(\mathrm{1RSB})}\mathbf{Q}_{\mu k}^t)^{\ell a \ell' a'} = \delta^{\ell\ell'}(^{(\mathrm{RS})}\mathbf{Q}_{\mu k}^t)^{aa'} + (1 - \delta^{\ell\ell'}) q_{2\mu k}^t,$$

where ℓ is a block index that runs from 1 to L and a is a intrablock replica index that runs form 1 to n where n is the number of variables per block. We also make the following reasonable assumption $q_{0\mu k}^t > q_{1\mu k}^t > q_{2\mu k}^t > 0$, as one expects correlations to gradually decrease between variables with nonidentical replica and block indices, respectively.

For both types of symmetries considered, the correlation matrix defined as

$$(\mathbf{Y}_{\mu k}^t)^{\mathbf{II'}} \equiv \langle\mathbf{\Delta}_{\mu k}^\mathbf{I}\mathbf{\Delta}_{\mu k}^\mathbf{I'}\rangle - \langle\mathbf{\Delta}_{\mu k}^\mathbf{I}\rangle\langle\mathbf{\Delta}_{\mu k}^\mathbf{I'}\rangle,$$

where $\mathbf{I}$ is an index or a pair of indices for RS and 1RSB, respectively. The correlation matrix is assumed to be self-averaging, i.e., $\mathbf{Y}_{\mu k}^t \simeq \mathbf{Y}^t$ and preserves the symmetry of the matrix $\mathbf{Q}_{\mu k}^t$. An explicit derivation of the entries of $\mathbf{Y}^t$ is presented in Appendixes A and B, for the RS and RSB-like correlation structures, respectively; the matrices take the following general form:

$$(^{(\mathrm{RS})}\mathbf{Y}^t)^{aa'} = \delta^{aa'} X^t + (1 - \delta^{aa'})\frac{1}{n}R^t,$$

$$(^{(\mathrm{1RSB})}\mathbf{Y}^t)^{a\ell'a'\ell'} = \delta^{\ell\ell'}\left(\delta^{aa'} X^t + (1 - \delta^{aa'})\frac{1}{n}V^t\right)$$

$$+ (1 - \delta^{\ell\ell'})\frac{1}{nL}(V^t - R^t).$$

Thus, for the appropriate center of the distribution $\mathbf{u}_{\mu k}^t$ [see Eqs. (A8) and (B13)], the probability of $\mathbf{\Delta}_{\mu k}$ can be expressed as

$$P(\mathbf{\Delta}_{\mu k}|\mathbf{B}) = \sqrt{\frac{1}{(2\pi)^n \det(\mathbf{Y}^t)}} \exp\left[-\frac{1}{2}(\mathbf{\Delta}_{\mu k} - \mathbf{u}_{\mu k}^t)^\mathsf{T}(\mathbf{Y}^t)^{-1}(\mathbf{\Delta}_{\mu k} - \mathbf{u}_{\mu k}^t)\right]$$

$$\propto \begin{cases} \displaystyle\int d\vartheta \exp\left(-n\frac{(\vartheta - u_{\mu k}^t)^2}{2R^t}\right) \prod_{a=1}^n \exp\left(-\frac{(\Delta_{\mu k}^a - \vartheta)^2}{2X^t}\right) & (\mathrm{RS}), \\[4mm] \displaystyle\int d\Theta \prod_{\ell=1}^L \exp\left[-\frac{n}{2}\left(\frac{(\vartheta^0)^2}{V^t - R^t} + \frac{(\vartheta^\ell)^2}{V^t - L^{-1}(V^t - R^t)}\right)\right] \prod_{a=1}^n \exp\left(-\frac{(\Delta_{\mu k}^{\ell a} - \vartheta_{\mu k}^{0\ell t})^2}{2(X^t - n^{-1}V^t)}\right) & (\mathrm{RSB}), \end{cases} \quad (7)$$

for the RS and RSB-like correlation matrices, respectively, where $\vartheta_{\mu k}^{0 \ell t} \equiv \vartheta^0 + \vartheta^\ell + u_{\mu k}^t$ and $\boldsymbol{\Theta}^{\mathsf{T}} = (\vartheta^0, \vartheta^1, \ldots, \vartheta^L)$.

B. Messages

Having obtained the conditional probability distribution $P(\boldsymbol{\Delta}_{\mu k} | \mathbf{B})$ one can then derive explicit expressions for the messages $m_{\mu k}$ (magnetization) and $\hat{m}_{\mu k}$ that can be viewed as parameters in the corresponding marginalized binary distributions $P^t(y_\mu | b_k, \{y_{\nu \neq \mu}\}) \propto (1 + \hat{m}_{\mu k}^t b_k)/2$ and $P^t(b_k | \{y_{\nu \neq \mu}\}) = (1 + m_{\mu k}^t b_k)/2$.

The messages from nodes y_μ to nodes $\mathbf{b}_k$, as derived in Appendix C, Eqs. (C1)–(C8),

$$
\hat{m}_{\mu k}^{t+1} = \begin{cases} \varepsilon_{\mu k} \dfrac{\widetilde{\vartheta}_{\mu k}^t - u_{\mu k}^t}{R^t} & \text{(RS)}, \\[3mm] \varepsilon_{\mu k} \dfrac{\widetilde{\vartheta}_{\mu k}^t - u_{\mu k}^t}{2V^t - R^t} + \dfrac{\varepsilon_{\mu k}}{2n} \dfrac{\mathcal{P}_2 V^t}{1 - \mathcal{P}_1 V^t} & \text{(RSB)}, \end{cases} \tag{8}
$$

where $\mathcal{P}_j = \frac{\partial^j \mathcal{P}}{\partial \vartheta^j} \big|_{\vartheta = \widetilde{\vartheta}_{\mu k}^t}$, $\mathcal{P}$ is defined in Eq. (C3) and $\widetilde{\vartheta}_{\mu k}^t$ is obtained from the saddle point equations given by Eq. (D1) in the RS case and by Eq. (D2) in the 1RSB case. The messages from nodes $\mathbf{b}_k$ to y_μ are given in both cases by the expression $m_{\mu k}^t \simeq \tanh(\Sigma_{\nu \neq \mu} \hat{m}_{\nu k}^t)$.

For the gauged field $b_k h_{\mu k}^t$ where $h_{\mu k}^t \equiv \operatorname{arctanh}(m_{\mu k}^t) = \Sigma_{\nu \neq \mu} \operatorname{arctanh}(\hat{m}_{\nu k}^t) \simeq \Sigma_{\nu \neq \mu} \hat{m}_{\nu k}^t$. The distribution of this field is well approximated by a Gaussian as a result of the central limit theorem. The mean and variance of the Gaussian are E^t and F^t, respectively,

$$
E^t = \frac{1}{K} \sum_{k=1}^{K} \sum_{\mu=1}^{N} b_k \hat{m}_{\mu k}^t,
$$

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

$$
\simeq \frac{1}{K} \sum_{k=1}^{K} \sum_{\mu=1}^{N} (\hat{m}_{\mu k}^t)^2. \tag{9}
$$

Both E^t and F^t are assumed to be independent of the index μ by virtue of the self-averaging property. For the same reason we expect the macroscopic variables defined as $M_\mu^t \equiv \Sigma_{k=1}^{K} b_k m_{\mu k}^t / K \simeq \Sigma_{k=1}^{K} b_k m_k^t / K = M^t$ and $N_\mu^t \equiv \Sigma_{k=1}^{K} (m_{\mu k}^t)^2 / K \simeq \Sigma_{k=1}^{K} (m_k^t)^2 / K = N^t$, where $m_k^t \simeq \tanh(\Sigma_{\nu=1}^{N} \hat{m}_{\nu k}^t)$, to be independent of the index μ. Thus, these macroscopic variables can be evaluated by the following integrals:

$$
M^t = \int \mathcal{D}u \, \tanh(\sqrt{F^t} u + E^t), \quad N^t = \int \mathcal{D}u \, \tanh^2(\sqrt{F^t} u + E^t),
$$

where $\mathcal{D}u = du \exp(-u^2/2)/\sqrt{2\pi}$ represents the product of a differential (du) times a probability distribution (Gaussian).

C. Optimization

The structure of the correlation matrix used introduces free variables in the form of the correlation terms between replicated solutions. These are used for optimizing the estimation provided with respect to a given performance measure.

Since the MPM estimator is given by $\hat{b}_k^t = \operatorname{sgn}(m_k^t) \simeq \operatorname{sgn}(m_{\mu k}^t) = \operatorname{sgn}(h_{\mu k}^t)$, the expression for the error per bit rate takes the form

$$
P_b^t = \frac{1}{2K} \sum_{k=1}^{K} [1 - \operatorname{sgn}(b_k m_k^t)], \tag{10}
$$

which is minimized when the true message vector $\mathbf{b}$ and the vector of messages $\mathbf{m}^t$ are parallel. Therefore, the error rate per bit decreases as the ratio $M^t/\sqrt{N^t} = \cos(\widehat{\mathbf{bm}^t})$ increases. The optimal value is reached when $E^t(\gamma^f) = F^t(\gamma^f)$ and $\frac{\partial E^t}{\partial \gamma_i} - \frac{1}{2} \frac{E^t}{F^t} \frac{\partial F^t}{\partial \gamma_i} \big|_{\gamma_i^f} = 0$ as derived in Appendix E.

V. CDMA AND LINEAR ISING PERCEPTRON

Using this notation one defines $\varepsilon_{\mu k} = s_{\mu k} / \sqrt{N}$ for the CDMA problem and $\varepsilon_{\mu k} = s_{\mu k} / \sqrt{K}$ for the Ising perceptron. The goal is to get an accurate estimate of the vector $\mathbf{b}$ for all users given the received message vector $\mathbf{y}$ via a principled approximation of 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). If the arithmetic variance over replicas of the macroscopic message $\Delta_{\mu k}^a$ is finite and independent of the subindexes μ and k, i.e., $\Sigma^2 \equiv \frac{1}{n} \Sigma_a (\Delta_{\mu k}^a)^2 - (\frac{1}{n} \Sigma_a \Delta_{\mu k}^a)^2 < \infty \, \forall \, \mu k$, then $P(y_\mu | \mathbf{B})$ can be expanded as

$$
P(y_\mu | \mathbf{B}) \simeq \sqrt{\frac{n}{2\pi\sigma^2}} e^{\Sigma^2/2\sigma^2} \exp\left(-\frac{(\mathbf{y}_\mu - \boldsymbol{\Delta}_{\mu k})^{\mathsf{T}} (\mathbf{y}_\mu - \boldsymbol{\Delta}_{\mu k})}{2\sigma^2} \right)
$$

$$
\times \left(1 + \frac{\varepsilon_{\mu k}}{\sigma^2} \mathbf{b}_k^{\mathsf{T}} (\mathbf{y}_\mu - \boldsymbol{\Delta}_{\mu k}) \right), \tag{11}
$$

where $\mathbf{y}_\mu = y_\mu \mathbf{u}$ and

$$
\mathbf{u}^{\mathsf{T}} \equiv (\overbrace{1, 1, \ldots, 1}^{nL}).
$$

The function $\mathcal{P}(\vartheta, y_\mu)$, defined in Eq. (C4), and obtained from this distribution is linear in ϑ; therefore, the second derivative used for calculating the messages in Eq. (8) $\mathcal{P}_2 = 0$ and the corresponding structure of the correlation matrix is RS-like.

To calculate correlations between replica we expand $P(y_\mu | \mathbf{B})$ in the large N limit in (11), as shown in Eq. (5). According to the RS correlation assumption, the macroscopic variables satisfy the following relation:

$$
u_{\mu k}^t = \frac{1}{\sqrt{e_1 N}} \sum_{l \neq k} s_{\mu l} m_{\mu l}^t,
$$

$$
X^t \simeq e_2 (1 - N^t),
$$

where $e_1 = 1(\beta)$ for the CDMA (ILP) system and $e_2 = \beta(1)$ for the CDMA (ILP) systems, respectively, due to the change in scaling. The saddle point equation (C6) provides a dominant value for the variable ϑ,
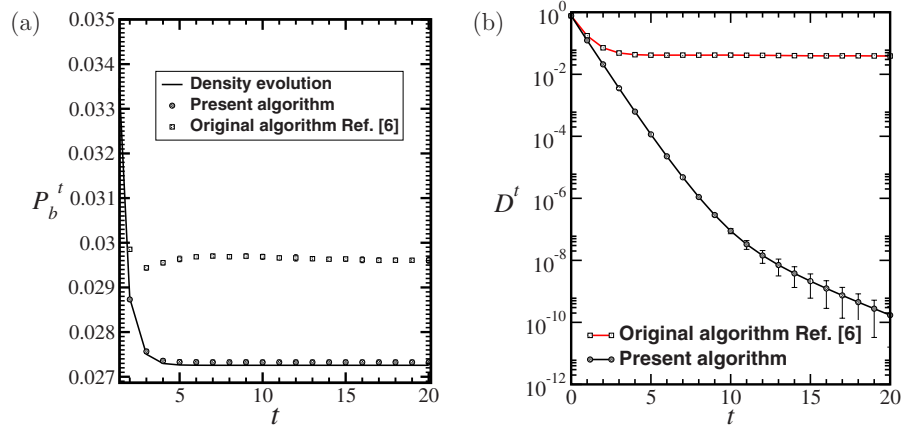
FIG. 3. (Color online) (a) Error probability of the inferred solution evolving in time. The system load $\beta=0.25$, true noise level σ_0^2 $=0.25$, and estimated noise $\sigma^2=0.01$. Squares represent results of the original algorithm [6], solid line represents the dynamics obtained from our equations; circles represent results obtained from the suggested practical algorithm. Variances are smaller than the symbol size. (b) D^t, a measure of convergence for the obtained solutions, as a function of time; symbols are as in (a).

$$\tilde{\vartheta} = \frac{R^t}{\sigma^2 + X^t + R^t} \left(\frac{\sigma^2 u_{\mu k}^t}{X^t + R^t} + y_\mu \right).$$

A. Message

The message from y_μ to b_k^a at time $t+1$ is then given by

$$\hat{m}_{\mu k}^{t+1} = \varepsilon_{\mu k} \frac{y_\mu - u_{\mu k}^t}{\sigma^2 + X^t + R^t}. \tag{12}$$

The main difference between Eq. (12) and the equivalent equation in [6] is the dependence of the prefactor on R^t, reflecting correlations between different solution groups (replica). To determine this term we optimize the choice of σ^2 by applying the condition $E^t=F^t$. Forcing this condition 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 Eq. (12) and using $E^t=F^t$ and $M^t=N^t$ one obtains

$$E^{t+1} = \frac{e_1^{-1}}{\sigma^2 + X^t + R^t}, \quad F^{t+1} = e_1(\sigma_0^2 + X^t)(E^{t+1})^2,$$

which imply, 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 us with a practical way to estimate the true noise variance. Notice that for calculating E^t and F^t we used the limits $K,N \to \infty$ with $K/N=\beta$. So that σ_0^2, which appears in the expression for F^t, can be obtained from the signal vector of y_μ with an infinite number of entries. Thus

$$\lim_{N\to\infty} \frac{1}{N}\sum_{\mu=1}^{N}(y_\mu)^2 = e_2 + \sigma_0^2.$$

Using this expression we can finally express the message as

$$\hat{m}_{\mu k}^{t+1} \simeq \varepsilon_{\mu k} \frac{y_\mu - u_{\mu k}^t}{\frac{1}{N}\sum_{\mu=1}^{N}(y_\mu)^2 - e_2 N^t}, \tag{13}$$

where no prior belief of σ is required.

B. Steady state and critical analysis

The steady state equations for the macroscopic variables N^t and E^t are obtained by taken the limit $t\to\infty$. Let us define $\bar{N} \equiv \lim_{t\to\infty} N^t$ and $\bar{E} \equiv \lim_{t\to\infty} E^t$. In the asymptotic regime the following relations hold:

$$\bar{N}(\sigma_0^2,\beta) = \int \mathcal{D}u \tanh^2[\sqrt{\bar{E}(\sigma_0^2,\beta)}u + \bar{E}(\sigma_0^2,\beta)],$$

$$\bar{E}(\sigma_0^2,\beta) = \frac{e_1^{-1}}{\sigma_0^2 + e_2[1 - \bar{N}(\sigma_0^2,\beta)]} \tag{14}$$

and from these expressions one can obtain the full expression for the error per bit rate,

$$\bar{P}_b(\sigma_0^2,\beta) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\sqrt{\frac{\bar{E}(\sigma_0^2,\beta)}{2}}\right)\right]. \tag{15}$$

C. CDMA signal detection—numerical results

The inference algorithm requires an iterative update of Eqs. (C9) and (13) and converges to a reliable estimate of the signal, with no need for prior information of the noise level. The computational complexity of the algorithm is of $O(K^2)$.

To test the performance of our algorithm we carried out a set of experiments of the CDMA signal detection problem under typical conditions. Error probability of the inferred signals was 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 Fig. 3(a). The solid line represents the
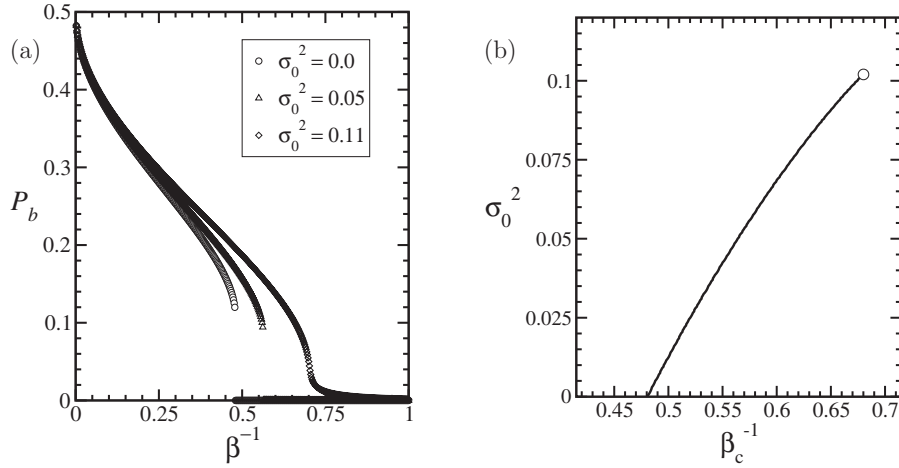
FIG. 4. (a) The error probability $\bar{P}_b$ at the steady state, Eq. (15), as a function of β^{-1} for different values of the noise parameter. For values of σ_0^2 below 0.1025 the curves show discontinuity at certain β values, which becomes continuous but nonanalytic at $\sigma_0^2=0.1025$ around $\beta^{-1}\simeq0.68$. For noise variance values above $\sigma_0^2=0.1025$ the curves become analytical. (b) Position of the nonanalyticity of the error rate curve β_C^{-1} as a function of the noise parameter σ_0^2. This first-order phase-transitionlike curve ends in a second-order phase-transitionlike point marked by ($\bigcirc$).

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 [6].

Another performance measure one should consider is

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

that provides an indication to the stability of the solutions obtained. In Fig. 3(b) we see that results obtained from our algorithm show convergence to a reliable solution in contrast to the original algorithm [6]. The physical interpretation of the difference between the two results is assumed to be related to a replica symmetry breaking phenomenon.

D. Ising linear perceptron—numerical results

For the ILP, the $K>N$ regime of high interest as the system develops a critical behavior for a range of σ_0^2 values. We carried out a set of experiments for this system based on density evolution. In Fig. 4(a) we present curves of the bit error probability $\bar{P}_b$, defined in Eq. (15), as a function of the inverse load β^{-1} for different values of σ_0^2. Three different regimes have been observed: For $\sigma_0^2<0.1025$ the curves exhibit a discontinuity at a value of β that varies with σ_0^2 (first-order phase-transitionlike behavior). At $\sigma_0^2=0.1025$ the curve becomes continuous but its slope diverges (second-order phase-transition-like behavior). The $\bar{P}_b$ curves show analytical behavior for noise values above 0.1025. Figure 4(b) exhibits a phase diagram of the ILP system; it shows the dependency of the critical load β_C^{-1} as a function of the noise parameter. The first-order transition line ends in a second-order transition point marked by a circle. The results ob-

tained, and in particular the critical β value, are consistent with those derived using the replica symmetric statistical mechanics-based analysis of the problem [13].

Another indication for the critical behavior is the number of steps required for the recursive update of Eq. (14) to convergence. In Fig. 5(a) we present the number of iterations required to reach a steady state as a function of β^{-1} when the noise parameter is set to $\sigma_0^2=0.1$. The number of iterations diverges 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. 5(b) we present the result of iterating equations (C9) and (13) for a system size of $K=500$. The curve presents mean values and error bars 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.

VI. CDMA SIGNAL DETECTION WITH DUAL-PEAKED GAUSSIAN NOISE

To demonstrate the suitability of the method for more complex inference problems that require a system with 1RSB-like structures, we will consider the CDMA signal of Eq. (1) where the noise n_μ is drawn from a bi-Gaussian distribution

$$P(n_\mu) = \frac{1-r_0}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(n_\mu+\varepsilon_0/\sigma_0)^2}{2}\right)$$
$$+ \frac{1+r_0}{2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(n_\mu-\varepsilon_0/\sigma_0)^2}{2}\right), \quad (16)$$

where $r_0 \in (-1,1)$ represents the bias and $\pm\varepsilon_0/\sigma_0$ the positions of the Gaussian peaks. We consider the particular case where $|\varepsilon_0/\sigma_0|\ll1$, so that the Gaussian peaks are slightly off center. For this model the likelihood expression takes the form
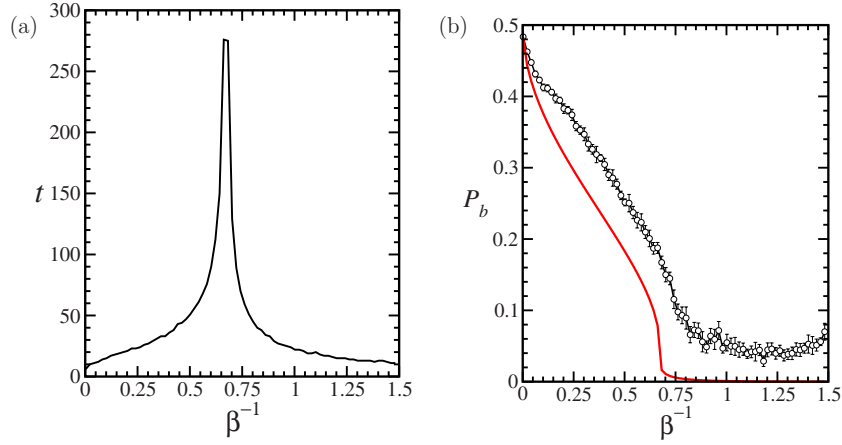
FIG. 5. (Color online) (a) Number of iterations of Eq. (14) required for convergence as a function of β, for $\sigma_0^2=0.10$; one clearly identifies the β value where the error rate curve exhibits a discontinuity. (b) Finite size effects are observed at all β values. The noise level used is $\sigma_0^2=0.10$ with $K=500$. The curves provide mean values and error bars over 1000 experiments. The solid curve obtained from the iteration of the steady state equations is presented as a reference.

$$P(y_\mu|\Delta_\mu;r,\varepsilon,\sigma^2) \propto \prod_{\ell=1}^{L}\prod_{a=1}^{n}\left[\frac{1-r}{2}\exp\left(-\frac{(y_\mu-\Delta_\mu^{\ell a}+\varepsilon)^2}{2\sigma^2}\right)\right.$$
$$\left.+\frac{1+r}{2}\exp\left(-\frac{(y_\mu-\Delta_\mu^{\ell a}-\varepsilon)^2}{2\sigma^2}\right)\right],$$

where r, ε, and σ^2 are estimates of the true parameters r_0, ε_0, and σ_0^2.

To derive the messages in this case we first calculate the function $\mathcal{P}(\vartheta,y_\mu)$ of Eq. (C4), which has the form

$$\mathcal{P}(\vartheta,y_\mu)=\frac{y_\mu-\vartheta}{\sigma^2+X^t}-\frac{\varepsilon}{\sigma^2+X^t}\tanh\left(\varepsilon\frac{y_\mu-\vartheta}{\sigma^2+X^t}+\text{arctanh}(r)\right),$$

where $X^t=\beta(1-N^t)$.

Following the derivation of Appendix C, the saddle point equations (D1) and (D2) can be expressed as

$$\tilde{\vartheta}_{\mu k}^t=u_{\mu k}^t+W^t\mathcal{P}(\tilde{\vartheta}_{\mu k}^t,y_\mu),$$

$$y_\mu-\tilde{\vartheta}_{\mu k}^t=y_\mu-u_{\mu k}^t-W^t\frac{y_\mu-\tilde{\vartheta}_{\mu k}^t}{\sigma^2+X^t}$$
$$+\varepsilon\frac{W^t}{\sigma^2+X^t}\tanh\left(\varepsilon\frac{y_\mu-\tilde{\vartheta}_{\mu k}^t}{\sigma^2+X^t}+\text{arctanh}(r)\right),$$

$$\frac{y_\mu-\tilde{\vartheta}_{\mu k}^t}{\sigma^2+X^t}=\frac{y_\mu-u_{\mu k}^t}{\sigma^2+X^t+W^t}+\frac{\varepsilon}{\sigma^2+X^t}\frac{W^t}{\sigma^2+X^t+W^t}$$
$$\times\tanh\left(\varepsilon\frac{y_\mu-\tilde{\vartheta}_{\mu k}^t}{\sigma^2+X^t}+\text{arctanh}(r)\right),$$

$$z=\rho_W(y_\mu-u_{\mu k}^t)+\varepsilon(\rho_0-\rho_W)\tanh[\varepsilon z+\text{arctanh}(r)]$$
$$\simeq z_0+r\Delta\rho_W\varepsilon+(1-r^2)(1-3r^2)\Delta\rho_W z\varepsilon^2$$
$$-r(1-r^2)\Delta\rho_W z^2\varepsilon^3-\frac{1}{3}(1-r^2)(1-3r^2)\Delta\rho_W z^3\varepsilon^4,$$

where we denote $W^t=R^t$ for the RS case and $W^t=2V^t-R^t$ for the 1RSB case, $z\equiv\frac{y_\mu-\tilde{\vartheta}_{\mu k}^t}{\sigma^2+X^t}$, $\rho_A\equiv(\sigma^2+X^t+A)^{-1}$, $z_0\equiv\rho_W(y_\mu-u_{\mu k}^t)$, and $\Delta\rho_W\equiv\rho_0-\rho_W$.

The solution of this equation provides, up to order $O(\varepsilon^4)$,

$$z(\varepsilon)\simeq z_0+r\Delta\rho_W\varepsilon+(1-r^2)\Delta\rho_W\big[z_0\varepsilon^2+r(\Delta\rho_W-z_0^2)\varepsilon^3$$
$$+(1-3r^2)z_0\big(\Delta\rho_W-\tfrac{1}{3}z_0^2\big)\varepsilon^4\big].$$

The function $\mathcal{P}$ and its two first derivatives at the saddle point value are

$$\mathcal{P}_0=-r[1+(1-r^2)\Delta\rho_W\varepsilon^2]\rho_W\varepsilon+[1-(1-r^2)\rho_W^2\varepsilon^2$$
$$-(1-r^2)(1-3r^2)\Delta\rho_W\rho_W^2\varepsilon^4]$$
$$\times(y_\mu-u_{\mu k}^t)+r(1-r^2)\rho_W^3(y_\mu-u_{\mu k}^t)^2\varepsilon^3$$
$$+\tfrac{1}{3}(1-r^2)(1-3r^2)\rho_W^4(y_\mu-u_{\mu k}^t)^3\varepsilon^4,$$

$$\mathcal{P}_1\simeq-\rho_0+O(\varepsilon^2),$$

$$\mathcal{P}_2=2\rho_0^3(1-r^2)[r\varepsilon^3+(1-3r^2)\rho_W(y_\mu-u_{\mu k}^t)\varepsilon^4],$$

therefore, one can obtain the following expression, required for calculating the messages in the 1RSB case (C8):

$$\frac{1}{2}\frac{\mathcal{P}_2 V^t}{1-\mathcal{P}_1 V^t}=(1-r^2)\rho_0\Delta\rho_V[r\varepsilon^3+(1-3r^2)\rho_W(y_\mu-u_{\mu k}^t)\varepsilon^4],$$

where $\Delta\rho_V\equiv\rho_0-\rho_V$. This straightforwardly leads to the following expression for the message:

$$^{(1RSB)}\hat{m}_{\mu k}^{t+1} = \frac{s_{\mu k}}{\sqrt{N}}\Bigg(-[\rho_W + (1-r^2)(\Upsilon_n - \rho_W^2)\varepsilon^2]r\varepsilon + \rho_W[1-(1$$

$$-r^2)\rho_W\varepsilon^2 - (1-r^2)(1-3r^2)(\Upsilon_n - \rho_W^2)\varepsilon^4](y_\mu$$

$$-u_{\mu k}^t) + r(1-r^2)\rho_W^3\varepsilon^3(y_\mu - u_{\mu k}^t)^2 + \frac{1}{3}(1-r^2)(1$$

$$-3r^2)\rho_W^4\varepsilon^4(y_\mu - u_{\mu k}^t)^3\Bigg), \qquad (17)$$

where $\Upsilon_n \equiv \rho_0(\rho_W - \frac{1}{n}\Delta\rho_V)$. The expression for the message in the RS case is recovered from Eq. (17) in the limit $n \to \infty$.

A. Optimization and messages

Calculating the expressions for the macroscopic variables E^{t+1} and F^{t+1}, used in the optimization process, requires performing the following sums, in the limit of $K, N \to \infty$ with $K/N = \beta < \infty$:

$$A_j \equiv \lim_{K,N\to\infty} \sum_\mu^N \frac{1}{K}\sum_{k=1}^K \frac{s_{\mu k}b_k}{\sqrt{N}}(y_\mu - u_{\mu k}^t)^j,$$

$$B_l \equiv \lim_{K,N\to\infty} \frac{1}{N}\sum_\mu^N \frac{1}{K}\sum_{k=1}^K (y_\mu - u_{\mu k}^t)^l,$$

where $j = 0, \ldots, 3$ and $l = 0, \ldots, 4$. From the definition of the signal y_μ (1) and the expression for the noise (16) we find that $A_0 = 0$, $A_1 = 1$, $A_2 = 2B_1$, $A_3 = 3B_2$, $B_0 = 1$, $B_1 = r_0\varepsilon_0$, $B_2 = \beta(1 - 2M^t + N^t) + \sigma_0^2 + \varepsilon_0^2$, $B_3 = B_1(3B_2 - 2\varepsilon_0^2)$, and $B_4 = 3B_2^2 - 2\varepsilon_0^4$. The explicit expressions derived for the macroscopic variables are

$$E^{t+1} = \rho_W - (1-r^2)\rho_W^2\varepsilon^2 + 2r(1-r^2)B_1\rho_W^3\varepsilon^3$$
$$- (1-r^2)(1-3r^2)[\Upsilon_n - (1 + B_2\rho_r)\rho_W^2]\rho_W\varepsilon^4,$$

$$F^{t+1} = B_2\rho_W^2 - 2rB_1\rho_W^3\varepsilon + [r^2 - 2(1-r^2)B_2\rho_W]\rho_W^2\varepsilon^2$$
$$- 2r(1-r^2)B_1[\Upsilon_n - (2 + 3B_2\rho_W)\rho_W^2]\rho_W\varepsilon^3$$
$$+ (1-r^2)[2r^2(\Upsilon_n - \rho_W^2)\rho_W + (1-3r^2)$$
$$\times B_2(3\rho_W^2 + 2B_2\rho_W^3 - 2\Upsilon_n)\rho_W^2]\varepsilon^4.$$

Applying the optimization conditions of Appendix E, $E^t(\gamma^f) = F^t(\gamma^f)$ and $\frac{\partial E^t}{\partial\gamma_i} - \frac{1}{2}\frac{E^t}{F^t}\frac{\partial F^t}{\partial\gamma_i}|_{\gamma_i^f} = 0$, where $\gamma^{\mathsf{T}} = (r, \varepsilon, \sigma^2, \frac{1}{n})$ one obtains the following conditions:

$$\rho_W = \frac{1}{B_2} + \frac{\varepsilon^2}{B_2^2} - \frac{\varepsilon^4}{B_2^3} + (1-r^2)^2\frac{1 - B_2\rho_0\left(1 - \frac{1}{n}B_2\Delta\rho_V\right)}{B_2^3}\varepsilon^4, \qquad (18)$$

$$r\varepsilon = B_1 + r(1-r^2)\frac{1 - B_2\rho_0\left(1 - \frac{1}{n}B_2\Delta\rho_V\right)}{B_2}\varepsilon^3. \qquad (19)$$

In the 1RSB case one can further simplify these expressions by a suitable choice of V^t and the number of replicas

per block n. Optimization with respect to the latter results in

$$1 = B_2\rho_0\left(1 - \frac{1}{n}B_2\Delta\rho_V\right), \qquad (20)$$

which implies

$$V^t = \frac{(X^t + \sigma^2)^2(\sigma_0^2 - \sigma^2)}{\frac{1}{n}(X^t + \sigma_0^2)^2 - (X^t + \sigma^2)(\sigma_0^2 - \sigma^2)},$$

that by definition is larger than zero. This condition is satisfied if our estimate for the noise variance is smaller than the true parameter ($\sigma^2 < \sigma_0^2$). In this case the number of replicas per block must satisfy the condition

$$1 \leqslant n \leqslant f(X^t; \sigma_0^2, \sigma^2) \equiv \frac{(X^t + \sigma_0^2)^2}{(X^t + \sigma^2)(\sigma_0^2 - \sigma^2)}.$$

Interestingly this ties the noise level mismatch to the number of replicas, thus giving further insight to the role played by the structure of the inter-replica correlation matrix.

For $0 \leqslant X^t$, the minimum value of $f(X^t; \sigma_0^2, \sigma^2)$ is reached at $X_{min} = \max(0, \sigma_0^2 - 2\sigma^2)$. It is also possible to prove that $4 \leqslant f(X_{min}; \sigma_0^2, \sigma^2)$. Although V^t and n will not be explicitly used in the following expressions, the correct choice of the value for these parameters allows one to use Eqs. (18) and (19) in order to find the final expression for the macroscopic variable E^{t+1}, where no estimates are needed for the noise parameters:

$$^{(1RSB)}E^{t+1} = \frac{1}{B_2 - B_1^2}.$$

Note that in the RS case we do not have the freedom to choose the number of replicas per block, given that this case is equivalent to take $n \to \infty$ in the absence of the additional replica $l = 1, \ldots, L$. For this reason Eqs. (18) and (19) take the form

$$\rho_W = \frac{1}{B_2} + \frac{\varepsilon^2}{B_2^2} - \frac{\varepsilon^4}{B_2^3} + (1-r^2)^2\frac{1 - B_2\rho_0}{B_2^3}\varepsilon^4, \qquad (21)$$

$$r\varepsilon = B_1 + r(1-r^2)\frac{1 - B_2\rho_0}{B_2}\varepsilon^3, \qquad (22)$$

and the macroscopic variable

$$^{(RS)}E^{t+1} = {}^{(1RSB)}E^{t+1} + \frac{2B_1^2(\varepsilon^2 - B_1^2)}{B_2^3}\left(\frac{B_2}{X^t + \sigma^2} - 1\right),$$

which depends on both estimates of the noise variance σ^2 and bias ε.

Given that the algorithm deals with finite signal vectors ($N < \infty$), the quantities B_1 and B_2 must be approximated by the correspondent finite sums. Therefore, we have

$$B_1 = \lim_{N,K\to\infty} \frac{1}{N}\sum_{\mu=1}^N \frac{1}{K}\sum_{k=1}^K (y_\mu - u_{\mu k}^t) \approx \frac{1}{N}\sum_{\mu=1}^N y_\mu \equiv \bar{B}_1, \qquad (23)$$
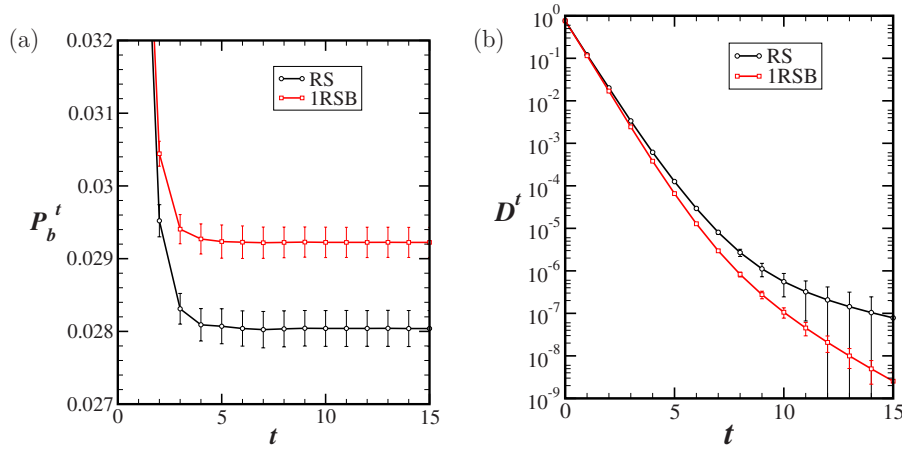
FIG. 6. (Color online) (a) Error probability of the inferred solution evolving in time, for the bi-Gaussian noise case. The system load $\beta=0.25$, true noise level $\sigma_0^2=0.25$ and estimated noise $\sigma^2=0.01$. Squares represent results of the RS algorithm and circles represent results obtained from the 1RSB algorithm. (b) D^t, a measure of convergence in the obtained solutions, as a function of time; symbols are as in (a).

$$B_2 = \lim_{N,K\to\infty} \frac{1}{N}\sum_{\mu=1}^{N}\frac{1}{K}\sum_{k=1}^{K}(y_\mu - u_{\mu k}^t)^2 \approx \frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2 + \beta N^t \equiv \bar{B}_2,$$

where we used the fact that $\lim_{N,K\to\infty}\frac{1}{NK}\Sigma_{\mu,k}u_{\mu k}^t=0$. Observe that no information about the true noise has been used to derive these expressions.

Having the estimates (23) we can write the messages explicitly,

$$^{(1RSB)}\hat{m}_{\mu k}^{t+1} = \frac{s_{\mu k}}{\sqrt{N}}\left[-\frac{\bar{B}_1}{\bar{B}_2} + \frac{\bar{B}_1}{\bar{B}_2^2}\varepsilon^2 + \left(\frac{1}{\bar{B}_2} + \frac{\bar{B}_1^2}{\bar{B}_2^2} - \frac{3\varepsilon^2 - 2\bar{B}_1^2}{\bar{B}_2^3}\varepsilon^2\right) \right.$$

$$\times(y_\mu - u_{\mu k}^t) + \frac{\bar{B}_1(\varepsilon^2 - \bar{B}_1^2)}{\bar{B}_2^3}(y_\mu - u_{\mu k}^t)^2$$

$$\left. + \frac{1}{3}\frac{(\varepsilon^2 - \bar{B}_1^2)(\varepsilon^2 - 3\bar{B}_1^2)}{\bar{B}_2^4}(y_\mu - u_{\mu m}^t)^3 \right],$$

$$^{(RS)}\hat{m}_{\mu k}^{t+1} = {}^{(1RSB)}\hat{m}_{\mu k}^{t+1} + \frac{s_{\mu k}}{\sqrt{N}}\left(1 - \frac{\bar{B}_2}{X^t + \sigma^2}\right)\frac{\varepsilon^2 - \bar{B}_1^2}{\bar{B}_2^2}$$

$$\times\left(\bar{B}_1 + 2\frac{\varepsilon^2 - 2\bar{B}_1^2}{\bar{B}_2^2}(y_\mu - u_{\mu k}^t)\right),$$

which can be now used recursively for obtaining the inferred solutions for this problem. Notice that an estimate of *both* ε and σ in required in the RS case.

B. Numerical results

To test the performance of the 1RSB algorithm we carried out a set of experiments of the CDMA signal detection problem with bi-Gaussian noise. The results shown in Fig. 6(a) describe the error probability of the inferred signals as a function of the number of iterations has been calculated using both RS and 1RSB-like correlation matrices for the case of parameters mismatch. The system load used in the simu-

lations was $\beta=0.25$, the true noise level $\sigma_0^2=0.25$, Gaussian bias of $\varepsilon_0=0.06$ and weight $r_0=0.6$. The estimated noise parameters are $\sigma^2=0.01$ and $\varepsilon=0.2$. The circles represent simulation results obtained via the 1RSB algorithm while the squares correspond to the RS-like structure. The results presented are based on 10^5 trials per point and a system size $N=1000$; error bars are also provided. The results obtained using the 1RSB-like structure are superior to those obtained using the RS algorithm. As shown in Fig. 6(b) using the stability measure D^t, both RS and 1RSB-based algorithms converge to reliable solutions; the 1RSB-based algorithm is slightly slower to converge, presumably due to the more complex message passing scheme.

VII. CONCLUSIONS

We present and methodologically develop an algorithm for using BP in densely connected systems that enables one to obtain reliable solutions even when the solution space is fragmented. The algorithm relies on the introduction of a large number of replicated variable systems exposed to the same evidential nodes. Messages are obtained by averaging over all replicated systems leading to pseudoposterior, that is then used to infer the variable nodes most probable values. This is done with no actual replication, by introducing an assumption about correlations between the replicated variables and exploiting the high number of replicated systems. The algorithm was developed in a systematic manner to accommodate more complex correlation matrices. It was successfully applied to the CDMA signal detection and ILP learning problems, using the RS-like correlation matrix, and to the CDMA inference problem with bimodal Gaussian noise model in the 1RSB-like correlation matrix. The algorithm provides superior results to other existing algorithms [6,20] and a systematic improvement where more complex correlation matrices are introduced, where required.

Further research is required to fully determine the potential of the algorithm. Two particular areas which we consider as particularly promising are inference problems character-

ized by discrete data variables and noise model and problems that can be mapped onto *sparse* graphs. Both activities are currently underway.

APPENDIX: A THE REPLICA SYMMETRIC ANSATZ

Within the RS setting, the interaction term in Eq. (6) is

$$\mathbf{b}_k^{\mathsf{T}}\mathbf{Q}_{\mu k}^t \mathbf{b}_k = n(q_{0\mu k}^t - q_{1\mu k}^t) + q_{1\mu k}^t \left(\sum_{a=1}^n b_k^a\right)^2,$$

A simplified expression for Eq. (6) immediately follows

$$P^t(\mathbf{b}_k|\{y_{\nu \neq \mu}\}) = [\mathcal{Z}_{\mu k}^t]^{-1} \exp\left[h_{\mu k}^t \sum_{a=1}^n b_k^a + \frac{1}{2} q_{1\mu k}^t \left(\sum_{a=1}^n b_k^a\right)^2\right]$$

$$= [\mathcal{Z}_{\mu k}^t]^{-1} \int_{-\infty}^{\infty} dx \exp\left[-\frac{x^2}{2q_{1\mu k}^t} + (x + h_{\mu k}^t)\sum_{a=1}^n b_k^a\right],$$

where $\mathcal{Z}_{\mu k}^t$ is a normalization constant. The diagonal elements $q_{0\mu k}^t$ only affect the normalization term and can therefore be taken to zero with no loss of generality.

We expect the logarithm of the normalization term $\mathcal{Z}_{\mu k}^t$ (linked to the free energy), obtained from the well-behaved distribution P^t, to be self-averaging. We therefore expect

$$\lim_{n\to\infty} \frac{1}{n} \ln(\overline{\mathcal{Z}_{\mu k}^t}) = \lim_{n\to\infty} \frac{1}{n} \ln[\mathcal{Z}_{\mu k}^t(\hat{h}, \hat{q}_1)],$$

where $\hat{h}$ and $\hat{q}_1$ are the mean values of the parameters drawn for some suitable distributions and the overline represents the mean value of the partition function over these distributions.

In the following we will drop the upper-index t and the subindices μ and k for brevity. To obtain the scaling behavior of the various parameters one calculates $\mathcal{Z}(h, q_1)$ explicitly, assuming the parameter q_1 is taken from a normal distribution $\mathcal{N}(\hat{q}_1, \sigma_q^2)$. The partition function takes the form

$$\mathcal{Z}(h, q_1) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi q_1}} \exp\left(-\frac{(x-h)^2}{2q_1} + n \ln[2 \cosh(x)]\right).$$

$$(A1)$$

Thus, the mean value of the partition function over the set of parameters is

$$\overline{\mathcal{Z}(h, q_1)} = \int \mathcal{D}q_1 \mathcal{Z}(h, q_1),$$

where $\mathcal{D}q_1 = dq_1 \mathcal{N}(\hat{q}_1, \sigma_{q_1}^2)$. The normalization can be expressed as

$$\overline{\mathcal{Z}(h, q_1)} = \sum_{a=0}^n \binom{n}{a} \exp\left\{n\left[h\left(1 - \frac{2a}{n}\right) + \frac{\hat{q}_1}{2}\left(1 - \frac{2a}{n}\right)^2 n\right.\right.$$

$$\left.\left. + \frac{\sigma_{q_1}^2}{8}\left(1 - \frac{2a}{n}\right)^4 n^3\right]\right\} = \mathcal{A}(n)(n+1)$$

$$\times \binom{n}{n/2} \exp\left[n\left(|h| + n\frac{\hat{q}_1}{2} + n^3\frac{\sigma_{q_1}^2}{8}\right)\right]$$

$$\simeq \sqrt{\frac{2}{\pi}} \mathcal{A}(n) \exp\left[n\left(\ln(2) + |h| + n\frac{\hat{q}_1}{2} + n^3\frac{\sigma_{q_1}^2}{8}\right)\right],$$

where $\mathcal{A}(n) \sim O(1)$. Thus, $h \sim O(1)$, $\hat{q}_1 \sim O(n^{-1})$, and $\sigma_{q_1}^2 \sim O(n^{-3})$. From now on we will take the off-diagonal elements of the RS matrix $\mathbf{Q}_{\mu k}^t$ equal to $g_{1\mu k}^t/n$, where $g_{1\mu k}^t \sim O(1)$.

The form of the marginalized posterior at time t is then

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

$$(A2)$$

where

$$\Phi(x; h_{\mu k}^t, g_{1\mu k}^t) = \frac{(x - h_{\mu k}^t)^2}{2g_{1\mu k}^t} - \ln[2 \cosh(x)].$$

The function $\Phi(x; h, g_1)$ presents one or two minima according to the following:

h	g_1	Number of minima
$h \in \mathbb{R}$	$0 < g_1 \leq 1$	one minima
$|h| = h_c$	$g_1 > 1$	one minima and one hump
$|h| < h_c$	$g_1 > 1$	two minima

where $h_c = \sqrt{g_1(g_1 - 1)} - \cosh^{-1}(\sqrt{g_1})$; the coefficient g_1 plays the role of the inverse temperature. Below the critical value $g_{1c} = 1$ a spontaneous magnetization appears.

This results from analyzing the equation

$$\frac{\partial \Phi(x; h, g_1)}{\partial x} = \frac{x - h}{g_1} - \tanh(x) = 0. \qquad (A3)$$

The case of two maxima is presented in Fig. 7.

We define the mean values from the distribution equation (A2). If the field h is not zero, as shown in Fig. 7, $[\exp(-\Phi)]^n$ develops one dominant maximum as $n \to \infty$. For large enough n, only this maximum contributes to the integrals (A2) and the algorithm obtained from this assumption turns out to be the same as the one presented in [6]. However, if the field is sufficiently small it gives rise to a regime where the two maxima contribute. At the same time, it is important to note that a small, nonzero field favors the solution of Eq. (A3) that satisfies sgn(x)=sgn(h). To analyze the behavior of the field, we will explore the solutions of Eq.
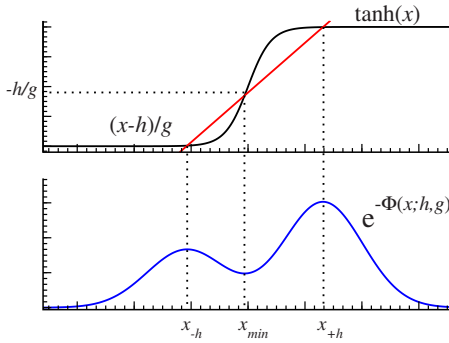
FIG. 7. (Color online) Solutions for the mean field equation (A3) with two maxima and one minimum for a positive value of the field h.

(A3) in the regime $0 \lesssim |h| \ll 1$. With this aim, suppose that the solutions for Eq. (A3) at zero field are $x_0 = \pm g_1|m|$ where $m \equiv \tanh(x_0)$ and $\text{sgn}(m) = \text{sgn}(h)$. If the field is sufficiently small one can expand the solutions of Eq. (A3) as $x_{\pm h} = \pm g_1 m + \xi(m, g_1)h$ where $\xi(m, g_1)h$ is expected to be small and satisfies $\text{sgn}[\xi(m, g_1)h] = \text{sgn}(h)$. Observe that if the field is positive (negative), both roots are displaced to the right (left) with respect to the zero field solutions. Using this expression for the roots in Eq. (A3) and disregarding terms of $O(h^2)$ one finds that

$$\xi(m, g_1) = \frac{1}{1 - g_1(1 - m^2)}. \tag{A4}$$

The expression for the exponent Φ near the roots and in the $0 \lesssim |h| \ll 1$ regime is then $\Phi(x_{\pm h}; h \to 0, g_1) \simeq \Phi(x_0; 0, g_1) \mp mh = \Phi_0 \mp mh$, and, by the definition of the m, the product mh is positively defined.

Let us define $\beta_{\pm h}(m, g_1) \equiv (1 - m^2)[1 \mp 2\xi(m, g_1)mh]$. We expect that, for large n, the following approximation is valid:

$$\exp\{-n\Phi(x; h \to 0, g_1)\} \simeq e^{-n\Phi_0}\left[e^{nmh} \exp\left(-\frac{n}{2}[g_1^{-1}\right.\right.$$
$$\left. - \beta_h(m, g_1)](x - x_h)^2\right) + e^{-nmh} \exp\left(-\frac{n}{2}[g_1^{-1}\right.$$
$$\left.\left. - \beta_{-h}(m, g_1)](x - x_{-h})^2\right)\right]. \tag{A5}$$

Using Eq. (A5) one can calculate the normalization in Eq. (A1),

$$\mathcal{Z}(h \to 0, g_1) \simeq e^{-n(\Phi_0 - mh)} \int dx \exp\left(-\frac{n}{2}[g_1^{-1} - \beta_h(m, g_1)](x\right.$$
$$\left. - x_h)^2\right) + e^{-n(\Phi_0 + mh)} \int dx \exp\left(-\frac{n}{2}[g_1^{-1} - \beta_{-h}(m, g_1)]\right.$$
$$\left.\times (x - x_{-h})^2\right) \simeq \sqrt{\frac{2\pi g_1 \xi(m, g_1)}{n}} e^{-n\Phi_0}\{e^{nmh}[1 - g_1(1$$
$$- m^2)\xi^2(m, g_1)mh] + e^{-nmh}[1 + g_1(1$$
$$- m^2)\xi^2(m, g_1)mh]\}. \tag{A6}$$

The mean value of a given function $f(x)$ with respect to the conditional probability distribution defined in Eq. (A2) is then

$$\langle f(x)|h \to 0, g_1\rangle \simeq \mathcal{Z}^{-1}e^{-n(\Phi_0 - mh)} \int dx \exp\left(-\frac{n}{2}\{g_1^{-1}\right.$$
$$\left. - (1 - m^2)[1 - 2\xi(m, g_1)mh]\}(x - x_h)^2\right)$$
$$\times \left(f(x_h) + (x - x_h)f'(x_h) + \frac{1}{2}(x - x_h)^2 f''(x_h)\right)$$
$$+ \mathcal{Z}^{-1}e^{-n(\Phi_0 + mh)} \int dx \exp\left(-\frac{n}{2}\{g_1^{-1} - (1 - m^2)\right.$$
$$\left.\times [1 + 2\xi(m, g_1)mh]\}(x - x_{-h})^2\right)\left(f(x_{-h}) + (x - x_{-h})\right.$$
$$\left.\times f'(x_{-h}) + \frac{1}{2}(x - x_{-h})^2 f''(x_{-h})\right),$$

which implies, considering that the integrals of the linear terms are zero and keeping only the leading terms in the expansions, that the expectation values takes the form

$$\langle f(x)|h \to 0, g_1\rangle \simeq \{1 - e^{-2nmh}[1 + 2\xi^2(m, g_1)mh]\}$$
$$\times \left(f(x_h) + \frac{g_1}{2n}\xi(m, g_1)f''(x_h)\right)$$
$$+ e^{-2nmh}[1 + 2\xi^2(m, g_1)mh]f(x_{-h}).$$

Considering the expansion of $f(x_{\pm h}) \simeq f[\pm g_1 m + \xi(m, g_1)h] \simeq f(\pm g_1 m) + \xi(m, g_1)f'(\pm g_1 m)h$ and disregarding terms of $O(he^{-2nmh})$, one can write

$$\langle f(x)|h \to 0, g_1\rangle \simeq f(mg_1) + \frac{g_1}{2n}\xi(m, g_1)f''(mg_1)$$
$$- e^{-2nmh}[f(mg_1) - f(-mg_1)] + f'(mg_1)\xi(m, g_1)h. \tag{A7}$$

The resulting one and two variable expectation values become

$$\langle b_k^a|h_{\mu k}^t \to 0, g_{\mu k}^t\rangle = \sum_{\{\mathbf{b}_k\}} P^t(\mathbf{b}_k|\{y_{\nu \neq \mu}\})b_k^a$$
$$= \langle \tanh(x)|h_{\mu k}^t \to 0, g_{1\mu k}^t\rangle$$
$$\simeq \left(1 - \frac{g_{1\mu k}^t}{n}[1 - (m_{\mu k}^t)^2]\xi(m_{\mu k}^t, g_{1\mu k}^t)\right.$$
$$\left. - 2e^{-2nm_{\mu k}^t h_{\mu k}^t}\right)m_{\mu k}^t + \xi(m_{\mu k}^t, g_{1\mu k}^t)[1 - (m_{\mu k}^t)^2]h_{\mu k}^t$$

and

$$\langle b_k^a b_k^b|h_{\mu k}^t \to 0, g_{1\mu k}^t\rangle = P^t(\mathbf{b}_k|\{y_{\nu \neq \mu}\})b_k^a b_k^b$$
$$= \delta^{ab} + (1 - \delta^{ab})\langle \tanh^2(x)|h_{\mu k}^t \to 0, g_{1\mu k}^t\rangle,$$

where

$$\langle \tanh^2(x) | h_{\mu k}^t \to 0, g_{1\mu k}^t \rangle = (m_{\mu k}^t)^2 + \xi(m_{\mu k}^t, g_{1\mu k}^t)[1 - (m_{\mu k}^t)^2]$$

$$\times \left(\frac{g_{1\mu k}^t}{n}[1 - 3(m_{\mu k}^t)^2] + 2m_{\mu k}^t h_{\mu k}^t \right),$$

and

$$\langle b_k^a b_l^b | h_{\mu k}^t \to 0, g_{1\mu k}^t \rangle = \langle b_k^a | h_{\mu k}^t \to 0, g_{1\mu k}^t \rangle \langle b_l^b | h_{\mu k}^t \to 0, g_{1\mu k}^t \rangle.$$

Thus, the leading terms for the covariance matrix of the replicated variables are

$$(\mathbf{\Psi}_{\mu kl}^t)^{ab} \equiv \langle b_k^a b_l^b | h_{\mu k}^t \to 0, g_{1\mu k}^t; h_{\mu l}^t \to 0, g_{1\mu l}^t \rangle - \langle b_k^a | h_{\mu k}^t$$
$$\to 0, g_{1\mu k}^t \rangle \langle b_l^b | h_{\mu l}^t \to 0, g_{1\mu l}^t \rangle = \delta_{kl}(\mathbf{\Psi}_{\mu k}^t)^{ab}$$

$$(\mathbf{\Psi}_{\mu k}^t)^{ab} \simeq \delta^{ab}[1 - (m_{\mu k}^t)^2] + (1 - \delta^{ab})\left(\frac{g_{1\mu k}^t}{n}\xi(m_{\mu k}^t, g_{1\mu k}^t)[1 \right.$$

$$\left. - (m_{\mu k}^t)^2]^2 + 4e^{-2nm_{\mu k}^t h_{\mu k}^t}(1 - e^{-2nm_{\mu k}^t h_{\mu k}^t})(m_{\mu k}^t)^2 \right).$$

If one requires the nondiagonal elements of this covariance matrix to have the same scaling as the inter-replica interaction matrix, the field must behave in such a way that the exponential term contributes at most in $O(n^{-1})$. One thus expects the field to obey $m_{\mu k}^t h_{\mu k}^t < \frac{1}{n} \ln \left| \frac{2n}{n_{\mu k}^t} \right|$, where the $n_{\mu k}^t$ are appropriate constants. With this asymptotic behavior, the expression for the entries in the covariance matrix is

$$(\mathbf{\Psi}_{\mu k}^t)^{ab} \simeq \delta^{ab}[1 - (m_{\mu k}^t)^2]$$

$$+ (1 - \delta^{ab})\frac{g_{1\mu k}^t \xi(m_{\mu k}^t, g_{1\mu k}^t)}{n}[1 - (m_{\mu k}^t)^2]^2,$$

which serves to define the probability distribution for the macroscopic variable $\Delta_{\mu k}^a = \Sigma_{l \neq k} \varepsilon_{\mu l} b_l^a$.

As $\varepsilon_{\mu k}$ and b_k^a are unbiased variables, the variable $\Delta_{\mu k}^a$, by virtue of the central limit theorem, obeys a normal distribution, with mean value and covariance matrix given by (to highest order)

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

$$(A8)$$

$$(\mathbf{Y}_{\mu k}^t)^{ab} \equiv \langle \Delta_{\mu k}^a \Delta_{\mu k}^b \rangle - \langle \Delta_k^a \rangle \langle \Delta_k^b \rangle$$

$$= \sum_{\{\mathbf{b}_{l \neq k}\}} \prod_{l \neq k} P^t(\mathbf{b}_l | \{y_{\nu \neq \mu}\}) \sum_{\substack{l \neq k \\ j \neq k}} \varepsilon_{\mu l} \varepsilon_{\mu j} b_l^a b_j^b$$

$$- \left(\sum_{l \neq k} \varepsilon_{\mu l} m_{\mu l}^t \right)^2$$

$$= \sum_{l \neq k} \varepsilon_{\mu l}^2 (\mathbf{\Psi}_{\mu lj}^t)^{ab}$$

$$= \delta^{ab} X_{\mu k}^t + (1 - \delta^{ab})\frac{1}{n} R_{\mu k}^t,$$

where

$$X_{\mu k}^t \equiv \sum_{l \neq k} \varepsilon_{\mu l}^2 [1 - (m_{\mu l}^t)^2],$$

$$R_{\mu k}^t \equiv \sum_{l \neq k} \varepsilon_{\mu l}^2 g_{1\mu l}^t \xi(m_{\mu l}^t, g_{1\mu l}^t)[1 - (m_{\mu l}^t)^2]^2, \quad (A9)$$

are macroscopic variables of $O(1)$. In particular, $R_{\mu k}^t$ is a free variable that can be used later on to optimize a given performance measure. This variables have the property of being self-averaging, therefore we can drop the subindices μ and k.

APPENDIX B: THE ONE-STEP REPLICA SYMMETRY BREAKING ANSATZ

Under a solution correlation matrix that resembles the 1RSB structure, the system comprises nL variables, where both the number of blocks L and the number of variables per block n are considered large. As before we are interested in the regime where L and $n \to \infty$.

With this setting, the interaction term in Eq. (6) is now

$$\mathbf{b}_k^\mathsf{T} \mathbf{Q}_{\mu k}^t \mathbf{b}_k = - q_{1\mu k}^t nL + (q_{1\mu k}^t - q_{2\mu k}^t) \sum_{\ell=1}^L \left(\sum_{a=1}^n b_k^{\ell a} \right)^2$$

$$+ q_{2\mu k}^t \left(\sum_{\ell=1}^L \sum_{a=1}^n b_k^{\ell a} \right)^2,$$

thus we have now $L+1$ squared sums in the exponent that can be replaced by integrals

$$P^t(\mathbf{b}_k | \{y_{\nu \neq \mu}\}) = [\mathcal{Z}_{\mu k}^t]^{-1} \int dx \exp\left(- \frac{x_0^2}{2q_{2\mu k}^t} - \sum_{\ell=1}^L \frac{x_\ell^2}{2\Delta q_{\mu k}^t} \right.$$

$$\left. + \sum_{\ell=1}^L (x_0 + x_\ell + h_{\mu k}^t) \sum_{a=1}^n b_k^{\ell a} \right),$$

where $\Delta q_{\mu k}^t \equiv q_{1\mu k}^t - q_{2\mu k}^t > 0$ and $x^\mathsf{T} = (x_0, x_1, \dots, x_L)$. Also here we expect the logarithm of the normalization term (linked to the free energy) obtained from the well-behaved distribution P^t to be self-averaging, thus

$$\lim_{n \to \infty} \lim_{L \to \infty} \frac{1}{nL} \ln(\overline{\mathcal{Z}_{\mu k}^t}) = \lim_{n \to \infty} \lim_{L \to \infty} \frac{1}{nL} \ln[\mathcal{Z}_{\mu k}^t(h_{\mu k}^t, q_{1\mu k}^t, q_{2\mu k}^t)],$$

which is satisfied if the entries behave like $q_{2\mu k}^t \sim g_{2\mu k}^t / nL$ and $\Delta q_{\mu k}^t \sim g_{1\mu k}^t / n$, where $g_{1\mu k}^t$ and $g_{2\mu k}^t \sim O(1)$. Using this scaled parameter, the expression for the normalization is $\mathcal{Z}_{\mu k}^t = \int d\mathbf{x} \exp[-nL\Phi(\mathbf{x}; h_{\mu k}^t, g_{1\mu k}^t, g_{2\mu k}^t)]$ where

$$\Phi(\mathbf{x}; h_{\mu k}^t, g_{1\mu k}^t, g_{2\mu k}^t) \equiv \frac{x_0^2}{2g_{1\mu k}^t} + \frac{1}{L}\sum_{\ell=1}^L \frac{x_\ell^2}{2g_{2\mu k}^t}$$

$$- \frac{1}{L}\sum_{\ell=1}^L \ln[2\cosh(x_0 + x_\ell + h_{\mu k}^t)].$$

As before, we drop the indexes μ, k, and t for brevity. The critical points of the function $\Phi(\mathbf{x}; h, g_1, g_2)$ satisfy the following set of equations:

$$\frac{\partial \Phi}{\partial x_0} = \frac{x_0}{g_1} - \frac{1}{L}\sum_{\ell=1}^L \tanh(x_0 + x_\ell + h) = 0,$$

$$\frac{\partial \Phi}{\partial x_\ell} = \frac{1}{L}\left(\frac{x_\ell}{g_2} - \tanh(x_0 + x_\ell + h)\right) = 0,$$

which are satisfied for the following values:

$$x_0^* = \frac{g_1}{g_2}\frac{1}{L}\sum_{\ell=1}^{L} x_\ell^* = \frac{g_1}{g_2}x^*,$$

$$\frac{x_\ell^*}{g_2} = \tanh\left(x_\ell^* + \frac{g_1}{g_2}x^* + h\right),$$ (B1)

where $x^* \equiv \frac{1}{L}\Sigma_{\ell=1}^{L} x_\ell^*$. The second equation in the set, Eq. (B1), has the same form for all $\ell = 1, \ldots, L$ and in the small field regime it has at most three different solutions. From the three possible solutions, one is a local maximum; of the other two, the one that has the same sign as h is dominant. Thus we can expect, for all ℓ, $x_\ell^* = x^*$. This reduces the set of $L+1$ equations to one,

$$\frac{x^*}{g_2} = \tanh\left(\frac{G}{g_2}x^* + h\right),$$

where $G \equiv g_1 + g_2$. With the substitution $u = (G/g_2)x^*$ the equation has the same form as Eq. (A3), i.e., $u = G\tanh(u + h)$. If one considers again the field h to be small, the solutions can be expressed as an expansion of the zero field solutions $u_{\pm h} \simeq \pm Gm + \xi(m, G)h$, where $\xi(m, G)$ is given by Eq. (A4), and $\text{sgn}(m) = \text{sgn}(h)$. Using these expansions the critical values are given by $x_{0,\pm h}^* \simeq g_1[\pm m + G^{-1}\xi(m, G)h]$ and $x_{\ell,\pm h}^* \simeq g_2[\pm m + G^{-1}\xi(m, G)h]$ for all $\ell = 1, \ldots, L$.

As in the RS case, the expansion of Φ around the critical points in the small field regime is $\Phi(\mathbf{x}_{\pm h}^*; h \to 0, g_1, g_2) \simeq \Phi(\mathbf{x}_0^*; 0, g_1, g_2) \mp mh = \Phi_0 \mp mh$. So the dominant solution is the one that shares the sign with the field.

For a sufficiently large system with nL variables, one expects the following expansion to be valid:

$$\exp\{-nL\Phi(\mathbf{x}; h \to 0, g_1, g_2)\} \simeq e^{-nL\Phi_0}\left[e^{nLmh}\exp\left(-\frac{nL}{2}(\mathbf{x}\right.\right.$$
$$\left.- \mathbf{x}_h^*)^\mathsf{T}\mathbf{H}_{\Phi,h}(\mathbf{x} - \mathbf{x}_h^*)\right) + e^{-nLmh}\exp\left(-\frac{nL}{2}(\mathbf{x}\right.$$
$$\left.\left.- \mathbf{x}_{-h}^*)^\mathsf{T}\mathbf{H}_{\Phi,-h}(\mathbf{x} - \mathbf{x}_{-h}^*)\right)\right],$$ (B2)

where $\mathbf{H}_{\Phi,\pm h}$ is the Hessian of Φ in $\mathbf{x}_{\pm h}^*$.

Defining $\beta_{\pm h} \equiv (1 - m^2)\{1 \mp 2[\xi(m, G) + 1]mh\}$, the entries of the Hessian become

$$\frac{\partial^2 \Phi}{\partial x_0^2}\bigg|_{\mathbf{x}_{\pm h}^*} \simeq g_1^{-1} - \beta_{\pm h} \equiv \alpha_{\pm h},$$

$$\frac{\partial^2 \Phi}{\partial x_\ell^2}\bigg|_{\mathbf{x}_{\pm h}^*} \simeq \frac{1}{L}(g_2^{-1} - \beta_{\pm h}) \equiv \frac{1}{L}\gamma_{\pm h},$$

$$\frac{\partial^2 \Phi}{\partial x_0 \partial x_\ell}\bigg|_{\mathbf{x}_{\pm h}^*} \simeq -\frac{1}{L}\beta_{\pm h},$$

$$\frac{\partial^2 \Phi}{\partial x_\ell \partial x_{\ell'}}\bigg|_{\mathbf{x}_{\pm h}^*} = 0.$$

The corresponding characteristic equation is

$$\det(\mathbf{H}_{\Phi,\pm h} - \lambda\mathbb{1}) = \left(\frac{1}{L}\gamma_{\pm h} - \lambda\right)^{L-1}\left[(\alpha_{\pm h} - \lambda)\left(\frac{1}{L}\gamma_{\pm h} - \lambda\right)\right.$$
$$\left.- \frac{1}{L}\beta_{\pm h}^2\right] = 0.$$

The solutions for this equation, disregarding terms of $O(L^{-2})$ and $O(hL^{-1})$, are

$$\lambda_{0,\pm h} = \alpha_{\pm h} + \frac{1}{L}\frac{\beta_{\pm h}^2}{\alpha_{\pm h}}$$
$$\simeq \lambda_0\{1 \pm 2[\xi(m, g_1) - 1][\xi(m, G) + 1](1 - m^2)mh\},$$

$$\lambda_{1,\pm h} = \frac{1}{L}\left(\gamma_{\pm h} - \frac{\beta_{\pm h}^2}{\alpha_{\pm h}}\right)$$
$$\simeq \lambda_1\left(1 \pm 2\frac{\xi(m, g_1)^2[\xi(m, g_2) - 1]}{1 - [\xi(m, g_1) - 1][\xi(m, g_2) - 1]}\right.$$
$$\left.\times[\xi(m, G) + 1](1 - m^2)mh\right),$$

$$\lambda_{\ell,\pm h} = \frac{1}{L}\gamma_{\pm h} \simeq \lambda_\ell\{1 \pm 2[\xi(m, g_2) - 1][\xi(m, G) + 1](1$$
$$- m^2)mh\} \quad \forall \ \ell = 2, \ldots, L,$$ (B3)

where $\lambda_0 \equiv \alpha_0 + \frac{1}{L}\frac{\beta_0^2}{\alpha_0}$, $\lambda_1 \equiv \frac{1}{L}\left(\gamma_0 - \frac{\beta_0^2}{\alpha_0}\right)$ and $\lambda_\ell \equiv \frac{1}{L}\gamma_0$. The corresponding eigenvectors, up to order L^{-1}, are

$$\mathbf{u}_{0,\pm h} = \left(1, \overbrace{-\frac{1}{L}\frac{\beta_{\pm h}}{\alpha_{\pm h}}, -\frac{1}{L}\frac{\beta_{\pm h}}{\alpha_{\pm h}}, \ldots, -\frac{1}{L}\frac{\beta_{\pm h}}{\alpha_{\pm h}}}^{L \text{ times}}\right)^\mathsf{T},$$

$$\mathbf{u}_{1,\pm h} = \frac{1}{\sqrt{L}}\left(\frac{\beta_{\pm h}}{\alpha_{\pm h}}, \overbrace{1, 1, \ldots, 1}^{L \text{ times}}\right)^\mathsf{T},$$ (B4)

$$\mathbf{u}_{\ell,\pm h} = \frac{1}{\sqrt{\ell(\ell-1)}}(0, \overbrace{1, 1, \ldots, 1}^{\ell-1 \text{ times}}, -(\ell-1), \overbrace{0, 0, \ldots, 0}^{L-\ell \text{ times}})^\mathsf{T}$$
$$\forall \ \ell = 2, \ldots, L.$$

These vectors satisfy the normalization condition $\mathbf{u}_{\ell,\pm h}^{\mathsf{T}}\mathbf{u}_{\ell',\pm h}=\delta^{\ell\ell'}[1+O(L^{-1})]\forall \ell,\ell'=0,1,\ldots,L$. The linear transformation from the canonical basis to the basis of eigenvectors is then represented by a matrix with the entries

$$(\mathbf{U}_{\pm h})_{ij} \simeq (\mathbf{U}_0)_{ij} = \delta_{0i}\delta_{0j} + \frac{1}{\sqrt{j(j-1)}}\left(\sum_{k=1}^{j-1}\delta_{ki} - (1-\delta_{0j})\right.$$

$$\times(1-\delta_{1j})\delta_{ij}(j-1)\Big) + \frac{1}{\sqrt{L}}\delta_{1j}\left(\delta_{0i}\frac{\beta_0}{\alpha_0} + (1-\delta_{0i})\right)$$

$$-\frac{1}{L}\delta_{0j}(1-\delta_{0i})\frac{\beta_0}{\alpha_0}, \tag{B5}$$

ignoring terms of $O(hL^{-1/2})$. Because this transformation is a rigid rotation, the following properties are satisfied: $|\det(\mathbf{U}_\pm)|=1$ and $\mathbf{U}_{\pm h}^{\mathsf{T}}\mathbf{U}_{\pm h}=\mathbf{U}_{\pm h}\mathbf{U}_{\pm h}^{\mathsf{T}}=\mathbb{1}$.

Second-order terms in Eq. (B2) can be rewritten using the diagonal representation of the Hessian. Therefore, keeping only terms of order $O(L^{-1})$ we have that $(\mathbf{x}-\mathbf{x}_{\pm h})^{\mathsf{T}}\mathbf{H}_{\Phi,\pm h}(\mathbf{x}-\mathbf{x}_{\pm h})=(\mathbf{x}-\mathbf{x}_{\pm h})^{\mathsf{T}}\mathbf{U}_0\mathbf{U}_0^{\mathsf{T}}\mathbf{H}_{\Phi,\pm h}\mathbf{U}_0\mathbf{U}_0^{\mathsf{T}}(\mathbf{x}-\mathbf{x}_{\pm h})$ $=(\mathbf{y}-\mathbf{y}_{\pm h})^{\mathsf{T}}\mathbf{H}'_{\Phi,\pm h}(\mathbf{y}-\mathbf{y}_{\pm h})$, where $\mathbf{y}\equiv\mathbf{U}_0^{\mathsf{T}}\mathbf{x}$ and $\mathbf{H}'_{\Phi,\pm h}\equiv\mathbf{U}_0^{\mathsf{T}}\mathbf{H}_{\Phi,\pm h}\mathbf{U}_0$ is the diagonal representation of the Hessian, i.e., $(\mathbf{H}'_{\Phi,\pm h})_{ij}=\delta_{ij}\lambda_{i,\pm h}$. Using the diagonal representation in conjunction with Eq. (B2) one obtains an expression for the normalization term,

$$\mathcal{Z}(h\to 0,g_1,g_2) \simeq e^{-nL(\Phi_0-mh)}\int d\mathbf{y}\exp\left(-\frac{nL}{2}(\mathbf{y}-\mathbf{y}_h^*)^{\mathsf{T}}\mathbf{H}'_{\Phi,h}(\mathbf{y}-\mathbf{y}_h^*)\right) + e^{-nL(\Phi_0+mh)}$$

$$\times\int d\mathbf{y}\exp\left(-\frac{nL}{2}(\mathbf{y}-\mathbf{y}_{-h}^*)^{\mathsf{T}}\mathbf{H}'_{\Phi,-h}(\mathbf{y}-\mathbf{y}_{-h}^*)\right) \simeq e^{-nL\Phi_0}\left(\frac{2\pi}{nL}\right)^{(L+1)/2}\left(e^{nLmh}\prod_{\ell=0}^{L}\lambda_{\ell,h}^{-1/2} + e^{-nLmh}\prod_{\ell=0}^{L}\lambda_{\ell,-h}^{-1/2}\right).$$

For a small field, the product of the eigenvalues can be approximated by

$$\prod_{\ell=0}^{L}\lambda_{\ell,\pm h}^{-1/2} \simeq \{1\mp[\xi(m,g_2)-1][\xi(m,G)+1]Lmh\}\prod_{\ell=0}^{L}\lambda_\ell^{-1/2}.$$

Thus, the expression for $\mathcal{Z}$ reduces to

$$\mathcal{Z}(h\to 0,g_1,g_2) \simeq e^{-nL\Phi_0}\left(\frac{2\pi}{nL}\right)^{(L+1)/2}\prod_{\ell=0}^{L}\lambda_\ell^{-1/2}(e^{nLmh}\{1-[\xi(m,g_2)-1][\xi(m,G)+1]Lmh\}$$

$$+e^{-nLmh}\{1+[\xi(m,g_2)-1][\xi(m,G)+1]Lmh\}).$$

The mean value of a given function $f(\mathbf{x})$ is then given by

$$\langle f(\mathbf{x})|h\to 0,g_1,g_2\rangle \simeq \mathcal{Z}^{-1}e^{-nL(\Phi_0-mh)}\int d\mathbf{y}\exp\left(-\frac{nL}{2}(\mathbf{y}-\mathbf{y}_h^*)^{\mathsf{T}}\mathbf{H}'_{\Phi,h}(\mathbf{y}-\mathbf{y}_h^*)\right)\left(f(\mathbf{x}_h)+\frac{1}{2}(\mathbf{y}-\mathbf{y}_h^*)^{\mathsf{T}}\mathbf{H}'_{f,h}(\mathbf{y}-\mathbf{y}_h^*)\right)$$

$$+\mathcal{Z}^{-1}e^{-nL(\Phi_0+mh)}\int d\mathbf{y}\exp\left(-\frac{nL}{2}(\mathbf{y}-\mathbf{y}_{-h}^*)^{\mathsf{T}}\mathbf{H}'_{\Phi,-h}(\mathbf{y}-\mathbf{y}_{-h}^*)\right)\left(f(\mathbf{x}_{-h})+\frac{1}{2}(\mathbf{y}-\mathbf{y}_{-h}^*)^{\mathsf{T}}\mathbf{H}'_{f,-h}(\mathbf{y}-\mathbf{y}_{-h}^*)\right),$$

where $\mathbf{H}'_{f,\pm h}$ is the Hessian of the function f in the basis of eigenvectors of $\mathbf{H}_{\Phi\pm h}$, evaluated at the critical points. The linear terms in the expansion of $f(\mathbf{x})$ do not contribute to the expectation value. The Gaussian integral of the cross products of the type $(y_i-y_{i,\pm h}^*)(y_j-y_{j,\pm h}^*)$ with $i\neq j$ are zero, thus the Gaussian integral of the second term in the expansion of $f(\mathbf{x})$ becomes

$$I_\pm = \frac{1}{2}\mathcal{Z}^{-1}e^{-nL(\Phi_0\mp mh)}\int d\mathbf{y}\exp\left(-\frac{nL}{2}(\mathbf{y}-\mathbf{y}_{\pm h}^*)^{\mathsf{T}}\mathbf{H}'_{\Phi,\pm h}(\mathbf{y}\right.$$

$$\left.-\mathbf{y}_{\pm h}^*\right)\Big)(\mathbf{y}-\mathbf{y}_{\pm h}^*)^{\mathsf{T}}\mathbf{H}'_{f,\pm h}(\mathbf{y}-\mathbf{y}_{\pm h}^*),$$

$$I_+ \simeq \frac{1}{2}\{1-e^{-2nLmh}\{1+2[\xi(m,g_2)-1][\xi(m,G)+1]Lmh\}\}$$

$$\times\frac{1}{nL}\sum_{\ell=0}^{L}\lambda_{\ell,h}^{-1}(\mathbf{H}'_{f,h})_{\ell\ell},$$

$$I_- \simeq \frac{1}{2}e^{-2nLmh}\{1+2[\xi(m,g_2)-1][\xi(m,G)+1]Lmh\}$$

$$\times\frac{1}{nL}\sum_{\ell=0}^{L}\lambda_{\ell,-h}^{-1}(\mathbf{H}'_{f,-h})_{\ell\ell}. \tag{B6}$$

Using the expansion $f(\mathbf{x}_\pm) \simeq f[\pm\mathbf{x}_0 + h\boldsymbol{\xi}(m,G)]$ $\simeq f(\pm\mathbf{x}_0) + h\boldsymbol{\xi}^\mathsf{T}(m,G)\nabla f(\pm\mathbf{x}_0) = f(\pm\mathbf{x}_0) + \delta f(\pm\mathbf{x}_0)h$ where

$$\mathbf{x}_0^\mathsf{T} = m(g_1,\overbrace{g_2,g_2,\ldots,g_2}^{L \text{ times}}) \qquad \text{and} \qquad \boldsymbol{\xi}^\mathsf{T}(m,G) = G^{-1}\xi(m,G)$$

$$\times(g_1,\overbrace{g_2,g_2,\ldots,g_2}^{L \text{ times}}),$$ the diagonal entries of the transformed Hessian are

$$(\mathbf{H}'_{f,\pm h})_{\ell\ell} = \sum_{i,j=0}^{L} (\mathbf{U}_{\pm h})_{\ell i}(\mathbf{U}_{\pm h})_{\ell j}(\mathbf{H}_{f,\pm h})_{ij}$$

$$= \sum_{i,j=0}^{L} (\mathbf{U}_{\pm h})_{\ell i}(\mathbf{U}_{\pm h})_{\ell j} \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{x}_\pm}$$

$$\simeq \sum_{i,j=0}^{L} (\mathbf{U}_0)_{\ell i}(\mathbf{U}_0)_{\ell j}\left(\left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\pm\mathbf{x}_0}\right.$$

$$\left. + h\ \boldsymbol{\xi}^\mathsf{T}(m,G)\nabla \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\pm\mathbf{x}_0}\right)$$

$$\simeq (\left.\mathbf{H}'_f\right|_{\pm\mathbf{x}_0})_{\ell\ell} + (\left.\delta\mathbf{H}'_f\right|_{\pm\mathbf{x}_0})_{\ell\ell}h, \qquad (B7)$$

with $(\left.\delta\mathbf{H}'_f\right|_{\pm\mathbf{x}_0})_{\ell\ell}$ defined by the second term in (B7). Using the entries of the diagonalized Hessian, the last term in the integrals (B6) becomes

$$\frac{1}{nL}\sum_{\ell=0}^{L} \lambda_{\ell,-h}^{-1}(\mathbf{H}'_{f,\pm h})_{\ell\ell} \simeq \frac{1}{nL}\frac{1}{\alpha_0}(\left.\mathbf{H}'_f\right|_{\mathbf{x}_0})_{00}$$

$$+ \frac{1}{n}\frac{\alpha_0}{(\alpha_0\gamma_0 - \beta_0^2)}(\left.\mathbf{H}'_f\right|_{\mathbf{x}_0})_{11}$$

$$+ \frac{1}{n}\frac{1}{\gamma_0}\sum_{\ell=2}^{L}(\left.\mathbf{H}'_f\right|_{\mathbf{x}_0})_{\ell\ell},$$

disregarding terms of $O\left(\frac{h}{n} + \frac{h}{L}\right)$. The expectation value of an arbitrary function f can then be approximated by

$$\langle f(\mathbf{x})|h \to 0^+, g_1, g_2\rangle \simeq f(\mathbf{x}_0) + \frac{1}{2}\frac{1}{nL}\sum_{\ell=0}^{L}\lambda_\ell^{-1}(\left.\mathbf{H}'_f\right|_{\mathbf{x}_0})_{\ell\ell}$$

$$- e^{-2nLmh}[f(\mathbf{x}_0) - f(-\mathbf{x}_0)] + \delta f(\mathbf{x}_0)h, \qquad (B8)$$

where we have disregarded terms of $O\left(\frac{h}{n} + \frac{h}{L}\right), O\left(\frac{1}{n}e^{-2nLmh}\right)$, and $O(Lhe^{-2nLmh})$. By simple inspection, Eq. (B8) is equivalent to the RS mean value equation (A7).

The single variable mean value is then

$$\langle b_k^{\ell a}|h_{\mu k}^t \to 0, g_{1\mu k}^t, g_{2\mu k}^t\rangle = \sum_{\{\mathbf{b}_k\}} P^t(\mathbf{b}_k|\{y_{\nu\neq\mu}\})b_k^{\ell a}$$

$$= \langle\tanh(x_0 + x_\ell + h_{\mu k}^t)|h_{\mu k}^t \to 0, g_{1\mu k}^t, g_{2\mu k}^t\rangle.$$

The expansion for $f(\mathbf{x}) = \tanh(x_0 + x_\ell + h_{\mu k}^t)$ is

$$f(\mathbf{x}) \simeq m_{\mu k}^t + [1 - (m_{\mu k}^t)^2]$$

$$\times(1,\overbrace{0,0,\ldots,0}^{\ell-1 \text{ times}},1,\overbrace{0,0,\ldots,0}^{L-\ell \text{ times}})^\mathsf{T}\boldsymbol{\xi}(m_{\mu k}^t, G_{\mu k}^t)h_{\mu k}^t,$$

which results in the following expression for the single variable mean value:

$$\langle b_k^{\ell a}|h_{\mu k}^t \to 0, g_{1\mu k}^t, g_{2\mu k}^t\rangle \simeq (1 - 2e^{-2nLm_{\mu k}^t h_{\mu k}^t})m_{\mu k}^t$$

$$+ \xi(m_{\mu k}^t, G_{\mu k}^t)[1 - (m_{\mu k}^t)^2]h_{\mu k}^t - m_{\mu k}^t[1$$

$$- (m_{\mu k}^t)^2]\frac{1}{nL}\sum_{k=0}^{L}\lambda_k^{-1}(\mathbf{M}'_{0\ell})_{kk},$$

where $(\mathbf{M}_{0\ell})_{ij} = \delta_{0i}\delta_{0j} + \delta_{0i}\delta_{\ell j} + \delta_{\ell i}\delta_{0j} + \delta_{\ell i}\delta_{\ell j}$ is a matrix such that $\mathbf{H}_{\tanh(x_0+x_\ell)}|_{\mathbf{x}_0} = -2m_{\mu k}^t[1 - (m_{\mu k}^t)^2]\mathbf{M}_{0\ell}$. In the basis of the $\mathbf{H}_\Phi$ eigenvalues, the expressions for the diagonal elements of this matrix are

$$(\mathbf{M}'_{0\ell})_{kk} = \sum_{i,j=0}^{L} (\mathbf{U}_{\pm h})_{ik}(\mathbf{U}_{\pm h})_{jk}(\delta_{0i}\delta_{0j} + \delta_{0i}\delta_{\ell j} + \delta_{\ell i}\delta_{0j} + \delta_{\ell i}\delta_{\ell j})$$

$$= [(\mathbf{U}_{\pm h})_{0k} + (\mathbf{U}_{\pm h})_{\ell k}]^2,$$

$$(\mathbf{M}'_{0\ell})_{00} \simeq 1 - \frac{2}{L}\frac{\beta_0}{\alpha_0},$$

$$(\mathbf{M}'_{0\ell})_{11} \simeq \frac{1}{L}\left(\frac{\alpha_0 + \beta_0}{\alpha_0}\right)^2,$$

$$(\mathbf{M}'_{0\ell})_{kk} = \delta_{k\ell}\frac{\ell - 1}{\ell} + \Theta(k - \ell - 1)\frac{1}{k(k-1)} \quad \forall\ \ell = 2,\ldots,L,$$

$$(B9)$$

where $\Theta(n) = 1$ if $n > 0$ and 0 otherwise. The sum of the eigenvalues' inverse times the diagonal elements equation (B9) results in

$$\frac{1}{nL}\sum_{k=0}^{L}\lambda_k^{-1}(\mathbf{M}'_{0\ell})_{kk}$$

$$\simeq \frac{1}{n}\frac{1}{\gamma_0}\sum_{k=2}^{L}\left(\delta_{k\ell}\frac{\ell - 1}{\ell} + \Theta(k - \ell - 1)\frac{1}{k(k-1)}\right)$$

$$+ \frac{1}{nL}\frac{1}{\alpha_0}\left(1 + \frac{(\alpha_0 + \beta_0)^2}{\alpha_0\gamma_0 - \beta_0^2}\right)$$

$$= \frac{1}{n}\frac{1}{\gamma_0}\left(\sum_{k=2}^{L}\frac{1}{k(k-1)}\right) + \frac{1}{nL}\frac{1}{\alpha_0}\left(1 + \frac{(\alpha_0 + \beta_0)^2}{\alpha_0\gamma - \beta_0^2}\right)$$

$$= \frac{1}{n}\frac{1}{\gamma_0} + \frac{1}{nL}\frac{1}{\alpha_0}\left(1 + \frac{(\alpha_0 + \beta_0)^2}{\alpha_0\gamma_0 - \beta_0^2} - \frac{\alpha_0}{\gamma_0}\right)$$

$$= \frac{1}{n}g_{2\mu k}^t\xi(m_{\mu k}^t, g_{2\mu k}^t) + \frac{1}{nL}[G_{\mu k}^t\xi(m_{\mu k}^t, G_{\mu k}^t)$$

$$- g_{2\mu k}^t\xi(m_{\mu k}^t, g_{2\mu k}^t)],$$

where we have used that $\sum_{k=2}^{L}[k(k-1)]^{-1} = (L-1)/L$, $\gamma_0^{-1} = g_{2\mu k}^t\xi(m_{\mu k}^t, g_{2\mu k}^t)$ and $\frac{1}{\alpha_0}\left(1 + \frac{(\alpha_0+\beta_0)^2}{\alpha_0\gamma_0 - \beta_0^2} - \frac{\alpha_0}{\gamma_0}\right) = G_{\mu k}^t\xi(m_{\mu k}^t, G_{\mu k}^t)$ $- g_{2\mu k}^t\xi(m_{\mu k}^t, g_{2\mu k}^t)$. The final expression for the expectation value of a single variable is

$$\langle b_k^{\ell a}|h_{\mu k}^t \to 0, g_{1\mu k}^t, g_{2\mu k}^t\rangle \simeq (1 - 2e^{-2nm_{\mu k}^t h_{\mu k}^t})m_{\mu k}^t - \frac{g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{n}[1 - (m_{\mu k}^t)^2]m_{\mu k}^t - \frac{G_{\mu k}^t \xi(m_{\mu k}^t, G_{\mu k}^t) - g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{nL}$$

$$\times [1 - (m_{\mu k}^t)^2]m_{\mu k}^t + \xi(m_{\mu k}^t, G_{\mu k}^t)[1 - (m_{\mu k}^t)^2]h_{\mu k}^t. \tag{B10}$$

To calculate $\langle b_k^{\ell a} b_k^{\ell a'}|h_{\mu k}^t \to 0, g_{\mu k}^t, \Delta g_{\mu k}^t\rangle$, an off-diagonal element $(a \neq a')$ in the same block ℓ, we can apply Eq. (B8) with $f(\mathbf{x}) = \tanh^2(x_0 + x_\ell + h_{\mu k}^t)$, thus the Hessian matrix is $\mathbf{H}_{\tanh^2(x_0 + x_\ell)}|_{\mathbf{x}_0} = 2[1 - (m_{\mu k}^t)^2][1 - 3(m_{\mu k}^t)^2]\mathbf{M}_{0\ell}$, thus

$$\langle b_k^{\ell a} b_k^{\ell a'}|h_{\mu k}^t \to 0, g_{1\mu k}^t, g_{2\mu k}^t\rangle \simeq (m_{\mu k}^t)^2 + \frac{g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{n}[1 - (m_{\mu k}^t)^2][1 - 3(m_{\mu k}^t)^2] + \frac{G_{\mu k}^t \xi(m_{\mu k}^t, G_{\mu k}^t) - g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{nL}$$

$$\times [1 - (m_{\mu k}^t)^2][1 - 3(m_{\mu k}^t)^2] + 2\xi(m_{\mu k}^t, G_{\mu k}^t)[1 - (m_{\mu k}^t)^2]m_{\mu k}^t h_{\mu k}^t. \tag{B11}$$

Finally, to calculate the expectation value for the product of two variables belonging to different blocks $\ell \neq \ell'$ (the sub-block index a is insignificant in this case), $\langle b_k^{\ell a} b_k^{\ell' a}|h_{\mu k}^t \to 0, g_{1\mu k}^t, g_{2\mu k}^t\rangle$. We set $f(\mathbf{x}) = \tanh(x_0 + x_\ell + h_{\mu k}^t)\tanh(x_0 + x_{\ell'} + h_{\mu k}^t)$, thus the Hessian matrix

$$(\mathbf{H}_{\tanh(x_0 + x_\ell)\tanh(x_0 + x_{\ell'})}|_{\mathbf{x}_0})_{ij} = \mathcal{M}_0(m_{\mu k}^t)(2\delta_{i0}\delta_{j0} + \delta_{i0}\delta_{j\ell} + \delta_{i\ell}\delta_{j0} + \delta_{i0}\delta_{j\ell'} + \delta_{i\ell'}\delta_{j0}) + \mathcal{M}_1(m_{\mu k}^t)(\delta_{i\ell}\delta_{j\ell'} + \delta_{i\ell'}\delta_{j\ell})$$

$$- 2\mathcal{M}_2(m_{\mu k}^t)(\delta_{i\ell}\delta_{j\ell} + \delta_{i\ell'}\delta_{j\ell'}),$$

where $\mathcal{M}_0(m_{\mu k}^t) \equiv [1 - (m_{\mu k}^t)^2][1 - 3(m_{\mu k}^t)^2]$, $\mathcal{M}_1(m_{\mu k}^t) \equiv [1 - (m_{\mu k}^t)^2]^2$, and $\mathcal{M}_2(m_{\mu k}^t) \equiv (m_{\mu k}^t)^2[1 - (m_{\mu k}^t)^2]$. The diagonal elements $\mathcal{K}_{\ell\ell';k} \equiv (\mathbf{H}'_{\tanh(x_0 + x_\ell)\tanh(x_0 + x_{\ell'})}|_{\mathbf{x}_0})_{kk}$ in the basis of eigenvectors of $\mathbf{H}_\Phi$ are

$$\mathcal{K}_{\ell\ell';0} \simeq 2\mathcal{M}_0(m_{\mu k}^t),$$

$$\mathcal{K}_{\ell\ell';1} \simeq \frac{2\mathcal{M}_0(m_{\mu k}^t)}{L}\left[\frac{\beta_0}{\alpha_0}\left(\frac{\beta_0 + 2\alpha_0}{\alpha_0}\right) + 1\right],$$

$$\mathcal{K}_{\ell\ell';j} = -2\delta_{j\ell}\mathcal{M}_2(m_{\mu k}^t)\frac{\ell - 1}{\ell} - 2[\Theta(j - \ell)\Theta(\ell' - j)]\frac{\mathcal{M}_2(m_{\mu k}^t)}{j(j - 1)}$$

$$- 2\delta_{j\ell'}\left[\frac{\mathcal{M}_1(m_{\mu k}^t)}{\ell'} + \frac{\mathcal{M}_2(m_{\mu k}^t)}{\ell'}\left(\ell' - 1 + \frac{1}{\ell' - 1}\right)\right] + 2\Theta(j - \ell')\frac{\mathcal{M}_0(m_{\mu k}^t)}{j(j - 1)},$$

thus, the sum of the diagonal elements is

$$\frac{1}{2}\frac{1}{nL}\sum_{k=0}^{L}\lambda_{k,\pm h}^{-1}\mathcal{K}_{\ell\ell';k} \simeq \frac{1}{nL}\frac{\mathcal{M}_0(m_{\mu k}^t)}{\alpha_0}\left(1 + \frac{(\beta_0 + \alpha_0)^2}{(\alpha_0\gamma_0 - \beta_0^2)}\right) - \frac{1}{n}\frac{1}{\gamma_0}\left[\mathcal{M}_2(m_{\mu k}^t)\left(\frac{\ell - 1}{\ell} + \sum_{j=\ell+1}^{\ell'-1}\frac{1}{j(j - 1)}\right)\right.$$

$$\left. + \frac{\mathcal{M}_1(m_{\mu k}^t)}{\ell'} + \frac{\mathcal{M}_2(m_{\mu k}^t)}{\ell'}\left(\ell' - 1 + \frac{1}{\ell' - 1}\right) - \mathcal{M}_0(m_{\mu k}^t)\sum_{j=\ell'+1}^{L}\frac{1}{j(j - 1)}\right]$$

$$= -\frac{g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{n}[\mathcal{M}_1(m_{\mu k}^t) - \mathcal{M}_0(m_{\mu k}^t)] + \frac{G_{\mu k}^t \xi(m_{\mu k}^t, G_{\mu k}^t) - g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{nL}\mathcal{M}_0(m_{\mu k}^t).$$

Using the sum of diagonal terms one then derives the expected correlation for variables belonging to two different blocks

$$\langle b_k^{\ell a} b_k^{\ell' a}|h_{\mu k}^t \to 0, g_{\mu k}^t, \Delta g_{\mu k}^t\rangle \simeq (m_{\mu k}^t)^2 - 2\frac{g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{n}(m_{\mu k}^t)^2[1 - (m_{\mu k}^t)^2] + \frac{G_{\mu k}^t \xi(m_{\mu k}^t, G_{\mu k}^t) - g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{nL}$$

$$\times [1 - 3(m_{\mu k}^t)^2][1 - (m_{\mu k}^t)^2] + 2\xi(m_{\mu k}^t, G_{\mu k}^t)[1 - (m_{\mu k}^t)^2]m_{\mu k}^t h_{\mu k}^t. \tag{B12}$$

Keeping in mind that $\langle b_k^{\ell a} b_k^{\ell a}|h_{\mu k}^t, g_{1\mu k}^t, g_{2\mu k}^t\rangle = 1$ and using Eqs. (B10)–(B12), the covariance matrix entries can be then calculated as

$$(\Psi_{\mu kl}^t)^{\ell a\ell'a'} = \langle b_k^{\ell a} b_l^{\ell'a'}|h_{\mu k}^t \to 0, g_{1\mu k}^t, g_{2\mu k}^t; h_{\mu l}^t \to 0, g_{1\mu l}^t, g_{2\mu l}^t\rangle - \langle b_k^{\ell a}|h_{\mu k}^t \to 0, g_{1\mu k}^t, g_{2\mu k}^t\rangle\langle b_l^{\ell'a'}|h_{\mu l}^t \to 0, g_{1\mu l}^t, g_{2\mu l}^t\rangle$$

$$= \delta_{kl}(\Psi_{\mu kk}^t)^{\ell a\ell'a'}$$

$$(\Psi_{\mu k k}^t)^{\ell a \ell' a'} \simeq \delta^{\ell \ell'} \delta^{a a'} [1 - (m_{\mu k}^t)^2] + \delta^{\ell \ell'} (1 - \delta^{a a'}) \frac{g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{n} [1 - (m_{\mu k}^t)^2]^2$$

$$+ (1 - \delta^{\ell \ell'}) \frac{G_{\mu k}^t \xi(m_{\mu k}^t, G_{\mu k}^t) - g_{2\mu k}^t \xi(m_{\mu k}^t, g_{2\mu k}^t)}{nL} [1 - (m_{\mu k}^t)^2]^2,$$

where we have kept only the dominant terms at each entry, disregarding terms of order $O\left(e^{-2nm_{\mu k}^t h_{\mu k}^t + \frac{h}{n} + \frac{h}{L}}\right)$.

If the $\varepsilon_{\mu k}$ and b_k^a are unbiased variables, the variable $\Delta_{\mu k}^a = \Sigma_{l \neq k} \varepsilon_{\mu l} b_l^a$, by virtue of the central limit theorem, obeys a normal distribution, with mean value and covariance matrix that can be obtained by employing the expressions derived for Ψ,

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

$$= \sum_{l \neq k} \varepsilon_{\mu l} m_{\mu l}^t,$$

$$(\mathbf{Y}_{\mu k}^t)^{\ell a \ell' a'} \equiv \langle \Delta_{\mu k}^{\ell a} \Delta_{\mu k}^{\ell' a'} \rangle - \langle \Delta_{\mu k}^{\ell a} \rangle \langle \Delta_{\mu k}^{\ell' a'} \rangle$$

$$= \sum_{\{\mathbf{b}_{l \neq k}\}} \prod_{l \neq k} P^t(\mathbf{b}_l | \{y_{\nu \neq \mu}\}) \sum_{\substack{l \neq k \\ j \neq k}} \varepsilon_{\mu l} \varepsilon_{\mu j} b_l^{\ell a} b_j^{\ell' a'}$$

$$- \left(\sum_{l \neq k} \varepsilon_{\mu l} m_{\mu l}^t \right)^2 = \sum_{l \neq k} \varepsilon_{\mu l}^2 (\Psi_{\mu l j}^t)^{\ell a \ell' a'}$$

$$= \delta^{\ell \ell'} \delta^{a a'} X_{\mu k}^t + \delta^{\ell \ell'} (1 - \delta^{a a'}) \frac{1}{n} R_{\mu k}^t + (1$$

$$- \delta^{\ell \ell'}) \frac{1}{nL} (V_{\mu k}^t - R_{\mu k}^t), \quad (B13)$$

where $X_{\mu k}^t$ is given by Eq. (A9) and

$$R_{\mu k}^t \equiv \sum_{l \neq k} \varepsilon_{\mu l}^2 g_{2\mu l}^t \xi(m_{\mu l}^t, g_{2\mu l}^t) [1 - (m_{\mu l}^t)^2]^2,$$

$$V_{\mu k}^t \equiv \sum_{l \neq k} \varepsilon_{\mu l}^2 G_{\mu l}^t \xi(m_{\mu l}^t, G_{\mu l}^t) [1 - (m_{\mu l}^t)^2]^2,$$

are macroscopic variables of $O(1)$. In particular, $R_{\mu k}^t$ and $V_{\mu k}^t$ are free variables that can be used to optimize a given performance measure.

APPENDIX C: THE MESSAGES

From the conditional probabilities of equations (3) and (4) and with the application of the probability distributions $P(\Delta_{\mu k} | \mathbf{B})$ of Eq. (8) in (5) we can express the message from nodes y_μ to nodes b_k^a at time $t+1$ as

$$\hat{m}_{\mu k}^{t+1} = \frac{\sum_{\{\mathbf{B}\}} b_k^{a'} \prod_{a=1}^n P(y_\mu | \mathbf{b}^a) P(\mathbf{b}^a) \prod_{l \neq k} P(\mathbf{b}_l | \{y_{\nu \neq \mu}\})}{\sum_{\{\mathbf{B}\}} \prod_{a=1}^n P(y_\mu | \mathbf{b}^a) P(\mathbf{b}^a) \prod_{l \neq k} P(\mathbf{b}_l | \{y_{\nu \neq \mu}\})}$$

$$= \frac{\int d\Delta_{\mu k} P(\Delta_{\mu k} | \mathbf{B}) \sum_{\{\mathbf{b}_k\}} b_k^{a'} P(y_\mu | \Delta_{\mu k}; \gamma) [1 + \varepsilon_{\mu k} \mathbf{b}_k^\mathsf{T} \nabla_{\Delta_{\mu k}} \ln P(y_\mu | \Delta_{\mu k}; \gamma)]}{\int d\Delta_{\mu k} P(\Delta_{\mu k} | \mathbf{B}) \sum_{\{\mathbf{b}_k\}} P(y_\mu | \Delta_{\mu k}; \gamma) [1 + \varepsilon_{\mu k} \mathbf{b}_k^\mathsf{T} \nabla_{\Delta_{\mu k}} \ln P(y_\mu | \Delta_{\mu k}; \gamma)]}. \quad (C1)$$

If $P(y_\mu | \Delta_{\mu k}; \gamma) = \prod_{a=1}^n P(y_\mu | \Delta_{\mu k}^a; \gamma)$, and ignoring $O(\varepsilon_{\mu k}^2)$ terms, the traces on $\mathbf{b}_k$ can be written as

$$\sum_{\{\mathbf{b}_k\}} P(y_\mu | \Delta_{\mu k}; \gamma) [1 + \varepsilon_{\mu k} \mathbf{b}_k^\mathsf{T} \nabla_{\Delta_{\mu k}} \ln P(y_\mu | \Delta_{\mu k}; \gamma)] = 2^n P(y_\mu | \Delta_{\mu k}; \gamma)$$

$$\sum_{\{\mathbf{b}_k\}} b_k^{a'} P(y_\mu | \Delta_{\mu k}; \gamma) [1 + \varepsilon_{\mu k} \mathbf{b}_k^\mathsf{T} \nabla_{\Delta_{\mu k}} \ln P(y_\mu | \Delta_{\mu k}; \gamma)] = 2^n \varepsilon_{\mu k} P(y_\mu | \Delta_{\mu k}; \gamma) \frac{\partial}{\partial \Delta_{\mu k}^{a'}} \ln P(y_\mu | \Delta_{\mu k}^{\tilde{a}}; \gamma),$$

thus, following from (C1) and neglecting $O(1/n)$ terms

$$
{}^{(RS)}\hat{m}_{\mu k}^{t+1} \simeq \varepsilon_{\mu k} \frac{\int d\mathbf{\Delta}_{\mu k} P(\mathbf{\Delta}_{\mu k}|\mathbf{B}) \prod\limits_{a=1}^{n} P(y_\mu|\Delta_{\mu k}^a;\gamma) \frac{\partial}{\partial \Delta_{\mu k}^{a'}} \ln P(y_\mu|\Delta_{\mu k}^{a'};\gamma)}{\int d\mathbf{\Delta}_{\mu k} P(\mathbf{\Delta}_{\mu k}|\mathbf{B}) \prod\limits_{a=1}^{n} P(y_\mu|\Delta_{\mu k}^a;\gamma)}
$$

$$
= \frac{\varepsilon_{\mu k}}{{}^{(RS)}\mathcal{N}_{\mu k}^t} \int d\vartheta \exp\left(-n\frac{(\vartheta - u_{\mu k}^t)^2}{2R^t}\right) \left[\int d\Delta \exp\left(-\frac{(\Delta - \vartheta)^2}{2X^t} + \ln P(y_\mu|\Delta;\gamma)\right)\right]^{n-1}
$$

$$
\times \int d\Delta \exp\left(-\frac{(\Delta - \vartheta)^2}{2X^t}\right) \frac{\partial}{\partial \Delta} P(y_\mu|\Delta;\gamma) \tag{C2}
$$

and

$$
{}^{(1RSB)}\hat{m}_{\mu k}^{t+1} \simeq \varepsilon_{\mu k} \frac{\int d\mathbf{\Delta}_{\mu k} P(\mathbf{\Delta}_{\mu k}|\mathbf{B}) \prod\limits_{\ell=1}^{L} \prod\limits_{a=1}^{n} P(y_\mu|\Delta_{\mu k}^{\ell a};\gamma) \frac{\partial}{\partial \Delta_{\mu k}^{\ell' a'}} \ln P(y_\mu|\Delta_{\mu k}^{\ell' a'};\gamma)}{\int d\mathbf{\Delta}_{\mu k} P(\mathbf{\Delta}_{\mu k}|\mathbf{B}) \prod\limits_{a=1}^{n} P(y_\mu|\Delta_{\mu k}^a;\gamma)}
$$

$$
= \frac{\varepsilon_{\mu k}}{{}^{(1RSB)}\mathcal{N}_{\mu k}^t} \int d\Theta \prod\limits_{\ell=1}^{L} \exp\left[-\frac{n}{2}\left(\frac{(\vartheta^0)^2}{V^t - R^t} + \frac{(\vartheta^\ell)^2}{V^t - L^{-1}(V^t - R^t)}\right)\right] \prod\limits_{\ell \neq \ell'} \left[\int d\Delta \exp\left(-\frac{(\Delta - \vartheta_{\mu k}^{0\ell t})^2}{2X^t} + \ln P(y_\mu|\Delta;\gamma)\right)\right]^n
$$

$$
\times \left[\int d\Delta \exp\left(-\frac{(\Delta - \vartheta_{\mu k}^{0\ell' t})^2}{2X^t} + \ln P(y_\mu|\Delta;\gamma)\right)\right]^{n-1} \int d\Delta \exp\left(-\frac{(\Delta - \vartheta_{\mu k}^{0\ell' t})^2}{2X^t}\right) \frac{\partial}{\partial \Delta} P(y_\mu|\Delta;\gamma),
$$

where ${}^{(RS)}\mathcal{N}_{\mu k}^t$ and ${}^{(1RSB)}\mathcal{N}_{\mu k}^t$ are suitable normalization constants and $\vartheta_{\mu k}^{0\ell t} \equiv \vartheta^0 + \vartheta^\ell + u_{\mu k}^t$. One can then define

$$
\mathcal{G}(y_\mu, \vartheta) \equiv \int d\Delta \exp\left(-\frac{(\Delta - \vartheta)^2}{2X^t}\right) P(y_\mu|\Delta;\gamma), \tag{C3}
$$

$$
\mathcal{P}(y_\mu, \vartheta) \equiv [\mathcal{G}(y_\mu, \vartheta)]^{-1} \int d\Delta \exp\left(-\frac{(\Delta - \vartheta)^2}{2X^t}\right) \frac{\partial}{\partial \Delta} P(y_\mu|\Delta;\gamma)
$$

$$
= [\mathcal{G}(y_\mu, \vartheta)]^{-1} \int d\Delta \exp\left(-\frac{(\Delta - \vartheta)^2}{2X^t}\right) \frac{\Delta - \vartheta}{X^t} P(y_\mu|\Delta;\gamma) = \frac{\partial}{\partial \vartheta} \ln \mathcal{G}(y_\mu, \vartheta), \tag{C4}
$$

$$
{}^{(RS)}\mathcal{H}(\vartheta, y_\mu) \equiv \frac{(\vartheta - u_{\mu k}^t)^2}{2R_{\mu k}^t} - \ln \mathcal{G}(y_\mu, \vartheta), \tag{C5}
$$

$$
{}^{(1RSB)}\mathcal{H}(\vartheta^0, \vartheta^\ell, y_\mu) \equiv \frac{1}{2}\left(\frac{(\vartheta^0)^2}{V^t - R^t} + \frac{(\vartheta^\ell)^2}{V^t - L^{-1}(V^t - R^t)}\right)
$$

$$
- \ln \mathcal{G}(y_\mu, \vartheta_{\mu k}^{0\ell t}). \tag{C6}
$$

Thus the expression for the RS message is

$$
{}^{(RS)}\hat{m}_{\mu k}^{t+1} = \varepsilon_{\mu k} \frac{\int d\vartheta \exp[-n^{(RS)}\mathcal{H}(\vartheta, y_\mu)] \mathcal{P}(y_\mu, \vartheta)}{\int d\vartheta \exp[-n^{(RS)}\mathcal{H}(\vartheta, y_\mu)]}.
$$

In the large n limit, only the solutions $\tilde{\vartheta}_{\mu k}^t$ of $\frac{\partial}{\partial \vartheta}{}^{(RS)}\mathcal{H} = 0$, that correspond to the minimum of $\mathcal{H}$ contribute to the integral. The dominant term in the integral is obtained via saddle point methods, which leads to the final expression for the message

$$
{}^{(RS)}\hat{m}_{\mu k}^{t+1} = \varepsilon_{\mu k} \frac{\tilde{\vartheta}_{\mu k}^t - u_{\mu k}^t}{R^t}, \tag{C7}
$$

where $\tilde{\vartheta}_{\mu k}^t$ is given by Eq. (D1).

The 1RSB case is a little more delicate. The exponential is a sum over L functions ${}^{(1RSB)}\mathcal{H}(\vartheta^0, \vartheta^\ell, y_\mu)$. Therefore, a Taylor expansion close to the saddle point of Eq. (D2) is employed resulting in

$$\sum_{\ell=1}^{L} {}^{(1\text{RSB})}\mathcal{H}(\vartheta^0, \vartheta^\ell, y_\mu) \simeq LE_0 + \frac{L}{2} h_0 (\Delta\vartheta^0)^2 + h_1 \Delta\vartheta^0 \sum_{\ell=1}^{L} \Delta\vartheta^\ell + \frac{1}{2} h_2 \sum_{\ell=1}^{L} (\Delta\vartheta^\ell)^2 + O(\Delta\vartheta^3),$$

where $E_0 = {}^{(1\text{RSB})}\mathcal{H}(\tilde{\vartheta}_{\mu k}^{0t}, \tilde{\vartheta}_{\mu k}^{\ell t}, y_\mu)$ is the energy of the ground state, $\Delta\vartheta^i = \vartheta^i - \tilde{\vartheta}_{\mu k}^{it}$ $i = 0, \ell$ and the entries h_0, h_1, and h_2 satisfy the equation

$$\begin{pmatrix} h_0 & h_1 \\ h_1 & h_2 \end{pmatrix} = \begin{pmatrix} (V^t - R^t)^{-1} & 0 \\ 0 & (V^t)^{-1} \end{pmatrix} - \frac{\partial \mathcal{P}}{\partial \vartheta}\bigg|_{\vartheta = \tilde{\vartheta}_{\mu k}^t} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

where $\tilde{\vartheta}_{\mu k}^{0\ell t}$ is the solution of Eq. (D2). If $\boldsymbol{\Theta}^{\mathsf{T}} = (\vartheta^0, \vartheta^1, \dots, \vartheta^L)$ and $(\mathbf{H}_{\mathcal{H}})_{jk} = \delta_{jk}[\delta_{j0}h_0 + (1-\delta_{j0})L^{-1}h_2] + (\delta_{j0} + \delta_{k0})(1-\delta_{jk})L^{-1}h_1$ is the Hessian of $\Sigma_{\ell=1}^{L} {}^{(1\text{RSB})}\mathcal{H}(\vartheta^0, \vartheta^\ell, y_\mu)$, then

$$\sum_{\ell=1}^{L} {}^{(1\text{RSB})}\mathcal{H}(\vartheta^0, \vartheta^\ell, y_\mu) \simeq LE_0 + \frac{L}{2} \Delta\boldsymbol{\Theta}^{\mathsf{T}} \mathbf{H}_{\mathcal{H}} \Delta\boldsymbol{\Theta}.$$

The matrix $\mathbf{H}_{\mathcal{H}}$ has the same structure as $\mathbf{H}_\Phi$, therefore, the eigenvalues and eigenvectors of $\mathbf{H}_{\mathcal{H}}$ can be obtained adapting Eqs. (B3) and (B4) by the substitutions $\alpha_0 = h_0$, $-\beta_0 = h_1$ and $\gamma_0 = h_2$. Expanding $\mathcal{P}(\vartheta, y_\mu)$ at the saddle point $\tilde{\vartheta}_{\mu k}^{0\ell t}$ one obtains $\mathcal{P}(\vartheta_{\mu k}^{0\ell' t}, y_\mu) \simeq \mathcal{P}_0 + \mathcal{P}_1(\Delta\vartheta^0 + \Delta\vartheta^{\ell'}) + \frac{1}{2}\mathcal{P}_2(\Delta\vartheta^0 + \Delta\vartheta^{\ell'})^2$ where $\mathcal{P}_j \equiv \frac{\partial^j \mathcal{P}}{\partial \vartheta^j}\big|_{\vartheta = \tilde{\vartheta}_{\mu k}^t}$. The resulting messages are

$${}^{(1\text{RSB})}\hat{m}_{\mu k}^{t+1} = \varepsilon_{\mu k} \frac{\int d\boldsymbol{\Theta} \exp\left(-\frac{nL}{2}\Delta\boldsymbol{\Theta}^{\mathsf{T}}\mathbf{H}_{\mathcal{H}}\Delta\boldsymbol{\Theta}\right)\left(\mathcal{P}_0 + \mathcal{P}_1(\Delta\vartheta^0 + \Delta\vartheta^{\ell'}) + \frac{1}{2}\mathcal{P}_2(\Delta\vartheta^0 + \Delta\vartheta^{\ell'})^2\right)}{\int d\boldsymbol{\Theta} \exp\left(-\frac{nL}{2}\Delta\boldsymbol{\Theta}^{\mathsf{T}}\mathbf{H}_{\mathcal{H}}\Delta\boldsymbol{\Theta}\right)},$$

where the term proportional to $\mathcal{P}_1$ vanishes for parity reasons. In the basis of eigenvectors of $\mathbf{H}_{\mathcal{H}}$, i.e., $\boldsymbol{\Gamma} = \mathbf{U}^{\mathsf{T}}\boldsymbol{\Theta} = (\gamma_0, \gamma_1, \dots, \gamma_L)^{\mathsf{T}}$ where $\mathbf{U}$ is adapted from Eq. (B5), the message has the form

$${}^{(1\text{RSB})}\hat{m}_{\mu k}^{t+1} \simeq \varepsilon_{\mu k} \frac{\int d\boldsymbol{\Gamma} \exp\left(-\frac{nL}{2}\sum_{\ell=0}^{L}\lambda_\ell(\Delta\gamma^\ell)^2\right)\left(\mathcal{P}_0 + \frac{1}{2}\mathcal{P}_2\Delta\boldsymbol{\Gamma}^{\mathsf{T}}\mathbf{M}'_{0\ell'}\Delta\boldsymbol{\Gamma}\right)}{\int d\boldsymbol{\Gamma} \exp\left(-\frac{nL}{2}\sum_{\ell=0}^{L}\lambda_\ell(\Delta\gamma^\ell)^2\right)},$$

where λ_ℓ are the eigenvalues of $\mathbf{H}_{\mathcal{H}}$ and $\mathbf{M}'_{0\ell'}$ is adapted from Eq. (B9).

The expression for the message is reduced to

$${}^{(1\text{RSB})}\hat{m}_{\mu k}^{t+1} \simeq \varepsilon_{\mu k}\left[\mathcal{P}_0 + \frac{1}{n}\frac{\mathcal{P}_2}{2h_2} + O\left(\frac{1}{nL}\right)\right] \simeq \varepsilon_{\mu k}\frac{\tilde{\vartheta}_{\mu k}^t - u_{\mu k}^t}{2V^t - R^t} + \frac{\varepsilon_{\mu k}}{2n}\frac{\mathcal{P}_2 V^t}{1 - \mathcal{P}_1 V^t}. \tag{C8}$$

The expression for the messages from **b** nodes to **y** nodes is

$$m_{\mu k}^t = \sum_{\{\mathbf{b}_k\}} b_k^{a'} P^t(\mathbf{b}_k | \{y_{\nu \neq \mu}\}) = \frac{\sum_{\{\mathbf{b}_k\}} b_k^{a'} \prod_{\nu \neq \mu} \sum_{\{\mathbf{b}_{l \neq k}\}} P(y_\nu | \mathbf{B}) \prod_{l \neq k} P^{t-1}(\mathbf{b}_l | \{y_{\sigma \neq \nu}\})}{\sum_{\{\mathbf{b}_k\}} \prod_{\nu \neq \mu} \sum_{\{\mathbf{b}_{l \neq k}\}} P(y_\nu | \mathbf{B}) \prod_{l \neq k} P^{t-1}(\mathbf{b}_l | \{y_{\sigma \neq \nu}\})},$$

which can be approximated by

$$m_{\mu k}^t \simeq \frac{\sum\limits_{\{\mathbf{b}_k\}} b_k^{a'} \int d\boldsymbol{\Delta}_{\nu k} P(y_\nu|\boldsymbol{\Delta}_{\nu k};\gamma) P(\boldsymbol{\Delta}_{\nu k}|\mathbf{B})[1 + \varepsilon_{\nu k}\mathbf{b}_k^\top \nabla_{\boldsymbol{\Delta}_{\nu k}} \ln P(y_\nu|\boldsymbol{\Delta}_{\nu k};\gamma)]}{\sum\limits_{\{\mathbf{b}_k\}} \int d\boldsymbol{\Delta}_{\nu k} P(y_\nu|\boldsymbol{\Delta}_{\nu k};\gamma) P(\boldsymbol{\Delta}_{\nu k}|\mathbf{B})[1 + \varepsilon_{\nu k}\mathbf{b}_k^\top \nabla_{\boldsymbol{\Delta}_{\nu k}} \ln P(y_\nu|\boldsymbol{\Delta}_{\nu k};\gamma)]}$$

$$= \frac{\sum\limits_{b_k^{a'}=\pm 1} b_k^{a'} \int d\boldsymbol{\Delta}_{\nu k} P(y_\nu|\boldsymbol{\Delta}_{\nu k};\gamma) P(\boldsymbol{\Delta}_{\nu k}|b_k^{a'})\left(1 + \varepsilon_{\nu k}b_k^{a'}\frac{\partial}{\partial\Delta_{\mu k}^{a'}} \ln P(y_\nu|\boldsymbol{\Delta}_{\nu k};\gamma)\right)}{\sum\limits_{b_k^{a'}=\pm 1} \int d\boldsymbol{\Delta}_{\nu k} P(y_\nu|\boldsymbol{\Delta}_{\nu k};\gamma) P(\boldsymbol{\Delta}_{\nu k}|b_k^{a'})\left(1 + \varepsilon_{\nu k}b_k^{a'}\frac{\partial}{\partial\Delta_{\mu k}^{a'}} \ln P(y_\nu|\boldsymbol{\Delta}_{\nu k};\gamma)\right)} = \frac{\sum\limits_{b_k^a=\pm 1} b_k^a \prod\limits_{\nu\neq\mu} \frac{1 + \hat{m}_{\nu k}^t b_k^a}{\mathcal{N}_{\nu k}^t}}{\sum\limits_{b_k^a=\pm 1} \prod\limits_{\nu\neq\mu} \frac{1 + \hat{m}_{\nu k}^t b_k^a}{\mathcal{N}_{\nu k}^t}}$$

$$= \frac{\prod\limits_{\nu\neq\mu} \frac{1 + \hat{m}_{\nu k}^t}{\mathcal{N}_{\nu k}^t} - \prod\limits_{\nu\neq\mu} \frac{1 - \hat{m}_{\nu k}^t}{\mathcal{N}_{\nu k}^t}}{\prod\limits_{\nu\neq\mu} \frac{1 + \hat{m}_{\nu k}^t}{\mathcal{N}_{\nu k}^t} + \prod\limits_{\nu\neq\mu} \frac{1 - \hat{m}_{\nu k}^t}{\mathcal{N}_{\nu k}^t}} = \tanh\left(\sum_{\nu\neq\mu} \text{arctanh}(\hat{m}_{\nu k}^t)\right),$$

but since $\hat{m}_{\nu k}^t \sim O(\varepsilon_{\nu k})$ we have that

$$m_{\mu k}^t \simeq \tanh\left(\sum_{\nu\neq\mu} \hat{m}_{\nu k}^t\right). \tag{C9}$$

APPENDIX D: THE SADDLE POINT OF $\mathcal{H}$

For the RS case the equation to be solved is

$$\frac{\partial}{\partial\vartheta}{}^{\text{(RS)}}\mathcal{H}(\vartheta,y_\mu) = \frac{\vartheta - u_{\mu k}^t}{R_{\mu k}^t} - \frac{\partial}{\partial\vartheta} \ln \mathcal{G}(\vartheta,y_\mu)$$

$$= \frac{\vartheta - u_{\mu k}^t}{R_{\mu k}^t} - \mathcal{P}(\vartheta,y_\mu),$$

thus, the equation to be satisfied is

$$\tilde{\vartheta}_{\mu k}^t = u_{\mu k}^t + R^t \mathcal{P}(\tilde{\vartheta}_{\mu k}^t, y_\mu). \tag{D1}$$

For the 1RSB case we have that $\frac{\partial}{\partial\vartheta^0}{}^{\text{(1RSB)}}\mathcal{H} = \frac{\partial}{\partial\vartheta^\ell}{}^{\text{(1RSB)}}\mathcal{H} = 0$, resulting in the set of equations:

$$0 = \tilde{\vartheta}_{\mu k}^{0t} - (V^t - R^t)\mathcal{P}(\tilde{\vartheta}_{\mu k}^{0\ell t}, y_\mu),$$

$$0 = \tilde{\vartheta}_{\mu k}^{\ell t} - V^t \mathcal{P}(\tilde{\vartheta}_{\mu k}^{0\ell t}, y_\mu),$$

which is equivalent to

$$\tilde{\vartheta}_{\mu k}^{0\ell t} = u_{\mu k}^t + (2V^t - R^t)\mathcal{P}(\tilde{\vartheta}_{\mu k}^{0\ell t}, y_\mu), \tag{D2}$$

where $\vartheta_{\mu k}^{0\ell t} = \vartheta^0 + \vartheta^\ell + u_{\mu k}^t$. Observed that Eq. (D2) is Equivalent to Eq. (D1) and that the ground state $\tilde{\vartheta}_{\mu k}^t$ is independent of the indices 0 and ℓ.

APPENDIX E: THE OPTIMIZATION CONDITION

Our goal is to devise an algorithm that returns a better estimate of the message at each iteration; we therefore apply a variational approach that optimizes the free parameters of the model at each iteration. We expect to find a suitable set of parameters γ^c that maximizes the drop in error per bit rate.

The error function has the form

$$\mathcal{E}^t(\gamma) \equiv \lambda^2 P_b^t - M^t/\sqrt{N^t}, \tag{E1}$$

where λ^2 is a positive constant.

Observe that

$$M^t - N^t = \frac{1}{\sqrt{2\pi F^t}} \int dz \exp\left(-\frac{z^2 + (E^t)^2}{2F^t}\right.$$
$$\left. - \ln\cosh(z)\right)\tanh(z)\sinh\left(\frac{E^t - F^t}{F^t}z\right),$$

and that $\text{sgn}\left[\tanh(z)\sinh\left(\frac{E^t-F^t}{F^t}z\right)\right] = \text{sgn}(E^t - F^t)\,\forall\,z$. Therefore $\text{sgn}(E^t - F^t) = \text{sgn}(M^t - N^t)$.

The second term on the right-hand side of Eq. (E1) is an implicit function of the parameters γ through E^t and F^t, therefore

$$\frac{\partial}{\partial\gamma_i}\left(\frac{M^t}{\sqrt{N^t}}\right) = \frac{\partial}{\partial E^t}\left(\frac{M^t}{\sqrt{N^t}}\right)\frac{\partial E^t}{\partial\gamma_i} + \frac{\partial}{\partial F^t}\left(\frac{M^t}{\sqrt{N^t}}\right)\frac{\partial F^t}{\partial\gamma_i}, \tag{E2}$$

where the partial derivatives with respect to E^t and F^t are

$$\frac{\partial}{\partial E^t}\left(\frac{M^t}{\sqrt{N^t}}\right) = (N^t)^{-3/2} \int \mathcal{D}z[1 - \tanh^2(\sqrt{F^t}z + E^t)]$$
$$\times[N^t - M^t\tanh(\sqrt{F^t}z + E^t)]$$

$$\frac{\partial}{\partial F^t}\left(\frac{M^t}{\sqrt{N^t}}\right) = (N^t)^{-3/2} \int \mathcal{D}z\frac{z}{2\sqrt{F^t}}[1 - \tanh^2(\sqrt{F^t}z + E^t)]$$
$$\times[N^t - M^t\tanh(\sqrt{F^t}z + E^t)].$$

By the definition of the field $b_k h_{\mu k}^t$ we have that $\text{sgn}(b_k h_{\mu k}^t) = \text{sgn}(b_k m_{\mu k}^t) = \text{sgn}(b_k m_k^t)$. Exploiting Gaussian properties of the distribution of $h_{\mu k}^t$ (9),

$$P_b^t \simeq \frac{1}{2K}\sum_{k=1}^{K}\left[1 - \text{sgn}(b_k h_{\mu k}^t)\right] \sim \int_{-\infty}^{\infty}\frac{du}{\sqrt{2\pi F^t}}\exp\left(-\frac{(u - E^t)^2}{2F^t}\right)\frac{1}{2}\left[1 - \text{sgn}(u)\right] = \int_{-\infty}^{-E^t/\sqrt{F^t}}\mathcal{D}u,$$

and we suppose that E^t and F^t are both explicit functions of the parameters γ, therefore

$$\frac{\partial P_b^t}{\partial \gamma_i} = -\frac{1}{\sqrt{2\pi F^t}}\exp\left(-\frac{(E^t)^2}{2F^t}\right)\left(\frac{\partial E^t}{\partial \gamma_i} - \frac{1}{2}\frac{E^t}{F^t}\frac{\partial F^t}{\partial \gamma_i}\right).$$

By differentiation equation (E1) and using Eq. (E2) one obtains

$$\frac{\partial}{\partial \gamma_i}\mathcal{E}^t = -\frac{\lambda^2}{\sqrt{2\pi F^t}}\exp\left(-\frac{(E^t)^2}{2F^t}\right)\left(\frac{\partial E^t}{\partial \gamma_i} - \frac{1}{2}\frac{E^t}{F^t}\frac{\partial F^t}{\partial \gamma_i}\right) - (N^t)^{-3/2}\int \mathcal{D}z\frac{N^t - M^t\tanh(\sqrt{F^t}z + E^t)}{\cosh^2(\sqrt{F^t}z + E^t)}\left(\frac{\partial E^t}{\partial \gamma_i} + \frac{z}{2\sqrt{F^t}}\frac{\partial F^t}{\partial \gamma_i}\right)$$

$$= -(F^t N^t)^{-3/2}\int \frac{du}{\sqrt{2\pi}}\exp\left(-\frac{(u - E^t)^2}{2F^t}\right)\frac{u}{2}\frac{N^t - M^t\tanh(u)}{\cosh^2(u)} - \left(\frac{\partial E^t}{\partial \gamma_i} - \frac{1}{2}\frac{E^t}{F^t}\frac{\partial F^t}{\partial \gamma_i}\right)$$

$$\times\left[\frac{\lambda^2}{\sqrt{2\pi F^t}}\exp\left(-\frac{(E^t)^2}{2F^t}\right) + \int \frac{du}{\sqrt{2\pi F^t(N^t)^3}}\exp\left(-\frac{(u - E^t)^2}{2F^t}\right)\frac{N^t - M^t\tanh(u)}{\cosh^2(u)}\right]. \tag{E3}$$

To optimize $\mathcal{E}^t$ with respect to γ_i one requires $\frac{\partial}{\partial \gamma_i}\mathcal{E}^t = 0$. The first term on the right-hand side of Eq. (E3) is independent of the index i and is zero if and only if the integrand is an odd function. This is true if $\tanh(u) = \frac{N^t}{M^t}\tanh\left(\frac{uE^t}{F^t}\right)\forall u \in \mathbb{R}$. This condition is only satisfied if $E^t(\gamma^c) = F^t(\gamma^c)$ which automatically makes $M^t = N^t$. By the application of this condition, the sum between square brackets in the second term on the right-hand side of Eq. (E3) is always positive, which implies $\frac{\partial E^t}{\partial \gamma_i} - \frac{1}{2}\frac{E^t}{F^t}\frac{\partial F^t}{\partial \gamma_i}\big|_{\gamma_i^c} = 0$.

The conditions $E^t(\gamma^c) = F^t(\gamma^c)$ and $\frac{\partial E^t}{\partial \gamma_i} - \frac{1}{2}\frac{E^t}{F^t}\frac{\partial F^t}{\partial \gamma_i}\big|_{\gamma_i^c} = 0$ imply that

$$\ln E^t = e_0 + \mathbf{e}_1^\mathsf{T}(\gamma - \gamma^c) + \frac{1}{2}(\gamma - \gamma^c)^\mathsf{T}\mathbf{E}_2(\gamma - \gamma^c) + \cdots,$$

$$\ln F^t = e_0 + 2\mathbf{e}_1^\mathsf{T}(\gamma - \gamma^c) + \frac{1}{2}(\gamma - \gamma^c)^\mathsf{T}\mathbf{F}_2(\gamma - \gamma^c) + \cdots,$$

therefore, if the critical point is a minimum, then the expansion $E^t/\sqrt{F^t} = \exp\left[\frac{1}{2}e_0 + \frac{1}{2}(\gamma - \gamma^c)^\mathsf{T}\left(\mathbf{E}_2 - \frac{1}{2}\mathbf{F}_2\right)(\gamma - \gamma^c) + \dots\right]$ has a second term that satisfy the conditions: $\det\left(\mathbf{E}_2 - \frac{1}{2}\mathbf{F}_2\right) > 0$ and $\left(\mathbf{E}_2 - \frac{1}{2}\mathbf{F}_2\right)_{ii} < 0$, validating the optimization process.

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

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

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

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

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

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

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

[8] J. P. Neirotti and D. Saad, Europhys. Lett. **71**, 866 (2005).

[9] Although we will be using the terms RS and RSB, it should be clear that this is not directly related to the replica approach [1,7], but merely uses similar structures for the cross-replica correlations.

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

[11] A. Engel and C. Van den Broeck, *Statistical Mechanics of Learning* (Cambridge University Press, Cambridge, 2001).

[12] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).

[13] H. S. Seung, H. Sompolinsky, and N. Tishby, Phys. Rev. A **45**, 6056 (1992).

[14] Y. Weiss, Neural Comput. **12**, 1 (2000).

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

[16] J. S. Yedidia, W. T. Freeman, and Y. Weiss, Adv. Neural Inf. Process. Syst. **13**, 698 (2000).

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

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

[19] A. Braunstein and R. Zecchina, Phys. Rev. Lett. **96**, 030201 (2006).

[20] Y. Kabashima, J. Phys. Soc. Jpn. **74**, 2133 (2005).