

PAPER

# A theory of solving TAP equations for Ising models with general invariant random matrices

To cite this article: Manfred Oppen *et al* 2016 *J. Phys. A: Math. Theor.* **49** 114002

View the [article online](#) for updates and enhancements.

## You may also like

- [Extended Plefka expansion for stochastic dynamics](#)  
B Bravi, P Sollich and M Oppen
- [Perturbative construction of mean-field equations in extensive-rank matrix factorization and denoising](#)  
Antoine Maillard, Florent Krzakala, Marc Mézard et al.
- [Weighted community detection and data clustering using message passing](#)  
Cheng Shi, Yanchen Liu and Pan Zhang

# A theory of solving TAP equations for Ising models with general invariant random matrices

Manfred Oppen<sup>1,4</sup>, Burak Çakmak<sup>2</sup> and Ole Winther<sup>3</sup>

<sup>1</sup> Department of Artificial Intelligence, Technische Universität Berlin, Marchstrae 23, Berlin 10587, Germany

<sup>2</sup> Department of Electronic Systems, Aalborg Universitet, Fredrik Bajers Vej 7, Ålborg 9220, Denmark

<sup>3</sup> DTU Compute, Danmarks Tekniske Universitet, Anker Engelunds Vej 1, Lyngby 2800, Denmark

E-mail: [manfred.oppen@tu-berlin.de](mailto:manfred.oppen@tu-berlin.de), [buc@es.aau.dk](mailto:buc@es.aau.dk) and [olwi@dtu.dk](mailto:olwi@dtu.dk)

Received 1 September 2015, revised 7 December 2015

Accepted for publication 17 December 2015

Published 8 February 2016



CrossMark

## Abstract

We consider the problem of solving TAP mean field equations by iteration for Ising models with coupling matrices that are drawn at random from general invariant ensembles. We develop an analysis of iterative algorithms using a dynamical functional approach that in the thermodynamic limit yields an effective dynamics of a single variable trajectory. Our main novel contribution is the expression for the implicit memory term of the dynamics for general invariant ensembles. By subtracting these terms, that depend on magnetizations at previous time steps, the implicit memory terms cancel making the iteration dependent on a Gaussian distributed field only. The TAP magnetizations are stable fixed points if a de Almeida–Thouless stability criterion is fulfilled. We illustrate our method explicitly for coupling matrices drawn from the random orthogonal ensemble.

Keywords: Ising models, TAP equations, dynamical functional theory, iterative convergent algorithms, random matrices, free probability

(Some figures may appear in colour only in the online journal)

<sup>4</sup> Author to whom any correspondence should be addressed.

## 1. Introduction

TAP equations provide generalized mean field equations for statistical physics models with random, infinite range interactions which (under certain conditions) are assumed to be exact in the limit of an infinite system [1]. In recent years there has been an increasing interest in such equations within and also outside the statistical physics community. This is partly due to the fact that the TAP approach can be applied to statistical inference in probabilistic models in information theory [2, 3], statistics [4] and machine learning [5, 6]. Originally developed by D J Thouless, P Anderson and R Palmer for the Sherrington–Kirkpatrick (SK) model of an Ising spin glass [7], the TAP approach has been generalized to a variety of other problems [4]. This includes models with continuous variables rather than Ising spins but also cases where the independent random interactions are replaced by other, more structured statistical ensembles that allow for certain dependencies.

While methods for deriving the TAP approach for different models are now well established it is not necessarily clear how the resulting system of nonlinear equations can be solved efficiently. A naive algorithm based on a simple iteration of the equations usually fails to achieve convergence. This problem has been addressed by a paper of Bolthausen for the case of the SK model [8]. He has analyzed the dynamics of iterations rigorously and has shown how the iterations can be altered in order to achieve exponential convergence (above the so-called AT line of stability). Other ideas to arrive at a convergent method are based on taking the limit of dense couplings in belief propagation algorithms, see [2, 9] and [10], (for rigorous analyses [3] and [11]). Unfortunately, for this approach it is necessary to augment the original variables by auxiliary ones, such that the interactions in the new model are independent. For example, the Hopfield model can be represented by a bipartite graph of Ising spins and continuous variables. It is not clear how such a method should be set up for a matrix of interactions with more general statistical dependencies. Also, taking the dense coupling limit of an approximate message passing (AMP) algorithm valid for sparse coupling problems will not always lead to the correct dense coupling algorithm. Here we will construct a theory for dynamics using dense couplings as the starting point.

In this paper we will address the problem of solving TAP equations for Ising models with dense random coupling matrices with a general invariant probability distribution. Our analysis is based on dynamical mean field theory which allows the study of the dynamics of iterative algorithms in the thermodynamic limit by a suitable average over the ensemble of couplings. It turns out that by including certain memory terms in the iteration, the effective field in the dynamics becomes a simple Gaussian random variable suggesting that the dynamics might converge. The explicit form of the memory terms depends explicitly on the statistical ensemble of couplings. We show that our method reproduces previous convergent algorithms for SK and Hopfield models. We also work out the details of our theory for the spin model with orthogonal random couplings [12, 13]. Simulations of the resulting algorithms show exponential convergence above a line of stability which can be identified with the so-called de Almeida–Thouless (AT) line.

The paper is organized as follows: in section 2 we introduce the general random matrix formulation for the TAP equations. In section 3 we present the results of dynamical functional theory (DFT). In section 4, we introduce ‘the single-step memory construction’ iterative algorithm for solving TAP equations. Section 5 is devoted to the derivation of the AT stability condition. The discussion and outlook are presented in section 6. Lengthy technical derivations are deferred to the appendix.

## 2. General invariant random matrix ensembles

We will consider Ising models with pairwise interactions given by the Gibbs distribution for the spins  $\mathbf{S} = (S_1, \dots, S_N)$

$$P(\mathbf{S}) = \frac{1}{Z} \exp \left[ \sum_{i < j}^N J_{ij} S_i S_j + \sum_i^N h_i S_i \right] \quad (1)$$

with  $Z$  denoting the normalization constant. We are interested in the case where the matrix  $\mathbf{J}$  is random with the condition that the marginal density of couplings  $p(J_{ij})$  is the same for all pairs  $(i, j)$  but couplings might be dependent random variables. A simple way to define such a class of random matrices is via the so-called *invariant* ensembles [14]. A random matrix  $\tilde{\mathbf{J}}$  is called invariant if it has the same probability distribution as  $\mathbf{V}^\dagger \tilde{\mathbf{J}} \mathbf{V}$  for an orthogonal matrix  $\mathbf{V}$  which is independent of  $\tilde{\mathbf{J}}$ . Equivalently, it admits the spectral decomposition

$$\tilde{\mathbf{J}} = \mathbf{O}^\dagger \mathbf{\Lambda} \mathbf{O} \quad (2)$$

where  $\mathbf{O}$  is Haar distributed (i.e. it is a random orthogonal matrix) and independent of the diagonal matrix  $\mathbf{\Lambda}$ . This characterization of invariant matrices involves the diagonal elements  $\tilde{J}_{ii}$  which are absent in (1). However, one can show that as  $N$  tends to infinity if the spectrum of  $\tilde{\mathbf{J}}$  converges almost surely to a compactly supported distribution such that the smallest and largest eigenvalue of  $\tilde{\mathbf{J}}$  converge almost surely to the infimum and supremum of the support, respectively, we have (in the almost sure sense)  $\tilde{J}_{ii} - \frac{1}{N} \text{tr}(\tilde{\mathbf{J}}) \rightarrow 0$ ,  $\forall i$  as  $N$  tends to infinity, see A. In other words the diagonal elements of an invariant matrix converge to the same deterministic limit. Thus, we may define asymptotically invariant couplings  $\mathbf{J}$  as  $J_{ij} = \beta \tilde{J}_{ij}$  for  $i \neq j$  and  $J_{ii} = 0$ , where we include an inverse temperature factor  $\beta$  in the definition. The couplings for the standard SK and Hopfield models belong to this class of matrices for  $(-\tilde{\mathbf{J}})$  being the Gaussian Wigner matrix and the null Wishart matrix (i. e. a sample covariance matrix of independent Gaussian random vectors whose entries are zero mean, independent and identical distributed), respectively.

### 2.1. The generating function and the R-transform

We will later need the generating function of asymptotically invariant random matrices given by [13, 15]

$$G(\mathbf{Q}) \triangleq \lim_{N \rightarrow \infty} \frac{1}{N} \log \langle e^{\frac{N}{2} \text{tr}(\mathbf{Q} \mathbf{J})} \rangle_{\mathbf{J}} \quad (3)$$

with the symmetric matrix  $\mathbf{Q}$  having a finite rank. Since we believe that our paper might be of interest to researchers with an information theory background, we will briefly mention how  $G$  is related to quantities which are well known in the theory of free probability [16], which is a powerful approach to random matrix theory. Setting

$$G(x) = \frac{1}{2} \int_0^x d\omega \, R(\omega) \quad (4)$$

one can show that  $R$  equals the so-called R-transform (in the theory of free probability) of the limiting spectrum. Its formal definition can be given in terms of the Cauchy transform: let  $P$  denote the limiting spectrum of  $\mathbf{J}$ . Moreover let

<sup>†</sup> Here  $(\cdot)^\dagger$  denotes transposition.

$$M(z) \triangleq \frac{1}{z} + \sum_{n=1}^{\infty} \frac{m_n}{z^{n+1}}, \quad (5)$$

where  $m_n$  are  $n$ th order moments of the limiting spectrum  $P$ , i.e.  $m_n = \int dP(\lambda) \lambda^n$ . Then the  $R$ -transform of  $P$  is given by

$$R(x) = M^{-1}(x) - \frac{1}{x} \quad (6)$$

with  $M^{-1}$  denoting the composition inverse of  $M$ . It admits the power series expansion

$$R(x) = \sum_{n=1}^{\infty} c_n x^{n-1}, \quad (7)$$

where  $c_n$  are known as the *free cumulants* of  $P$ . For example, the first two free cumulants  $c_1$  and  $c_2$  are the mean and variance of the distribution  $P$ , respectively, i.e.  $c_1 = m_1$  and  $c_2 = m_2 - m_1^2$ . For details we refer the reader to [17].

In the following we give the  $R$ -transforms for some random matrix ensembles that we will discuss later. We first point out the simple identity

$$R(x) = \beta(\tilde{R}(\beta x) - \tilde{c}_1). \quad (8)$$

In this expression  $\tilde{R}$  denotes the  $R$ -transform of the limiting spectrum of  $\tilde{J}$  and  $\tilde{c}_1 = \tilde{R}(0)$ . We next provide the explicit form of  $R(x)$  for the SK, Hopfield and random orthogonal [12, 13] models, respectively: (i) for the SK case we have a (symmetric) matrix  $\tilde{J}$  whose diagonal entries are zero and upper-triangle entries are independent identically distributed (iid) Gaussian with zero mean and variance  $1/N$ . Note that in this case we have  $J = \beta\tilde{J}$ ; so that  $R(x) = \beta^2 x$  [17]; (ii) for the Hopfield model, we consider entries of an  $(N/\alpha) \times N$  matrix  $H$  that are iid and Gaussian with zero mean and variance  $\alpha/N$ . Moreover let  $\tilde{J} = -H^\dagger H$ , i.e.  $(-\tilde{J})$  is a null Wishart matrix whose limiting spectrum is given by the Marchenko–Pastur distribution. By invoking the  $R$ -transform of the Marchenko–Pastur distribution, see e.g. [17], we have

$$R(x) = \frac{\beta^2 \alpha x}{1 + \beta \alpha x}; \quad (9)$$

(iii) Finally, for the random orthogonal case we consider a spectral decomposition (2) such that  $\tilde{J} = O^\dagger \Lambda O$ . Here the diagonal entries of the diagonal matrix  $\Lambda$  are composed of  $\pm 1$  such that  $\text{tr}(\Lambda) = 0$ . Then, one can easily show that

$$R(x) = \frac{-1 + \sqrt{1 + 4\beta^2 x^2}}{2x}. \quad (10)$$

The latter result was given in [12].

## 2.2. TAP equations for general invariant couplings

TAP equations are a set of self-consistent equations for the vector of magnetizations  $\mathbf{m} = \langle \mathbf{S} \rangle$  where the brackets denote expectation w.r.t. the Gibbs distribution (1). For a general invariant ensemble they have been derived first in [12] using the large  $N$  scaling of a perturbation expansion. A second derivation using the cavity method and the large  $N$  limit of the ‘adaptive’ TAP equations can be found [4]. We have provided a more rigorous derivation of the transition from ‘adaptive TAP’ to the self-averaging limit using random matrix theory in appendix B. The resulting TAP equations read

$$\mathbf{m} = \tanh(\psi) \quad (11)$$

$$\psi = \mathbf{h} + \mathbf{J}\mathbf{m} - \mathbf{R}(1 - q)\mathbf{m} \quad (12)$$

where  $q \triangleq \frac{1}{N} \mathbf{m}^\dagger \mathbf{m}$  and  $\mathbf{h}$  is the vector of non-random external fields. Note that the only dependency on the random matrix ensemble is via the R-transform  $\mathbf{R}(1 - q)$  in the so-called Onsager term which is a correction to the naive mean field term  $\mathbf{J}\mathbf{m}$ . One can show that  $\Psi_i$  is the mean of the cavity field. Furthermore, following the calculations of [4] one finds that  $\Psi_i$  is Gaussian distributed (in the large  $N$  limit) with respect to the random couplings  $\mathbf{J}$  with mean  $h_i$  and variance

$$\langle (\Psi_i - h_i)^2 \rangle = q \mathbf{R}'(1 - q). \quad (13)$$

Hence, the subtraction of the Onsager term  $\mathbf{R}(1 - q)\mathbf{m}$  from the mean field  $\mathbf{J}\mathbf{m}$  makes the remainder Gaussian. We will next transfer the idea of a Gaussian field from the static solutions to the dynamics of an algorithm.

### 3. The results of dynamical functional theory

Dynamical properties of disordered systems can be computed by the method of dynamical functionals [18]. In the limit  $N \rightarrow \infty$  this method provides us with exact results for the marginal distribution of a trajectory of a single variable (in our case a magnetization  $m_i(t)$ ), when we define a dynamics of an algorithm for the solution of the TAP equations. As a typical result of such a calculation one finds that the ‘field’  $\mathbf{J}\mathbf{m}(t)$  becomes a sum of a Gaussian term and a memory term which includes the magnetizations at all previous times. This memory often makes the dynamics of disordered systems highly complex allowing, e.g. for a persistent dependency on the initial conditions and thus a failure to converge to a unique fixed point. Hence, we propose to introduce *explicit* memory terms which are chosen to cancel the *implicit* memory terms derived from the DFT. In such a way, at each time step, the update of the magnetization for the algorithm involve a Gaussian distributed random field only and we expect that we might obtain good convergence results. This Gaussian property of the effective dynamical field was already shown for a Hopfield model in [2] and [9] (and proved in [3]) and reappeared in Bolthausen’s iterative construction of solutions to the TAP equations for the SK model in [8].

We start with defining a set of dynamical equations which could serve as a candidate algorithm for solving the TAP equations for invariant random coupling matrices

$$\mathbf{m}(t) = f_t(\{\gamma(\tau), \mathbf{m}(\tau)\}_{\tau=0}^{t-1}) \quad (14)$$

$$\gamma(t) = \mathbf{h} + \mathbf{J}\mathbf{m}(t) \quad (15)$$

for  $\tau = 0, \dots, t$  which depend on the field  $\mathbf{J}\mathbf{m}(t)$  and the previous local magnetizations  $\mathbf{m}(\tau)$ . Here  $f_t$  is an appropriate sequence of nonlinear scalar functions. Our goal is to obtain the statistics of a single trajectory of (14)–(15), when  $\mathbf{J}$  is a random matrix with generating function (3). To do so we make use of the DFT analysis as described in [19, 20] which is a discrete time version of the method of [18]. We also refer the reader to [21] where DFT was used to analyze the AMP algorithm in the context of the CDMA communication algorithm.

We introduce the generating functional corresponding to the dynamics (14)–(15) as

$$Z(\{\mathbf{I}(t)\}) = \int \prod_{t=0}^{T-1} \{d\mathbf{m}(t)d\gamma(t) \delta(\mathbf{m}(t) - f_t(\{\gamma(\tau), \mathbf{m}(\tau)\}_{\tau=0}^{t-1})) \delta(\gamma(t) - \mathbf{h} - \mathbf{J}\mathbf{m}(t))e^{i\gamma(t)\mathbf{I}(t)}\}. \quad (16)$$

Note that  $Z(\{\mathbf{I}(t) = 0\}) = 1$ . The statistics of the variables can be computed from the averaged generating functional  $\langle Z(\{\mathbf{I}(t)\}) \rangle_J$ . In the large  $N$  limit we obtain that (see C)

$$\begin{aligned} \langle Z(\{\mathbf{I}(t)\}) \rangle_J &\simeq \prod_{n=1}^N \int d\mathcal{N}(\{\phi_n(t)\}; 0, \mathcal{C}_\phi) \\ &\prod_{t=0}^{T-1} \{dm_n(t)d\gamma_n(t) \delta(m_n(t) - f_t\{m_n(\tau), \gamma_n(\tau)\}_{\tau=0}^{t-1}) \\ &\delta\left(\gamma_n(t) - h_n - \sum_{s<t} \hat{\mathcal{G}}(t, s)m_n(s) - \phi_n(t)\right)e^{i\gamma_n(t)I_n(t)}\} \end{aligned} \quad (17)$$

with  $\mathcal{N}(\cdot; \mu, \Sigma)$  denoting the multivariate normal distribution with mean  $\mu$  and covariance  $\Sigma$ . This result shows that in the large  $N$  limit single trajectories can be treated as independent following the effective stochastic dynamical process given by

$$\mathbf{m}(t) = f_t(\{\gamma(\tau), \mathbf{m}(\tau)\}_{\tau=0}^{t-1}) \quad (18)$$

$$\gamma(t) = \mathbf{h} + \sum_{\tau=0}^{t-1} \hat{\mathcal{G}}(t, s)\mathbf{m}(\tau) + \phi(t). \quad (19)$$

Here  $\phi(t)$  is a vector of independent Gaussian random variables with covariance matrix  $\mathcal{C}_\phi$  given by

$$\mathcal{C}_\phi = \sum_{n=1}^{\infty} c_n \sum_{k=0}^{n-2} \mathcal{G}^k \mathcal{C}(\mathcal{G}^\dagger)^{n-2-k}, \quad (20)$$

where  $\mathcal{G}$  and  $\mathcal{C}$  are  $T \times T$  the response and the correlation matrices, respectively. With slight abuse of notation, their  $(t+1, \tau+1)$  indexed entries are given by

$$\mathcal{G}(t, \tau) = \frac{1}{N} \sum_{i=1}^N \left\langle \frac{\partial m_i(t)}{\partial \phi_i(\tau)} \right\rangle_{\phi_i} \quad (21)$$

$$\mathcal{C}(t, \tau) = \frac{1}{N} \sum_{i=1}^N \langle m_i(t)m_i(\tau) \rangle_{\phi_i}. \quad (22)$$

Moreover the specific random matrix ensemble enters the result through the coefficients  $c_n$ , see (7), and the memory matrix  $\hat{\mathcal{G}}$  given by

$$\hat{\mathcal{G}} = \mathbf{R}(\mathcal{G}). \quad (23)$$

So far we have not yet referred to the TAP equations in the DFT analysis. Instead we have considered a somewhat general dynamical system with disorder and memory. Such a formulation gives us enough freedom to construct a convenient dynamics which asymptotically converges to the solution of the TAP equations. We will define the dynamics to be of the form

$$\mathbf{m}(t+1) = \tanh(\psi(t)) \quad (24)$$

where the variables  $\psi_i(t)$  must be chosen to become independent Gaussian fields in the resulting effective single variable dynamics (19). In fact, there are actually various methods for doing so. In the following we will limit our attention to a method that we call the *single-step memory* construction.

#### 4. The single-step memory construction

In the single-step memory algorithm we will construct the update in such a way that the resulting memory term (23) satisfies the equation

$$\hat{\mathcal{G}}(t, \tau) = 0, \forall \tau \neq t-1. \quad (25)$$

Hence, if (25) holds, then using (19) we find that the variable

$$\gamma(t) - \hat{\mathcal{G}}(t, t-1)\mathbf{m}(t-1) = \phi(t) + \mathbf{h} \quad (26)$$

becomes a Gaussian field. We will choose the field  $\psi(t)$  in (24) as a linear combination of the Gaussian fields  $\phi(\tau)$ ,  $\tau = 1, \dots, t$  of the form

$$\psi(t) = \sum_{\tau=0}^t \mathcal{A}(t+1, \tau)(\phi(\tau) + \mathbf{h}) \quad (27)$$

where we have to construct the non-random terms  $\mathcal{A}(t+1, \tau)$  to make the dynamical order parameters consistent with the single-step memory condition (25). This condition leads to a very simple result for the response function (21) because there is no complicated propagation in the time of a response to an external field. In fact, from (24) we obtain for the response function (21)

$$\mathcal{G}(t, \tau) = \frac{1}{N} \sum_i \left\langle (1 - m_i^2(t)) \frac{\partial \psi_i(t-1)}{\partial \phi_i(\tau)} \right\rangle \quad (28)$$

$$= (1 - q(t)) \mathcal{A}(t, \tau) \quad (29)$$

with  $q(t) \triangleq \frac{1}{N} \mathbf{m}(t)^\dagger \mathbf{m}(t)$ . Thus we have the explicit result

$$\mathcal{A}(t, \tau) = \frac{\mathcal{G}(t, \tau)}{1 - q(t)}. \quad (30)$$

Finally, using (23) we obtain an explicit result for the response function in terms of the memory terms  $\hat{\mathcal{G}}(t, t-1)$ . Note that by construction of the single-step memory matrix  $\hat{\mathcal{G}}$  (25) we can write (23) as

$$\mathcal{G}(t, \tau) = a_{t-\tau} \prod_{s=\tau+1}^t \hat{\mathcal{G}}(s, s-1), \quad (31)$$

where the coefficients  $a_n$  are obtained from the power series expansion of the composition inverse of the R-transform:

$$\mathbf{R}^{-1}(x) = \sum_{n=1}^{\infty} a_n x^n. \quad (32)$$

By definition the trace of  $\mathbf{J}$  is zero, i.e.  $\mathbf{R}(0) = 0$ . Hence, the power series expansion in (32) starts from the first order term.

To complete the specification of the single-step memory construction we only need to specify  $\hat{\mathcal{G}}(t, t-1)$ . This will be chosen such that the method is asymptotically consistent



with the static TAP equations. Specifically, from (12) we should have

$$\lim_{t \rightarrow \infty} \hat{\mathcal{G}}(t, t-1) = R(1-q) \quad (33)$$

We choose the explicit form

$$\hat{\mathcal{G}}(t, t-1) = \frac{1-q(t)}{1-q(t-1)} R(1-q(t-1)), \quad (34)$$

which assuming convergence  $q(t) \rightarrow q$  as  $t \rightarrow \infty$  leads to (33). This form has also the advantage that in (30), for the update an unwanted factor  $1-q(t+1)$ , which would make  $\psi(t)$  depend on the future state  $\mathbf{m}(t+1)$ , cancels.

#### 4.1. Summary

Putting everything together the single-step memory algorithm for  $t \geq 0$  is defined as

$$\mathbf{m}(t+1) = \tanh(\psi(t)) \quad (35)$$

$$\psi(t) = Q(t) \sum_{\tau=0}^t a_{t+1-\tau} \mathbf{u}(\tau) \quad (36)$$

$$\mathbf{u}(t) = \frac{\mathbf{h} + \mathbf{J}\mathbf{m}(t) - \hat{\mathcal{G}}(t, t-1)\mathbf{m}(t-1)}{Q(t-1)(1-q(t))} \quad (37)$$

where we introduce

$$Q(t) = \prod_{\tau=0}^t R(1-q(\tau)) = Q(t-1)R(1-q(t)) \quad (38)$$

such that  $Q(-1) = 1$ . The memory term  $\hat{\mathcal{G}}(t, t-1)$  is given by (34). Moreover the algorithm initializes with  $\mathbf{m}(t) = 0$  for  $t \in \{-1, 0\}$ .

#### 4.2. Asymptotic consistency with TAP equations

In the following we show that if the single-step memory algorithm (35)–(37) converges, it solves the TAP equations (11)–(12). Let us assume that  $\mathbf{m}(t) \rightarrow \mathbf{m}$  as  $t$  tends to infinity. To have the convergence to the TAP equations we solely need to show that the sum in (36) converges to the proper limit. From (27) and (30) we must have

$$Q(t) \sum_{\tau=0}^t \frac{a_{t+1-\tau}}{(1-q(\tau))Q(\tau-1)} = \frac{1}{1-q(t+1)} \sum_{\tau=0}^t \mathcal{G}(t+1, \tau) \rightarrow 1. \quad (39)$$

We make the so-called weak long-term response assumption [22] that

$$\lim_{t \rightarrow \infty} \mathcal{G}(t, \tau) = 0, \quad \forall \text{ finite } \tau. \quad (40)$$

Hence, for sufficiently large  $t$  and  $\tau' < t$  such that  $t/\tau'$  being finite as  $t \rightarrow \infty$ , we can write

$$\sum_{\tau=0}^t \mathcal{G}(t+1, \tau) \simeq \sum_{\tau=\tau'}^t \mathcal{G}(t+1, \tau) \quad (41)$$

$$\simeq \sum_{\tau=\tau'}^t a_{t+1-\tau} R(1-q)^{t+1-\tau} \quad (42)$$

$$\simeq \sum_{n=1}^{\infty} a_n R(1 - q)^n \quad (43)$$

$$= R^{-1}(R(1 - q)) = 1 - q. \quad (44)$$

Next we will provide the details of the single-step memory algorithm for the SK, Hopfield and random orthogonal models.

#### 4.3. Example 1: the SK-model

Recall that for the standard SK model we have  $R(x) = \beta^2 x$ ; so that  $R^{-1}(x) = x/\beta^2$ . Hence,  $a_1 = 1/\beta^2$  and  $a_n = 0$  for  $n > 1$ . Thus, the single-step memory algorithm may written as

$$\mathbf{m}(t + 1) = \tanh(\psi(t)) \quad (45)$$

$$\psi(t) = \mathbf{h} + \mathbf{J}\mathbf{m}(t) - \beta^2(1 - q(t))\mathbf{m}(t - 1). \quad (46)$$

At first glance, these dynamical equations are similar but not exactly equal to those proposed by Bolthausen [8]. The difference is that instead of the dynamical order parameter  $q(t)$  the fixed point solution of  $q$  appears. Using the explicit form of the covariance of the field  $\psi_i(t)$  given by (60) in the next section, one finds for the field variance  $\langle (\psi_i(t) - \langle \psi_i(t) \rangle)^2 \rangle = \beta^2 q(t)$ . Hence, if we start the iteration (as in [8]) with  $m_i(1) = \sqrt{q}$  such that  $q(1) = q$ , then we find that in the large  $N$  limit, we also have  $q(t) = \langle \tanh^2(\psi_i(t - 1)) \rangle = q$  for all times  $t$  and we get agreement with [8].

#### 4.4. Example 2: the Hopfield model

For the Hopfield model from (9) we have

$$R^{-1}(x) = \frac{1}{\beta\alpha} \frac{x}{\beta - x}. \quad (47)$$

Thus the memory coefficients are given as  $a_n = 1/(\alpha\beta^{n+1})$  for  $n \geq 1$ . In the following we show that the single-step memory algorithm for the Hopfield model coincides with the AMP algorithm which was introduced in the context of the CDMA problem in [2] and compressed sensing in [9]. From (36) we first write

$$\psi(t) = \frac{Q(t)}{\alpha\beta^2} \mathbf{u}(t) + \frac{Q(t)}{\alpha\beta^{t+2}} \sum_{\tau=0}^{t-1} \beta^\tau \mathbf{u}(\tau) \quad (48)$$

$$= \frac{Q(t)}{\alpha\beta^2} \mathbf{u}(t) + \frac{1}{\beta} R(1 - q(t)) \psi(t - 1). \quad (49)$$

For convenience let us introduce

$$A(t) \triangleq \frac{R(1 - q(t))}{\beta\alpha(1 - q(t))} = \frac{\beta}{1 + \beta\alpha(1 - q(t))}. \quad (50)$$

Note that from (37) we may write (49) in the form of

$$\psi(t) = \frac{1}{\beta} A(t) [\mathbf{h} + \mathbf{J}\mathbf{m}(t)] + \alpha(1 - q(t)) A(t) [\psi(t - 1) - A(t - 1) \mathbf{m}(t - 1)]. \quad (51)$$

Then, defining  $z(t) \triangleq \psi(t) - A(t)\mathbf{m}(t)$ , we write the single-step memory algorithm as

$$\mathbf{m}(t+1) = \tanh(z(t) + A(t)\mathbf{m}(t)) \quad (52)$$

$$z(t) = \frac{1}{\beta}A(t)[\mathbf{h} + (\mathbf{J} - \beta\mathbf{I})\mathbf{m}(t)] + \alpha(1 - q(t))A(t)z(t-1) \quad (53)$$

where  $\mathbf{I}$  is the identity matrix of appropriate dimension. Note that  $(\mathbf{I} - \mathbf{J}/\beta)$  asymptotically coincides with the corresponding central Wishart matrix (see section 2). Thereby we exactly obtain the AMP iteration steps as introduced in [2]. We also refer the reader to the related works [3] and [21], where the dynamics of the AMP algorithm is analyzed by means of DFT and Bolthausen's conditioning technique [8], respectively.

Bolthausen's conditioning technique for the SK model [8] and Hopfield model [3] is based on the assumption that the entries of the underlying coupling matrix are defined via zero-mean iid and *Gaussian distributed* random variables, see section 2.1. Recently, it has been shown in [11] that the same analyses can be obtained without the need for the Gaussian distribution assumption but a sub-Gaussian tail condition of the distribution is required. Indeed, thanks to the central limit theorem, one can show that the generating function (3) also yields the same result regardless of whether the Gaussian distribution assumption is considered or not, see [23 section 5].

#### 4.5. Example 3: the random orthogonal model

For the random orthogonal model from (10) we have  $R^{-1}(x) = x/(\beta^2 - x^2)$ . This yields the memory coefficients as

$$a_n = \begin{cases} \frac{1}{\beta^{n+1}} & n \text{ is odd} \\ 0 & n \text{ is even.} \end{cases} \quad (54)$$

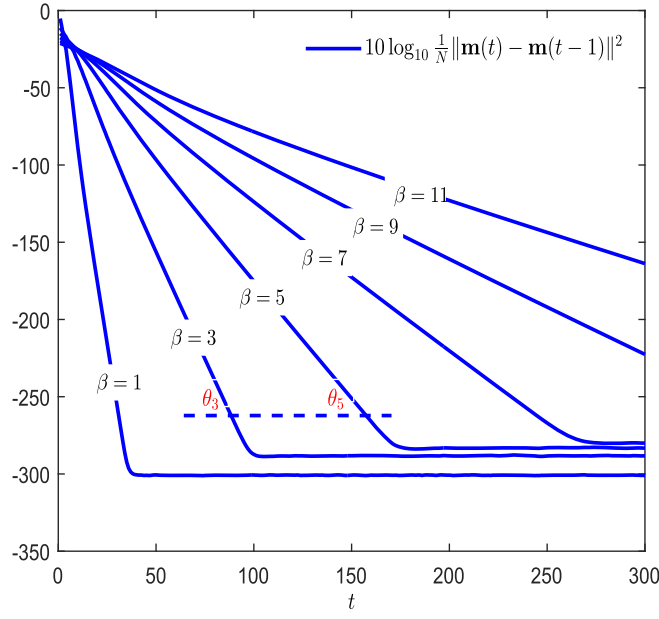
In figures 1 and 2 we illustrate the convergence of the single-step memory algorithm for the random orthogonal model obtained by running simulations. Note that after a few iteration steps the convergence becomes exponentially fast. The flat lines around  $(-300)$  dB, i.e.  $10^{-30}$ , are the consequence of the machine precision of the computer which was used.

The convergence improves with increasing temperature parameter  $1/\beta$ . On the other hand, for large enough  $\beta$ , the algorithm fails to converge. To estimate the critical parameter, we study the inverse decay time measured by the angle  $\theta$  as illustrated in figure 1 and extrapolate the simulational data to  $\theta = 0$  using a convenient range of  $\beta$ . One might expect that the critical  $\beta$  would coincide with the one obtained from an AT stability condition which can also be derived from the TAP approach [4]. The AT line is given by the equation

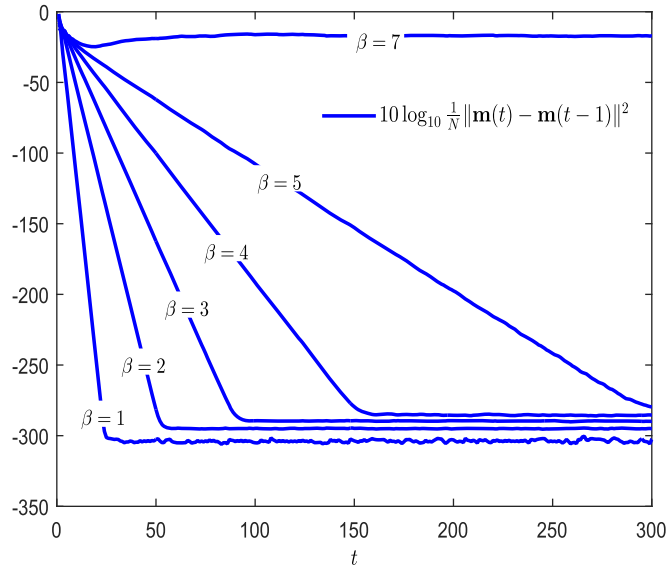
$$\alpha R'(1 - q) = 1 \quad \text{with} \quad \alpha \triangleq \frac{1}{N} \sum_{i=1}^N \langle (1 - \tanh^2(\psi_i))^2 \rangle \quad (55)$$

where the random variable  $\psi_i$  is a Gaussian with mean  $h_i$  and variance  $qR'(1 - q)$ . Note, that [4] contains a typo in the corresponding expression. In figure 3,

we present a comparison between the simulations and (55). This coincidence can be understood from a dynamical point of view by analyzing the stability of the dynamics close to the fixed point. The details will be postponed to section 5. Finally, it is also worth noting that the trajectories of the algorithm show a self-averaging behavior above the AT line for large  $N$ . On the other hand, we find that, below the AT line, there are strong sample to sample

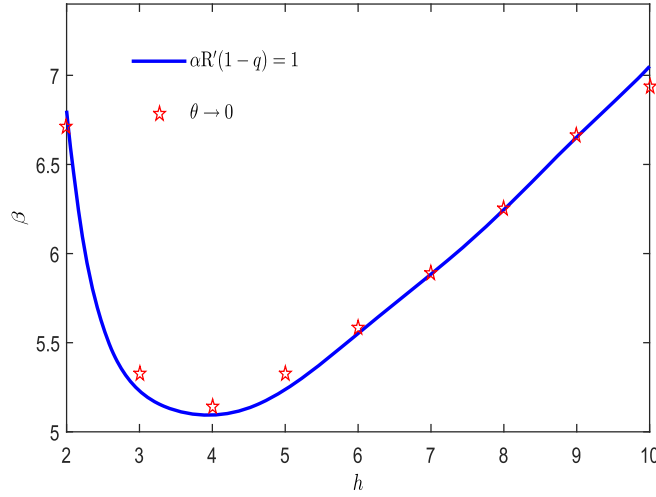


**Figure 1.** Random orthogonal model with  $\beta \in \{1, 3, \dots, 9, 11\}$ ,  $h_i = 1$  and  $N = 2^{14}$ . Here e.g.  $\theta \in \{\theta_3, \theta_5\}$  substitutes the respective linear decay time (in the log-domain), i.e. the convergence is faster the larger  $\theta$  is.



**Figure 2.** Random orthogonal model with  $\beta \in \{1, \dots, 5, 7\}$ ,  $h_i = 2$  and  $N = 2^{14}$ . The inverse temperature  $\beta = 6.9$  gives the AT line.

fluctuations. However, by averaging order parameters over many samples, we obtain a good agreement with the theory. Since the main goal of this paper is to present a convergent algorithm we will leave a more careful investigation of this point to future publications.



**Figure 3.** Consistency of the diverging decay time with the AT line: the simulations were based on the average over ten realizations of  $\mathbf{J}$  with  $N = 2^{14}$ .

#### 4.6. The field covariance matrix

In order to compare simulations of systems with analytical results obtained from the dynamical functional approach in the limit  $N \rightarrow \infty$  and to study the stability of TAP fixed points we have to perform expectations over the Gaussian random variables  $\psi_i(t)$  (see (27)). Specifically we write

$$\frac{1}{N} \langle \|\mathbf{m}(t) - \mathbf{m}(t-1)\|^2 \rangle_{\mathbf{J}} = q(t) + q(t-1) - 2\mathcal{C}(t, t-1) \quad (56)$$

where the order parameter  $\mathcal{C}$  is given by

$$\mathcal{C}(t+1, t'+1) = \frac{1}{N} \sum_{i=1}^N \int d\mathbf{P}(x, y) \tanh(\langle \psi_i(t) \rangle + x) \tanh(\langle \psi_i(t') \rangle + y). \quad (57)$$

Here  $\mathbf{P}$  denotes a two-dimensional Gaussian distribution with zero mean. The mean of the field  $\psi_i(t)$  follows from (27) as

$$\langle \psi_i(t) \rangle = \left\{ Q(t) \sum_{\tau=0}^t \frac{a_{t+1-\tau}}{(1-q(\tau))Q(\tau-1)} \right\} h_i. \quad (58)$$

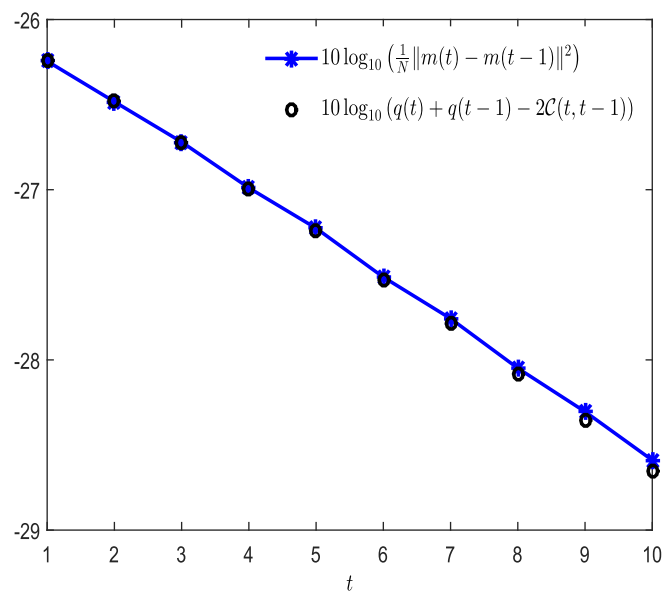
Hence, we need to compute the corresponding covariance matrix which is defined as

$$\mathcal{C}_{\psi}(t, t') = \langle (\psi_i(t) - \langle \psi_i(t) \rangle)(\psi_i(t') - \langle \psi_i(t') \rangle) \rangle. \quad (59)$$

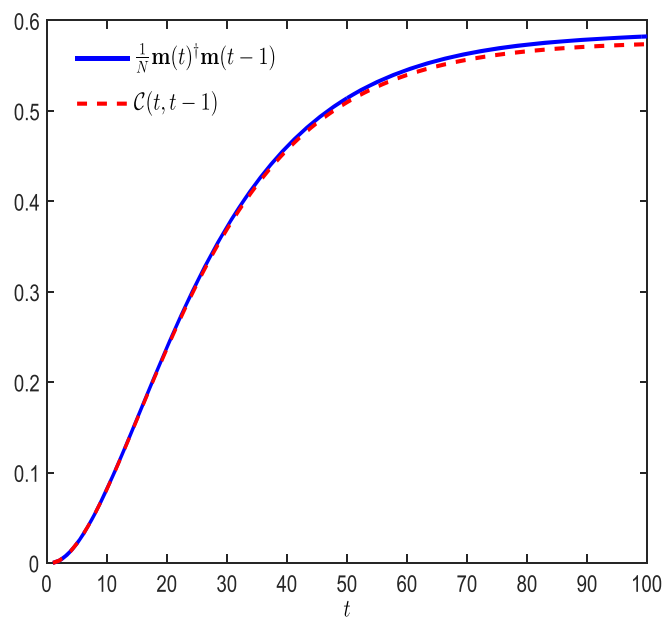
In appendix D we derive the expression

$$\mathcal{C}_{\psi}(t, t') = Q(t)Q(t') \sum_{l \leq t, m \leq t'} \frac{\text{Co}_{x^{t+1-l}y^{t'+1-m}}[A(x, y)]\mathcal{C}(l, m)}{(1-q(l))(1-q(m))Q(l-1)Q(m-1)}. \quad (60)$$

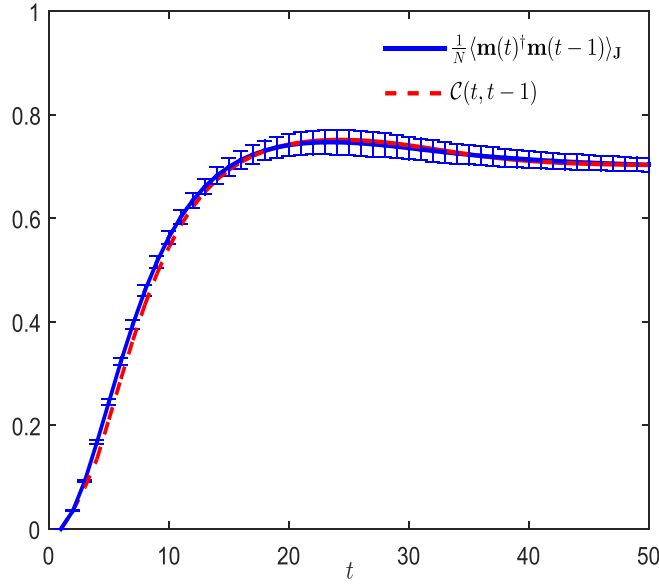
In this expression, for a power series  $f(x, y) = \sum_{n,k \geq 0} a_n b_k x^n y^k$ , we have introduced the symbol  $\text{Co}_{x^n y^k}[f(x, y)] \triangleq a_n b_k$  for its coefficients. Moreover the function  $A$  is defined as



**Figure 4.** Above the AT line: comparison of the theory and simulation for  $\beta = 20$  and  $h_i = 1$ ,  $N = 2^{14}$ .



**Figure 5.** Above the AT line: comparison of the theory and simulation for  $\beta = 20$  and  $h_i = 1$ ,  $N = 2^{14}$ .



**Figure 6.** Below the AT line: comparison of the theory and simulation for  $\beta = 10$  and  $h_i = 2$ ,  $N = 2^{12}$ .  $\langle \cdot \rangle_J$  is obtained by  $5 \times 10^3$  realizations of  $J$ .

$$A(x, y) \triangleq \left( \frac{1}{R^{-1}(x)} - \frac{1}{R^{-1}(y)} \right)^{-1} (y - x). \quad (61)$$

The function  $A$  has a relatively simple form for the three random matrix ensembles considered in this paper. For the SK and Hopfield models we have  $A(x, y) = xy/\beta^2$  and  $A(x, y) = xy/(\beta^2\alpha)$ , respectively. Moreover, for the random orthogonal model, from  $R^{-1}(x) = x/(\beta^2 - x^2)$  we have  $\text{Co}_{x^ny^k}[A(x, y)] = \delta_{nk}(-1)^{n+1}/\beta^{2n}$ . We next compare our simulations with theoretical results. We used the initializations  $\mathbf{m}(t) = 0$  for  $t \in \{0, 1\}$ , hence we assign  $\mathcal{C}(1, 0) = 0$ . In figures 4 and 5 we show such a comparison above the AT line. Note that no averaging over coupling matrices was used for the simulations. The integration over two-dimensional correlated Gaussian distribution used to calculate  $\mathcal{C}(t, t-1)$  was performed numerically. However the accuracy of the numerical method limits us from providing very precise results as  $t$  grows. In figure 4 we illustrate the theoretical prediction of  $\frac{1}{N}\mathbf{m}(t)^\dagger \mathbf{m}(t-1)$  by the order parameter  $\mathcal{C}(t, t-1)$  for a large range of  $t$ . Below the AT line, the single-step memory algorithm diverges and simulated trajectories show strong sample fluctuations. However, by taking an average over a large number of trajectories we obtain a good agreement with the theory (see figure 6).

#### 4.7. Asymptotic consistency with cavity variance

In section 4.2 we have demonstrated the convergence of the single-step memory algorithm to the TAP equations. In a similar way one can show that (60) converges to the variance of the static field variance in (13) as  $t$  and  $t'$  tend to infinity. Specifically, by invoking the weak long-term response assumption (40) in (20) and following similar steps as in appendix D, for sufficiently large  $t$ ,  $\tau < t$ ,  $t' < t$  and  $\tau' < t'$  such that  $\tau/t$  and  $\tau'/t'$  being finite as  $t$  and  $t'$  tend to infinity, one can show that

$$\mathcal{C}_\psi(t, t') \simeq \frac{q}{(1-q)^2} \sum_{\tau \leq t, \tau' \leq t'} \text{Co}_{x^{t+1-\tau}y^{t'+1-\tau'}}[A(x, y)]R(1-q)^{t+1-\tau}R(1-q)^{t'+1-\tau'} \quad (62)$$

$$\simeq \frac{q}{(1-q)^2} \sum_{n, k \leq 1} \text{Co}_{x^n y^k}[A(x, y)]R(1-q)^n R(1-q)^k \quad (63)$$

$$= \frac{q}{(1-q)^2} A(R(1-q), R(1-q)). \quad (64)$$

In this expression, by abuse of notation, we denote  $\lim_{y \rightarrow x} A(x, y)$  by  $A(x, x)$ . Its explicit form is given by

$$\lim_{y \rightarrow x} A(x, y) = (R^{-1}(x))^2 \frac{1}{(R^{-1}(x))'} = (R^{-1}(x))^2 R'(R^{-1}(x)). \quad (65)$$

Hence we have

$$\lim_{t, t' \rightarrow \infty} \mathcal{C}_\psi(t, t') = \frac{q}{(1-q)^2} (R^{-1}(R(1-q)))^2 R'(R^{-1}(R(1-q))) \quad (66)$$

$$= qR'(1-q). \quad (67)$$

## 5. The stability of TAP fixed points

In order to analyze the stability of the fixed points of the single-step algorithm we resort to a linear stability analysis. We add Gaussian white noise to the dynamics, i.e. we set  $\psi_i(t) \rightarrow \psi_i(t) + \epsilon_i(t)$  with  $\langle \epsilon_i(t)^2 \rangle = \epsilon$  and discuss the limit  $\epsilon \rightarrow 0$ . If the static TAP fixed point is stable, then the system should asymptotically show only small stationary fluctuations around this and we can work in the Fourier domain. Hence, we assume

$$\mathcal{C}(t, t') = \frac{1}{2\pi} \int d\omega \hat{\mathcal{C}}(\omega) e^{i\omega(t-t')} \quad (68)$$

$$\mathcal{C}_\psi(t, t') = \frac{1}{2\pi} \int d\omega \hat{\mathcal{C}}_\psi(\omega) e^{i\omega(t-t')}. \quad (69)$$

Inserting these Fourier representations into (60), for large  $t$  and  $t'$  we may write (see (62)–(64))

$$\hat{\mathcal{C}}_\psi(\omega) \simeq \sum_{l \leq t, m \leq t'} \frac{\text{Co}_{x^{l+1-l}y^{m+1-m}}[A(x, y)] \hat{\mathcal{C}}(\omega) e^{i\omega(l-t-(m-t'))}}{(1-q)^2 R(1-q)^{l-t-1} R(1-q)^{m-t'-1}} \quad (70)$$

$$\simeq \frac{A(e^{-i\omega} R(1-q), e^{i\omega} R(1-q))}{(1-q)^2} \hat{\mathcal{C}}(\omega). \quad (71)$$

For small noise  $\epsilon \rightarrow 0$ , the assumption of stability translates into small fluctuations around the static solution and we can write

$$\hat{\mathcal{C}}(\omega) \simeq 2\pi q \delta(\omega) + \epsilon \hat{c}(\omega) \quad (72)$$

$$\hat{\mathcal{C}}_\psi(\omega) \simeq 2\pi q R'(1-q) \delta(\omega) + \epsilon \hat{c}_\psi(\omega). \quad (73)$$

where we have separated fluctuations into static and dynamical parts. We will analyze the dynamical part next, but note that also the static part  $q$  will have contributions from  $\epsilon$ . Thus for  $\omega \neq 0$  we have



$$\hat{c}_\psi(\omega) = \hat{c}(\omega) \frac{A(e^{-i\omega}R(1-q), e^{i\omega}R(1-q))}{(1-q)^2}. \quad (74)$$

where now the value of  $q$  is computed for  $\epsilon = 0$ . We next express  $\hat{c}(\omega)$  in terms of  $\hat{c}_\psi(\omega)$  for small  $\epsilon$ . The calculation in [E](#) is based on expanding

$$\mathcal{C}(t+1, t'+1) = \frac{1}{N} \sum_i \langle \tanh(u(t) + h_i) \tanh(u(t') + h_i) \rangle_u \quad (75)$$

up to first order in  $\epsilon$ . The brackets denote expectations over the two-dimensional Gaussian field  $(u(t), u(t'))$  with  $\langle u(t)u(t') \rangle \simeq s_0 + \epsilon(s(t-t') + \delta_{t,t'})$  for  $\epsilon \rightarrow 0$ , where  $s_0 = qR'(1-q)$  and  $s(t-t') = c_\psi(t, t')$ . For  $t = t'$  the integral is over a single Gaussian only. The calculation shows that

$$\hat{c}(\omega) = \alpha(1 + \hat{c}_\psi(\omega)) \quad (76)$$

with  $\alpha$  is defined as in [\(55\)](#). Combining this relationship with [\(74\)](#) we have

$$\hat{c}(\omega) \simeq \alpha \left( 1 - \frac{\alpha A(e^{-i\omega}R(1-q), e^{i\omega}R(1-q))}{(1-q)^2} \right)^{-1}. \quad (77)$$

In fact, for the SK, Hopfield and random orthogonal models, we have  $\text{Co}_{x^ny^k}[A(x, y)] = 0$   $\forall n \neq k$ . Therefore [\(77\)](#) is actually independent of  $\omega$  and from [\(65\)](#) we explicitly have that

$$\hat{c}(\omega) = \frac{\alpha}{1 - \alpha R'(1-q)}. \quad (78)$$

In general, the right-hand side of [\(77\)](#) must be non-negative to have a valid representation as a Fourier-transform of a time dependent correlation function. While the term  $A(e^{-i\omega}R(1-q), e^{i\omega}R(1-q))$  is always positive, see [\(71\)](#), the second term is small and positive for sufficiently small  $\beta$ . But it changes sign and diverges. One expects that the divergence will occur first for the long range fluctuations, i.e. for the limit of low frequencies. Taking the limit yields

$$\lim_{\omega \rightarrow 0} \hat{c}(\omega) = \frac{\alpha}{1 - \alpha R'(1-q)}. \quad (79)$$

The condition

$$\alpha R'(1-q) = 1 \quad (80)$$

for the onset of instability agrees with the well-known AT stability criterion [\[4\]](#).

## 6. Discussion and outlook

In this paper we have presented a theoretical approach to the design of iterative algorithms for solving the TAP equations for Ising models with random couplings drawn from general invariant ensembles. We were guided by the idea that one needs to subtract terms from the internal field which depend on the values of the magnetizations at previous times. Using dynamical functional theory we have shown that in such a way, memory terms can be canceled and one arrives at a Gaussian distributed field, which eventually converges to the cavity field provided that a stability condition is fulfilled. We have presented a specific method which we have called the ‘single-step memory construction’. Our approach may be extended in several ways. For example other subtraction methods are possible. One might design an alternative scheme, where the response function is required to be zero after one time step leading to a somewhat different algorithm and we will give details elsewhere. It would be

interesting to see in which cases the explicit memory terms in the subtraction method can be simplified by introducing auxiliary variables as is possible for the Hopfield model. Other extensions of our method would be to more general probabilistic models beyond the simple Ising case. This would include continuous random variables and other forms of interactions. An application to models of compressed sensing would be interesting where certain random matrix ensembles (such as the random orthogonal ones) might be natural models for the observation matrix. Specifically one can trivially extend the random orthogonal ensemble by considering the more general spectrum such that the eigenvalues of  $\tilde{\mathbf{J}}$  are distributed as  $\alpha\delta(\lambda - 1) + (1 - \alpha)\delta(\lambda + 1)$ . In the context of compressed sensing this model coincides with the so-called random row-orthogonal ensemble [24, 25]. We will discuss details in a forthcoming publication. Finally, it would be important to address a drawback of our method which prevents an application to probabilistic inference problems with arbitrary data. Our subtraction scheme depends explicitly on the random matrix ensemble of couplings which may not be known in practice. Hence it would be interesting to develop schemes which adapt to the concrete data which would then achieve convergence to ‘adaptive TAP equations’ of [4], providing possible alternatives to the currently applied message passing algorithms [6].

## Acknowledgments

The authors would like to thank Florent Krzakala for inspiring us to do this study. This work was partially supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMunications NEWCOM# (Grant agreement no. 318306).

## Appendix A. The proof of $\tilde{J}_{ii} - \frac{1}{N}\text{tr}(\tilde{\mathbf{J}}) \rightarrow 0$

From (2) we have

$$\tilde{J}_{ii} = \sum_{n=1}^N \lambda_n [(\mathbf{O}_n \mathbf{O}_n^\dagger)]_{ii} = \sum_{n=1}^N \lambda_n O_{in}^2 \quad (\text{A.1})$$

where  $\mathbf{O}_n$  is the  $n$ th column vector of Haar (orthogonal)  $\mathbf{O}$  and  $\lambda_n = \Lambda_{nn}$ . Note that  $\Lambda$  is independent of  $\mathbf{O}$ . For convenience we treat the realizations of  $\Lambda$  from its probability space and denote as  $\Lambda(\omega)$ . We show the convergence for every realization  $\tilde{J}_{ii}(\omega) = \sum_{n=1}^N \lambda_n(\omega) O_{in}^2$ . The mean and variance of  $\tilde{J}_{ii}(\omega)$  are, respectively, given by

$$\langle \tilde{J}_{ii}(\omega) \rangle = \sum_{n=1}^N \lambda_n(\omega) \langle O_{in}^2 \rangle \quad (\text{A.2})$$

$$\text{Var}[\tilde{J}_{ii}(\omega)] = \sum_{n=1}^N \lambda_n^2(\omega) \text{Var}[O_{in}^2] + 2 \sum_{n < k} \lambda_n(\omega) \lambda_k(\omega) \text{Cov}[O_{in}^2, O_{ik}^2] \quad (\text{A.3})$$

where for the random variables  $X$  and  $Y$ ,  $\text{Var}[X]$  and  $\text{Cov}[X, Y]$  denoting the variance  $X$  and the covariance of  $X$  and  $Y$ , respectively. For the proof, we basically need to show that

$$\lim_{N \rightarrow \infty} \text{Var}[\tilde{J}_{ii}(\omega)] = 0. \quad (\text{A.4})$$

To do this we make use of the so-called (orthogonal) Weingarten calculus that allows to for calculate joint moments of Haar entries (analogous to Wick calculus for Gaussian matrices). For details we refer the reader to [26, 27]. From [27 theorem 2.1, example 2.1] we have

$$\langle O_{in}^2 O_{ik}^2 \rangle = \begin{cases} \frac{N+1}{N(N+2)(N-1)} & n \neq k \\ \frac{3}{N(N+2)} & n = k \end{cases}. \quad (\text{A.5})$$

Furthermore we have  $\langle O_{ij}^2 \rangle = 1/N$ ,  $\forall i, j$ . Thus, the variance (A.3) reads

$$\text{Var}[\tilde{J}_{ii}(\omega)] = \frac{2(N-1)}{N^2(N+2)} \sum_{n=1}^N \lambda_n^2(\omega) + \frac{4}{N^2(N+2)(N-1)} \sum_{n < k} \lambda_n(\omega) \lambda_k(\omega). \quad (\text{A.6})$$

Note that the spectrum  $\tilde{\mathbf{J}}$  is assumed to converge almost surely to a compactly supported probability distribution such that the smallest and largest eigenvalue of  $\tilde{\mathbf{J}}$  converge (almost surely) to the infimum and supremum of the compact support, respectively. This implies that the minimum and maximum of the eigenvalues of  $\tilde{\mathbf{J}}$  are uniformly bounded above for a sufficiently large  $N$ . This is sufficient to get (A.4) from (A.6). Thereby we complete the proof.

## Appendix B. The self-averaging limit of adaptive TAP equations

We will provide a derivation of the TAP equations (11)–(12) from the ‘adaptive TAP’ approach of [4]. Under the assumption of Gaussian distributed cavity fields and an approximate linear response argument one finds

$$m_i = \tanh \left( h_i + \sum_j J_{ij} m_j - V_i m_i \right) \quad (\text{B.1})$$

$$V_i = \Lambda_{ii} - \frac{1}{\chi_{ii}} \quad (\text{B.2})$$

$$\Lambda_{ii} = V_i + \frac{1}{1 - m_i^2} \quad (\text{B.3})$$

with the positive definite matrix  $\chi \triangleq (\mathbf{\Lambda} - \mathbf{J})^{-1}$ .

To obtain the TAP equations in (11)–(12) we basically need to show that  $V_i \simeq \mathbb{R}(1 - q)$ . To that end we write (B.2) in the form of

$$V_i = \Lambda_{ii} - \left( \frac{\partial \ln \det(\mathbf{\Lambda} - \mathbf{J})}{\partial \Lambda_{ii}} \right)^{-1}. \quad (\text{B.4})$$

Our basic idea is to simplify  $\ln \det(\mathbf{\Lambda} - \mathbf{J})$  using the results of *free probability theory* for random matrices. To that end we invoke an additional assumption that the empirical distribution function of  $\{\Lambda_{11}, \dots, \Lambda_{NN}\} \in \mathbb{R}^N$  converges weakly and almost surely to a compactly supported probability distribution as  $N \rightarrow \infty$ . Since  $\mathbf{J}$  is asymptotically invariant and has a compactly supported limiting spectrum,  $\mathbf{\Lambda}$  and  $\mathbf{J}$  are asymptotically (almost surely) free [26]. Thus, due to the uniform convergence property of the R-transform, see [17 lemma 3.3.4], for a sufficiently large  $N$  we have

$$\mathbf{R}_{\mathbf{\Lambda} - \mathbf{J}}^N(x) \simeq \mathbf{R}_{\mathbf{\Lambda}}^N(x) + \mathbf{R}_{-\mathbf{J}}^N(x) \quad (\text{B.5})$$

$$= \mathbf{R}_{\mathbf{\Lambda}}^N(x) - \mathbf{R}_{\mathbf{J}}^N(-x) \quad (\text{B.6})$$

where we denote the R-transform of the spectrum of an  $N \times N$  symmetric matrix  $\mathbf{X}$  by  $\mathbf{R}_{\mathbf{X}}^N$ . Note that we have  $1 - q = \frac{1}{N} \text{tr}(\mathbf{\Lambda} - \mathbf{J})^{-1}$ . Then, from lemma 1 below we have

$$\frac{1}{N} \ln \det(\Lambda - J) = -(1 + \ln(1 - q)) + \int_0^{1-q} d\omega R_{\Lambda-J}^N(-\omega) \quad (\text{B.7})$$

$$\simeq -(1 + \ln(1 - q)) + \int_0^{1-q} d\omega R_{\Lambda}^N(-\omega) - \int_0^{1-q} d\omega R_J^N(\omega) \quad (\text{B.8})$$

$$= \frac{1}{N} \ln \det(\Lambda - V\mathbf{I}) + (1 - q)V - \int_0^{1-q} d\omega R_J^N(\omega) \quad (\text{B.9})$$

where  $V$  is defined through the implicit equation  $(1 - q) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\Lambda_{ii} - V}$ . This implies that  $\text{tr}(\Lambda - J)^{-1} \simeq \text{tr}(\Lambda - V\mathbf{I})^{-1}$ . Here from (B.8) to (B.9) we make use of the identity (B.19) below. Note that for a non-negative  $N \times N$  matrix  $X$  we can write the Stieltjes transform of the eigenvalue distribution of  $X$ , say  $P_X^N$ , as  $M_X^N(\omega) = \int dP_X^N(x) / (\omega - x)$  with  $\omega \in (-\infty, 0)$ . Then from (B.5) we use the subordination property [28, Chapter 22] as  $M_{\Lambda-J}^N(\omega) \simeq M_{\Lambda}^N(\omega + R_J^N(-M_{\Lambda-J}^N(\omega)))$  and take the limit  $\omega \rightarrow 0$ . Doing so yields  $V \simeq R(1 - q)$  where we note that  $R(x) = \lim_{N \rightarrow \infty} R_J^N(x)$ . We also note that  $\frac{\partial q}{\partial \Lambda_{ii}} = O(\frac{1}{N})$ . Hence for large  $N$  we can write

$$\frac{\partial \ln \det(\Lambda - J)}{\partial \Lambda_{ii}} \simeq \frac{\partial \ln \det(\Lambda - V\mathbf{I})}{\partial \Lambda_{ii}} = \frac{1}{\Lambda_{ii} - V} \quad (\text{B.10})$$

which yields  $V_i \simeq V = R(1 - q)$ .

**Lemma 1.** *Let an  $N \times N$  matrix  $X$  be positive definite. Let  $Q = \frac{1}{N} \text{tr}(X^{-1})$ . Then,*

$$\frac{1}{N} \ln \det(X) = -(1 + \ln Q) + \int_0^Q d\omega R_X^N(-\omega). \quad (\text{B.11})$$

**Proof.** Note that

$$\ln \det(X) = \lim_{\epsilon \rightarrow \infty} \ln \det(\epsilon^{-1}\mathbf{I} + X). \quad (\text{B.12})$$

For convenience let  $\eta(\epsilon) \triangleq \frac{1}{N} \text{tr}((\mathbf{I} + \epsilon X)^{-1})$  for  $\epsilon > 0$ . Since  $X$  is positive definite we can write [29]

$$R_X^N(-\epsilon\eta(\epsilon)) = \frac{1 - \eta(\epsilon)}{\epsilon\eta(\epsilon)}. \quad (\text{B.13})$$

Applying the substitution  $\omega \triangleq t\eta(t)$  to the following integral we have

$$\int_0^{Q(\epsilon)} d\omega R_X^N(-\omega) = \int_0^\epsilon dt \frac{\eta(t) + t\eta'(t)}{t\eta(t)} [1 - \eta(t)] \quad (\text{B.14})$$

$$= \int_0^\epsilon dt \frac{\eta'(t)}{\eta(t)} (1 - \eta(t)) + \int_0^\epsilon dt \frac{1 - \eta(t)}{t} \quad (\text{B.15})$$

$$= \ln \eta(\epsilon) + 1 - \eta(\epsilon) + \int_0^\epsilon dt \frac{1 - \eta(t)}{t} \quad (\text{B.16})$$

$$= \ln \eta(\epsilon) + 1 - \eta(\epsilon) + \frac{1}{N} \ln \det(\mathbf{I} + \epsilon X) \quad (\text{B.17})$$

$$= \ln Q(\epsilon) + 1 - \eta(\epsilon) + \frac{1}{N} \ln \det(\epsilon^{-1} \mathbf{I} + \mathbf{X}) \quad (\text{B.18})$$

with  $Q(\epsilon) \triangleq e\eta(\epsilon)$ . In other words we have

$$\frac{1}{N} \ln \det(\epsilon^{-1} \mathbf{I} + \mathbf{X}) = (\eta(\epsilon) - 1 - \ln Q(\epsilon)) + \int_0^{Q(\epsilon)} d\omega \mathbf{R}_X^N(-\omega). \quad (\text{B.19})$$

Taking the limit  $\epsilon \rightarrow \infty$  we complete the proof.

### Appendix C. Derivation of DFT results

For the sake of notational compactness let

$$g(\{\mathbf{m}(\tau), \gamma(\tau)\}_{\tau=0}^t) \triangleq \delta(\mathbf{m}(t) - f(\{\gamma(\tau), \mathbf{m}(\tau)\}_{\tau=0}^{t-1})). \quad (\text{C.1})$$

By the Fourier representation of the Dirac function we write

$$Z(\{\mathbf{l}(t)\}) = \int \prod_{t=0}^{T-1} d\mathbf{m}(t) d\gamma(t) d\hat{\gamma}(t) g(\{\mathbf{m}(\tau), \gamma(\tau)\}_{\tau=0}^t) e^{i\hat{\gamma}(t)^\dagger (\gamma(t) - \mathbf{h} - \mathbf{J}\mathbf{m}(t))} e^{i\gamma(t)^\dagger \mathbf{l}(t)}. \quad (\text{C.2})$$

The derivation is separated into two parts: (i) the disorder average and (ii) the saddle-point method.

#### C.1. Disorder average

For convenience let us introduce  $N \times T$  matrices  $\mathbf{X}$  and  $\hat{\mathbf{X}}$  with  $X_{nt} = \frac{m_n(t+1)}{\sqrt{N}}$  and  $\hat{X}_{nt} = \frac{\hat{\gamma}_n(t+1)}{i\sqrt{N}}$ . We need to evaluate

$$\left\langle e^{-\frac{i}{2} \sum_t \{\hat{\gamma}(t)^\dagger \mathbf{J} \mathbf{m}(t) + \mathbf{m}(t)^\dagger \mathbf{J} \hat{\gamma}(t)\}} \right\rangle_{\mathbf{J}} \simeq e^{\frac{N}{2} \sum_{n \geq 1} \frac{c_n}{n} \text{tr}(\mathbf{Q}^n)} \quad (\text{C.3})$$

with  $\mathbf{Q} = \hat{\mathbf{X}}\mathbf{X}^\dagger + \mathbf{X}\hat{\mathbf{X}}^\dagger$ . Here (C.3) follows directly from (3)–(7). We will evaluate

$$\text{tr}(\mathbf{Q}^n) = \text{tr}((\hat{\mathbf{X}}\mathbf{X}^\dagger + \mathbf{X}\hat{\mathbf{X}}^\dagger)^n). \quad (\text{C.4})$$

in terms of the matrices (C.4)

$$\mathcal{G} \triangleq \mathbf{X}^\dagger \hat{\mathbf{X}} \quad (\text{C.5})$$

$$\mathcal{C} \triangleq \mathbf{X}^\dagger \mathbf{X} \quad (\text{C.6})$$

$$\tilde{\mathcal{C}} \triangleq \hat{\mathbf{X}}^\dagger \hat{\mathbf{X}} \quad (\text{C.7})$$

Then by using cyclic invariance of the trace we obtain the expression

$$\text{tr}(\mathbf{Q}^n) = 2\text{tr}(\mathcal{G}^n) + n \text{tr} \sum_{k=0}^{n-2} \{\mathcal{G}^k \mathcal{C} (\mathcal{G}^\dagger)^{n-2-k} \tilde{\mathcal{C}}\} + \text{I}(\mathcal{G}, \mathcal{C}, \tilde{\mathcal{C}}), \quad (\text{C.8})$$

where the function I satisfies

$$\left. \frac{\partial \text{I}(\mathcal{G}, \mathcal{C}, \tilde{\mathcal{C}})}{\partial \tilde{\mathcal{C}}} \right|_{\tilde{\mathcal{C}}=0} = 0. \quad (\text{C.9})$$

This means that I contains more than one factor  $\tilde{\mathcal{C}}$  and will thus at the saddle-point value  $\tilde{\mathcal{C}} = 0$  not contribute to saddle-point equations.

### C.2. The saddle-point calculation

We write as

$$\begin{aligned} \langle Z(\{I(t)\}) \rangle_J &\simeq \int d\mathcal{G} d\mathcal{C} d\tilde{\mathcal{C}} e^{\frac{N}{2} \sum_{n \geq 1} \frac{c_n}{n} \left( 2\text{tr}(\mathcal{G}^n) + n\text{tr} \sum_{k=0}^{n-2} \{ \mathcal{G}^k \mathcal{C} (\mathcal{G}^\dagger)^{n-2-k} \tilde{\mathcal{C}} \} + I(\mathcal{G}, \mathcal{C}, \tilde{\mathcal{C}}) \right)} \\ &\int \prod_{t=0}^{T-1} \{ d\mathbf{m}(t) d\boldsymbol{\gamma}(t) d\hat{\boldsymbol{\gamma}}(t) g(\{\mathbf{m}(\tau), \boldsymbol{\gamma}(\tau)\}_{\tau \leq t}) e^{i\hat{\boldsymbol{\gamma}}(t)^\dagger (\boldsymbol{\gamma}(t) - \mathbf{h})} e^{i\boldsymbol{\gamma}(t)^\dagger I(t)} \} \\ &\prod_{t,s} \delta(iN\mathcal{G}(t,s) - \mathbf{m}(t)^\dagger \hat{\boldsymbol{\gamma}}(s)) \delta(N\mathcal{C}(t,s) - \mathbf{m}(t)^\dagger \mathbf{m}(s)) \delta(N\tilde{\mathcal{C}}(t,s) + \hat{\boldsymbol{\gamma}}(t)^\dagger \hat{\boldsymbol{\gamma}}(s)). \end{aligned} \quad (\text{C.10})$$

By the Fourier representation of Dirac function we write the last line of (C.10) as

$$c \int d\hat{\mathcal{G}} d\hat{\mathcal{C}} d\hat{\tilde{\mathcal{C}}} e^{\sum_{(t,s)} i\hat{\mathcal{G}}(t,s)(iN\mathcal{G}(t,s) - \mathbf{m}(t)^\dagger \hat{\boldsymbol{\gamma}}(s)) + i\hat{\mathcal{C}}(t,s)(N\mathcal{C}(t,s) - \mathbf{m}(t)^\dagger \mathbf{m}(s)) - \hat{\tilde{\mathcal{C}}}(t,s)(N\tilde{\mathcal{C}}(t,s) + \hat{\boldsymbol{\gamma}}(t)^\dagger \hat{\boldsymbol{\gamma}}(s))}. \quad (\text{C.11})$$

Here  $c$  is a constant irrelevant for the saddle-point calculation. We define the auxiliary single-site partition function

$$\begin{aligned} \tilde{Z}_n(l_n, \hat{\mathcal{G}}, \hat{\mathcal{C}}, \hat{\tilde{\mathcal{C}}}) &\triangleq \int \prod_{t=0}^{T-1} \{ dm_n(t) d\gamma_n(t) d\hat{\gamma}_n(t) g(\{m_n(\tau), \gamma_n(\tau)\}_{\tau \leq t}) \\ &e^{i\hat{\gamma}_n(t)^\dagger (\gamma_n(t) - h_n)} e^{i\gamma_n(t)^\dagger l_n(t)} \} \\ &e^{-\sum_{(t,s)} [i\hat{\mathcal{G}}(t,s)m_n(t)\hat{\gamma}_n(s) + i\hat{\mathcal{C}}(t,s)m_n(t)m_n(s) + \hat{\tilde{\mathcal{C}}}(t,s)\hat{\gamma}_n(t)\hat{\gamma}_n(s)]}. \end{aligned} \quad (\text{C.12})$$

In this way we can write (C.10) as

$$\begin{aligned} \langle Z(\{I(t)\}) \rangle_J &\simeq \int d\mathcal{G} d\hat{\mathcal{G}} d\mathcal{C} d\hat{\mathcal{C}} d\tilde{\mathcal{C}} d\hat{\tilde{\mathcal{C}}} e^{\frac{N}{2} \sum_{n \geq 1} \frac{c_n}{n} \left( 2\text{tr}(\mathcal{G}^n) + n\text{tr} \sum_{k=0}^{n-2} \{ \mathcal{G}^k \mathcal{C} (\mathcal{G}^\dagger)^{n-2-k} \tilde{\mathcal{C}} \} + I(\mathcal{G}, \mathcal{C}, \tilde{\mathcal{C}}) \right)} \\ &e^{N \sum_{(t,s)} [-\hat{\mathcal{G}}(t,s)\mathcal{G}_{t,s} + i\hat{\mathcal{C}}(t,s)\mathcal{C}(t,s) - \hat{\tilde{\mathcal{C}}}(t,s)\tilde{\mathcal{C}}(t,s)] + \sum_n \log \tilde{Z}_n(l_n, \hat{\mathcal{G}}, \hat{\mathcal{C}}, \hat{\tilde{\mathcal{C}}})}. \end{aligned} \quad (\text{C.13})$$

In the large  $N$  limit we can perform the integrations over  $\mathcal{G}$ ,  $\hat{\mathcal{G}}$ ,  $\mathcal{C}$ ,  $\hat{\mathcal{C}}$ ,  $\tilde{\mathcal{C}}$ ,  $\hat{\tilde{\mathcal{C}}}$  with the saddle-point methods. Doing so yields:

$$\mathcal{G}(t, s) = -\frac{i}{N} \sum_n \langle m_n(t) \hat{\gamma}_n(s) \rangle_{\tilde{Z}_n} \quad (\text{C.14})$$

$$\mathcal{C}(t, s) = \frac{1}{N} \sum_n \langle m_n(t) m_n(s) \rangle_{\tilde{Z}_n} \quad (\text{C.15})$$

$$\tilde{\mathcal{C}}(t, s) = -\frac{1}{N} \sum_n \langle \hat{\gamma}_n(t) \hat{\gamma}_n(s) \rangle_{\tilde{Z}_n}. \quad (\text{C.16})$$

with  $\langle \rangle_{\tilde{Z}_n}$  denoting the average with respect to the single-site partition function. Here the quantity (C.16) has only the trivial solution that  $\tilde{\mathcal{C}}_{t,s} = 0$ . Other solutions may violate the normalization  $Z(\{I(t) = 0\}) = 1$ . Furthermore this solution leads to  $\hat{\mathcal{C}} = 0$ . By invoking (C.8) we have

$$\hat{\mathcal{G}} = R(\mathcal{G}) \quad (\text{C.17})$$

$$\hat{\mathcal{C}} = \frac{1}{2} \sum_{n=1}^{\infty} c_n \sum_{k=0}^{n-2} \mathcal{G}^k \mathcal{C}(\mathcal{G}^\dagger)^{n-2-k}. \quad (\text{C.18})$$

Thus, for large  $N$  we get the factorization of the generating function

$$\begin{aligned} \langle Z(\{I(t)\}) \rangle_J &\simeq \prod_{n=1}^N \int \prod_{t=0}^{T-1} \{dm_n(t) d\gamma_n(t) \mathbf{g}(\{m_n(\tau), \gamma_n(\tau)\}_{\tau \leq t}) e^{i\gamma_n(t) l_n(t)} \\ &\times e^{i\hat{\gamma}_n(t)(\gamma_n(t) - h_n - \sum_s \hat{\mathcal{G}}(t,s)m_n(s)) - \sum_s \hat{\mathcal{C}}(t,s)\hat{\gamma}_n(t)\hat{\gamma}_n(s)} \}. \end{aligned} \quad (\text{C.19})$$

We linearize the quadratic terms in  $\hat{\gamma}_n(t)$  by introducing auxiliary Gaussian random fields  $\phi_n(t)$  which are iid for each  $n$  with zero mean and covariance  $\mathcal{C}_\phi(t, s) \triangleq 2\hat{\mathcal{C}}(t, s) \triangleq \langle \phi_n(t) \phi_n(s) \rangle$ . In this way we can write

$$e^{-\sum_{t,s} \hat{\mathcal{C}}(t,s)\hat{\gamma}_n(t)\hat{\gamma}_n(s)} = \left\langle e^{-i\sum_t \phi_n(t)\hat{\gamma}_n(t)} \right\rangle_{\phi_n}. \quad (\text{C.20})$$

Doing so leads (C.19) to

$$\begin{aligned} \langle Z(\{I(t)\}) \rangle_J &\simeq \prod_{n=1}^N \int d\mathcal{N}(\{\phi_n(t)\}; 0, \mathcal{C}_\phi) \\ &\times \prod_{t=0}^{T-1} \{dm_n(t) d\gamma_n(t) \delta(m_n(t) - f_t\{m_n(\tau), \gamma_n(\tau)\}_{\tau=0}^{t-1}) \\ &\times \delta\left(\gamma_n(t) - h_n - \sum_{s < t} \hat{\mathcal{G}}(t, s)m_n(s) - \phi_n(t)\right) e^{i\gamma_n(t) l_n(t)} \}. \end{aligned} \quad (\text{C.21})$$

Finally, note that  $-i\langle m_n(t)\hat{\gamma}_n(s) \rangle_{\phi_n} = \langle \frac{\partial m_n(t)}{\partial \phi_n(s)} \rangle_{\phi_n}$ . Thus,  $\mathcal{G}$  equals the response function (21). This completes the derivation.

#### Appendix D. Derivation of equation (60)

First note that from (27) and (30) we have

$$\mathcal{C}_\psi(t-1, t'-1) = \frac{(\mathcal{G}\mathcal{C}_\phi\mathcal{G}^\dagger)(t, t')}{(1-q(t))(1-q(t'))} \quad (\text{D.1})$$

where  $\mathcal{C}_\phi$  is defined as in (20). For sake of compactness we let  $f(x) = \mathbf{R}^{-1}(x)$ . By elementary combinatorics and using  $\mathcal{G} = f(\hat{\mathcal{G}})$  we can show that any power of the matrix  $\mathcal{G}$  can be written as

$$\mathcal{G}^k = \sum_{n=1}^{\infty} \text{Co}_{x^n}(f(x)^k) \hat{\mathcal{G}}^n. \quad (\text{D.2})$$

where for any power series  $f(x) = \sum_n a_n x^n$  we represent the coefficient via the definition  $a_k \triangleq \text{Co}_{x^k}(f(x))$ . This means that we have

$$\mathcal{G}^k(t, \tau) = \text{Co}_{x^{t-\tau}}(f(x)^k) \prod_{s=\tau}^{t-1} \hat{\mathcal{G}}(s+1, s). \quad (\text{D.3})$$

Hence we also obtain

$$(\mathcal{G}^{k+1}\mathcal{C}(\mathcal{G}^\dagger)^{n-1-k})(t-1, t'-1) = \sum_{l < t, m < t'} \left( \prod_{s=l}^{t-1} \hat{\mathcal{G}}(s+1, s) \right) \left( \prod_{s=m}^{t'-1} \hat{\mathcal{G}}(s+1, s) \right) \\ \times \mathcal{C}(l, m) \text{Co}_{x^{t-l}y^{t'-m}}(f(x)^{k+1}f(y)^{n-1-k}). \quad (\text{D.4})$$

where we have extended the definitions of coefficients  $\text{Co}_{x^n y^k}[f(x, y)] \triangleq a_n b_k$  to double power series  $f(x, y) = \sum_{n,k \geq 0} a_n b_k x^n y^k$ . Summing up the geometric series, we have

$$\sum_{k=0}^{n-2} \text{Co}_{x^{t-l}y^{t'-m}}(f(x)^{k+1}f(y)^{n-1-k}) = \text{Co}_{x^{t-l}y^{t'-m}} \left( \frac{f(y)^{n-1} - f(x)^{n-1}}{[f(y) - f(x)][f(x)f(y)]} \right) \quad (\text{D.5})$$

and

$$\sum_{n=1}^{\infty} c_n \text{Co}_{x^{t-l}y^{t'-m}} \left( \frac{f(y)^{n-1} - f(x)^{n-1}}{[f(y) - f(x)][f(x)f(y)]} \right) \\ = \text{Co}_{x^{t-l}y^{t'-m}} \left( \frac{y - x}{[f(y) - f(x)][f(x)f(y)]} \right). \quad (\text{D.6})$$

Putting everything together completes the proof.

## Appendix E. Derivation of equation (76)

For the sake of compactness, without loss of generality, we may set  $h_i = h$ . Using the representation of the Gaussian density in terms of the characteristic function we have the expansion for  $t \neq t'$

$$\langle \tanh(u(t) + h) \tanh(u(t') + h) \rangle_u \\ \simeq \int du_1 du_2 dk_1 dk_2 \exp[i(k_1 u_1 + k_2 u_2) \\ - \frac{1}{2} \{s_0 + \epsilon(s(t, t) + 1)\}(k_1^2 + k_2^2) - \{s_0 + \epsilon s(t, t')\}k_1 k_2] \\ \times \tanh(u_1 + h) \tanh(u_2 + h) \\ \simeq q - \frac{\epsilon}{2} \int du_1 du_2 dk_1 dk_2 \exp \left[ i(k_1 u_1 + k_2 u_2) - \frac{s_0}{2}(k_1^2 + k_2^2 + 2k_1 k_2) \right] \\ \times \tanh(u_1 + h) \tanh(u_2 + h) \{ (s(t, t) + 1)(k_1^2 + k_2^2) + 2s(t, t')k_1 k_2 \} \\ = q + \epsilon(s(t, t) + 1) \left\langle \tanh(u + h) \frac{\partial^2 \tanh(u + h)}{\partial u^2} \right\rangle \\ + \epsilon s(t_1, t_2) \left\langle \left( \frac{\partial \tanh(u + h)}{\partial u} \right)^2 \right\rangle. \quad (\text{E.1})$$

The last line is obtained by representing  $k_1 k_2$ , etc, by derivatives with respect to  $u_1$  and  $u_2$ . Repeating a similar equation for  $t = t'$ , we obtain



$$\begin{aligned}
\langle \tanh^2(u(t) + h) \rangle_u &= q + \frac{\epsilon}{2} (s(t, t) + 1) \left\langle \frac{\partial^2 \tanh^2(u + h)}{\partial u^2} \right\rangle \\
&= q + \epsilon (s(t, t) + 1) \left\langle \tanh(u + h) \frac{\partial^2 \tanh(u + h)}{\partial u^2} + \left( \frac{\partial \tanh(u + h)}{\partial u} \right)^2 \right\rangle. \quad (\text{E.2})
\end{aligned}$$

Both expansions can be represented in the single equation

$$\begin{aligned}
\frac{d \langle \tanh(u(t) + h) \tanh(u(t') + h) \rangle_u}{d\epsilon} &= (s(t, t) + 1) \left\langle \tanh(u + h) \frac{\partial^2 \tanh(u + h)}{\partial u^2} \right\rangle \\
&\quad + (s(t, t') + \delta_{t,t'}) \left\langle \left( \frac{\partial \tanh(u + h)}{\partial u} \right)^2 \right\rangle. \quad (\text{E.3})
\end{aligned}$$

Note that only the second term contributes to the dynamic part of the fluctuations. Hence, by taking the Fourier-transform and noting that  $\frac{\partial \tanh(u + h)}{\partial u} = 1 - \tanh^2(u + h)$  the result is obtained.

## References

- [1] Mézard M, Parisi G and Virasoro M 1987 *Spin Glass Theory and Beyond (Lecture Notes in Physics vol 9)* (Singapore: World Scientific)
- [2] Kabashima Y 2003 *J. Phys. A: Math. Gen.* **36.43** 11111
- [3] Bayati M and Montanari A 2011 *IEEE Trans. Inform. Theory* **57** 764–85
- [4] Oppen M and Winther O 2001 *Phys. Rev. E* **64** 056131
- [5] Oppen M and Winther O 2000 *Neural Computation* **12** 2655–84
- [6] Minka T P 2001 Expectation propagation for approximate Bayesian inference *Proc. 17th Conf. on Uncertainty in Artificial Intelligence (UAI)* pp 362–9
- [7] Thouless D J, Andersen P W and Palmer R G 1977 *Phil. Mag.* **35.3** 593–601
- [8] Bolthausen E 2014 *Commun. Math. Phys.* **325.1** 333–66
- [9] Donoho D L, Maleki A and Montanari A 2009 *Proc. Natl Acad. Sci.* **106** 18914–9
- [10] Krzakala F, Mézard M, Sausset F, Sun Y and Zdeborová L 2012 *J. Stat. Mech.: Theory Exp.* **2012**
- [11] Bayati M, Lelarge M and Montanari A 2015 *Ann. Appl. Probab.* **25** 753–822
- [12] Parisi G and Potters M 1995 *J. Phys. A: Math. Gen.* **28.18** 5267
- [13] Marinari E, Parisi G and Ritort F 1994 *J. Phys. A: Math. Gen.* **27.23** 7647
- [14] Deift P and Gioev D 2009 *Random Matrix Theory: Invariant Ensembles and Universality* vol 18 (American Mathematical Society)
- [15] Guionnet A and Marˆda M 2005 *J. Funct. Anal.* **222** 435–90
- [16] Voiculescu D V, Dykema K J and Nica A 1992 *Free Random Variables* vol 1 (American Mathematical Society)
- [17] Hiai F and Petz D 2006 *The Semicircle Law Free Random Variables and Entropy* (American Mathematical Society)
- [18] Martin P C, Siggia E D and Rose H A 1973 *Phys. Rev. A* **8.1** 423
- [19] Eiseffler H and Oppen M 1992 *Phys. Rev. Lett.* **68** 2094
- [20] Henkel R D and Oppen M 1991 *J. Phys. A: Math. Gen.* **24.9** 2201
- [21] Mimuura K and Okada M 2014 *IEEE Trans. Inform. Theory* **60** 3645–70
- [22] Cugliandolo L F and Kurchan J 1994 *J. Phys. A: Math. Gen.* **27.17** 5749
- [23] Tanaka T 2008 *J. Phys.: Conf. Ser.* **95** 012002
- [24] Vehkaperä M, Kabashima Y and Chatterjee S 2013 arXiv:1312.0256
- [25] Tulino A M, Caire G, Shamai S and Verdú S 2013 *IEEE Trans. Inform. Theory* **59** 4243–71
- [26] Collins B and Sniady P 2006 *Commun. Math. Phys.* **264.3** 773–95
- [27] Collins B and Matsumoto S 2009 *J. Math. Phys.* **50** 113516
- [28] Akemann G, Baik J and Di Francesco P 2011 *The Oxford Handbook of Random Matrix Theory* (Oxford: Oxford University Press)
- [29] Tulino A M and Verdú S 2004 *Random Matrix Theory and Wireless Communications* vol 1 (Delft: Now)