Aalborg Universitet

Dynamical Functional Theory for Compressed Sensing

Cakmak, Burak; Opper, Manfred ; Winther, Ole; Fleury, Bernard Henri

Published in:
2017 IEEE International Symposium on Information Theory (ISIT)

DOI (link to publication from Publisher):
10.1109/ISIT.2017.8006908

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Cakmak, B., Opper, M., Winther, O., & Fleury, B. H. (2017). Dynamical Functional Theory for Compressed Sensing. In *2017 IEEE International Symposium on Information Theory (ISIT)* (pp. 2143-2147). IEEE (Institute of Electrical and Electronics Engineers). https://doi.org/10.1109/ISIT.2017.8006908

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: December 05, 2025

Dynamical Functional Theory for Compressed Sensing

Burak Çakmak
Aalborg University
9220 Aalborg, Denmark
Email: buc@es.aau.dk

Manfred Opper
Technical University of Berlin
Berlin 10587, Germany
Email: manfred.opper@tu-berlin.de

Ole Winther
Technical University of Denmark
2800 Lyngby, Denmark
Email: olwi@dtu.dk

Bernard H. Fleury
Aalborg University
9220 Aalborg, Denmark
Email: fleury@es.aau.dk

***Abstract*—We introduce a theoretical approach for designing generalizations of the approximate message passing (AMP) algorithm for compressed sensing which are valid for large observation matrices that are drawn from an invariant random matrix ensemble. By design, the fixed points of the algorithm obey the Thouless-Anderson-Palmer (TAP) equations corresponding to the ensemble. Using a dynamical functional approach we are able to derive an effective stochastic process for the marginal statistics of a single component of the dynamics. This allows us to design memory terms in the algorithm in such a way that the resulting fields become Gaussian random variables allowing for an explicit analysis. The asymptotic statistics of these fields are consistent with the replica ansatz of the compressed sensing problem.**

I. Introduction

The canonical problem of compressed sensing is concerned with the linear observation model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{A}$ is the $N \times K$ observation matrix, $\boldsymbol{x}$ is the independent and identically distributed (iid) (unobserved) signal vector, $\boldsymbol{y}$ is the observation vector and $\boldsymbol{n}$ is the iid vector of observation errors whose entries are zero-mean Gaussian with precision ξ. Here, all variables are real-valued. The dimensions N and K are both large, possibly arbitrary large with the aspect ratio $\alpha \triangleq N/K$ fixed. Compressed sensing aims to recover the signal $\boldsymbol{x}$ from a compressive sampling system, i.e. $N < K$, by exploiting the sparsity structure of $\boldsymbol{x}$.

Within the above context, the approximate message passing (AMP) algorithm [1] has drawn significant attention in the communities of information theory and signal processing. In particular, the algorithm has the virtue that the trajectories of its dynamics can be described by a simple *one-dimensional* iterative equation – called state evolution [2]. State evolution is not only useful for performing the signal recovery but also for the analysis of (1) within information theory [3].

The AMP method is based on the assumption that the entries of observation matrix are iid zero-mean Gaussian[1] with variance $1/N$. However, without a clear physical and/or mathematical motivation, the iid assumption can be artificial. Indeed, signals are often sparse in the Fourier domain, in which case the observation matrix is typically a $N \times K$ "corner" of the $K \times K$ randomly-permuted discrete-cosine transform (DCT) matrix. In this case, the optimality of the AMP algorithm is broken and state evolution becomes invalid.

It is the goal of this study to develop a theoretical framework that extends the AMP method in such a way that it is valid for general invariant matrix ensembles [5]. Specifically, we assume that Gramian $\tilde{\boldsymbol{J}} \triangleq \boldsymbol{A}^{\dagger}\boldsymbol{A}$ is invariant, i.e. it has the same distribution as $\boldsymbol{O}^{\dagger}\tilde{\boldsymbol{J}}\boldsymbol{O}$ for any orthogonal matrix $\boldsymbol{O}$ independent of $\tilde{\boldsymbol{J}}$.[2] These ensembles are now favoured in performance analyses of compressed sensing [8–10].

The AMP algorithm can be viewed as a system of dynamical equations that attempt to solve the so-called Thouless-Anderson-Palmer (TAP) (fixed-point) equations [11–13] of the observation model (1) under the iid assumption about the entries of $\boldsymbol{A}$, see [14–16]. Although the AMP method was originally derived under the assumption that the observation matrix is sparse, due to the iid property of the matrix elements, the method also leads to a correct algorithm when said matrix is dense. Since it is not clear how to deal with sparse observation matrices for more general random matrix ensembles, we develop in this study a method for describing the dynamics of iterative algorithms involving dense observation matrices as a starting point. Specifically, we first introduce general dynamical equations of a candidate algorithm that attempts to solve the TAP equations for general invariant matrix ensembles. Then, we make use of dynamical functional theory (DFT) [17, 18] to statistically characterize the trajectory generated by these equations in the large-system limit (more specifically, in the thermodynamic limit). This approach yields a method which allows for both designing and analyzing algorithms involving general invariant matrix ensembles.

This paper is organized as follows: In Section 2 we present the elementary tools from random matrix theory that we use in this contribution and summarize the TAP equations of (1) for general invariant ensembles. In Section 3 we present the results of DFT. Section 4 presents the "single-step memory" algorithm. Section 5 is devoted to the derivation of the statistical characterization of the trajectory of the single-step memory algorithm. Section 6 gives a summary and outlook.

This work was partially supported by the research project VIRTUOSO funded by Intel Mobile Communications, Keysight, Telenor, Aalborg University, and the Danish National Advanced Technology Foundation.

[1]The Gaussian condition can be relaxed to a sub-Gaussian distribution [4].

[2]Though a Gramian of the corner of randomly-permuted DCT matrix is not invariant, it asymptotically behaves as an invariant matrix, see e.g. [6, 7].

A. Related works

We refer to [19] which makes use of DFT to analyze an AMP-type iterative detection algorithm for code-division–multiple-access systems with the classical iid assumption about the entries of the observation matrix.

From the technical point of view, this paper can be viewed as an extension of [20] where DFT is used for solving the TAP equations of the Ising model of spin glass theory. Due to space limitations, we do not present the proofs of our results; they are obtained by following arguments similar to those used in [20], see also the paragraph below (12).

B. Notations

The entries of a $K \times N$ matrix $\boldsymbol{X}$ and a $K \times 1$ vector $\boldsymbol{x}$ are denoted as X_{nk} and x_k for $n \in [1, N]$ and $k \in [1, K]$. We use calligraphic notation, e.g. $\mathcal{G}$, for $T \times T$ matrices whose entry indices range (for convenience) in $[0, T-1]$, i.e. starting at 0. Thus, $\mathcal{G}(t, \tau)$ is the $(t+1, \tau+1)$ entry of $\mathcal{G}$. The transposition is denoted by $(\cdot)^\dagger$. The Gaussian distribution function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. For random variables x and y, $x \sim y$ implies that x and y are identically distributed. We indicate that x has a probability density function (pdf) p, or distribution function P, by $x \sim p$, or $x \sim \mathrm{P}$. Moreover, $\langle x \rangle = \langle x \rangle_p = \int x \mathrm{dP}(x)$.

II. Preliminaries

Throughout the paper, we assume the following.

ASSUMPTION 1 *Gramian $\tilde{\boldsymbol{J}} = \boldsymbol{A}^\dagger \boldsymbol{A}$ is invariant, has a uniformly bounded spectral norm and its empirical eigenvalue distribution converges almost surely to a limiting eigenvalue distribution (LED) as $K \to \infty$ (with $\alpha = N/K$ fixed). For convenience, we assume that $\mathrm{tr}(\tilde{\boldsymbol{J}})/K \to 1$ almost surely as $K \to \infty$.*

We shall exemplify our general results through two specific matrix ensembles: (i) $\boldsymbol{A}$ has iid zero-mean Gaussian entries with variance $1/N$; (ii) $\boldsymbol{A}$ is random row-orthogonal, namely $\boldsymbol{A} = \alpha^{-\frac{1}{2}} \boldsymbol{P}_\alpha \boldsymbol{O}$ where $\boldsymbol{O}$ is a $K \times K$ Haar matrix and $N \times K$ matrix $\boldsymbol{P}_\alpha$ removes $K - N$ rows of $\boldsymbol{U}$. Specifically, $[P_\alpha]_{ij} = \delta_{ij}$, $\forall i, j$, where δ_{ij} denotes the Kronecker delta.

For mathematical convenience, our random matrix arguments will be based on the LED of $\boldsymbol{J} \triangleq \xi \mathbf{I} - \xi \tilde{\boldsymbol{J}}$.

A. R-transform and its inverse

Our key tool from random matrix theory is the so-called R-transform [21] of the LED of $\boldsymbol{J}$ that is defined as

$$\mathrm{R}(\omega) \triangleq \sum_{n=1}^{\infty} c_n \omega^{n-1} \tag{2}$$

where c_n is the distribution's nth free-cumulant. Here, we underline the simple relationship $\mathrm{R}(\omega) = \xi - \xi \tilde{\mathrm{R}}(-\xi\omega)$ where $\tilde{\mathrm{R}}$ is the R-transform of the LED of $\tilde{\boldsymbol{J}}$. In particular, $\tilde{\mathrm{R}}$ can be conveniently obtained from the inverse (with respect to the composition of functions) of the Stieltjes transform of the LED of $\tilde{\boldsymbol{J}}$, for details we refer to [21].

For matrix ensemble (i), $\tilde{\mathrm{R}}$ is the R-transform of the Marčenko-Pastur distribution (see [21]) and thereby we have

$$\mathrm{R}(\omega) = \frac{\xi^2 \omega}{\alpha + \xi \omega}. \tag{3}$$

For matrix ensemble (ii), one can show that (see [8, Eq. (36)])

$$\mathrm{R}(\omega) = \xi - \frac{(\frac{\xi\omega}{\alpha} + 1) - \sqrt{(\frac{\xi\omega}{\alpha} + 1)^2 - 4\xi\omega}}{2\omega}. \tag{4}$$

By the definition of $\boldsymbol{J}$, we have $\mathrm{R}(0) = 0$. Hence, Lagrange's inversion theorem implies that the inverse R^{-1} exists and is analytic in a neighborhood of zero. Thus, we can write

$$\mathrm{R}^{-1}(\omega) = \sum_{n=1}^{\infty} a_n \omega^n. \tag{5}$$

For matrix ensembles (i) and (ii) we obtain respectively $a_n = (\alpha\xi^{-1})\xi^{-n}$ and

$$a_n = (\alpha\xi^{-1})\xi_1^{-n} - (\alpha\xi^{-1})\xi_2^{-n} \tag{6}$$

where for short we introduce the vector $\boldsymbol{\xi} \triangleq (\xi, \xi(\alpha - 1)/\alpha)$.

B. TAP equations with general invariant matrix ensembles

The TAP method [11, 12] – also known as expectation consistency in machine learning [13] – typically provides highly accurate approximations for probabilistic inference. In our context, the TAP equations refer to an approximate, denoted by $\boldsymbol{m}$, of the minimum-mean-square-error estimate of the signal, i.e. the mean of the posterior pdf of $\boldsymbol{x}$: $\boldsymbol{m} \approx \langle \boldsymbol{x} \rangle_{p(\boldsymbol{x}|\boldsymbol{y})}$.

For convenience, we introduce the auxiliary pdf

$$q_{\boldsymbol{\psi};v}(\boldsymbol{x}) \propto p(\boldsymbol{x}) \exp\left(-\frac{v}{2}\|\boldsymbol{x} - \boldsymbol{\psi}\|^2\right) \tag{7}$$

where $p(\boldsymbol{x})$ is the prior pdf of the iid signal vector $\boldsymbol{x}$. The mean and the normalized total variance of this pdf are denoted as

$$\eta_v(\boldsymbol{\psi}) \triangleq \langle \boldsymbol{x} \rangle_{q_{\boldsymbol{\psi};v}} \text{ and } \chi \triangleq \frac{1}{K} \langle \|\boldsymbol{x} - \eta_v(\boldsymbol{\psi})\|^2 \rangle_{q_{\boldsymbol{\psi};v}}. \tag{8}$$

Then, the TAP equations of (1) for general invariant matrix ensembles can be expressed as [12, 16]

$$\boldsymbol{m} = \eta_v(\boldsymbol{\psi}) \quad \text{with} \quad v \triangleq \xi - \mathrm{R}(\chi) \tag{9a}$$

$$\boldsymbol{\psi} = \frac{1}{v}\left(\boldsymbol{h} + \boldsymbol{J}\boldsymbol{m} - \mathrm{R}(\chi)\boldsymbol{m}\right) \quad \text{with} \quad \boldsymbol{h} \triangleq \xi \boldsymbol{A}^\dagger \boldsymbol{y}. \tag{9b}$$

Note that the dependency on the random matrix ensemble in the TAP equations is via the R-transform of the LED of $\boldsymbol{J}$.

The TAP equations are known to be consistent with the replica ansatz [8] – which is commonly assumed to be exact above the so-called Almedia-Thouless line of the stability [22]. Specifically, the entries of the static field $\boldsymbol{\psi}$ are asymptotically iid – as implied by the decoupling principle – and [14]

$$\psi_k \sim \theta + x \quad \text{with} \quad \theta \sim \mathcal{N}(0, 1/v) \text{ and } x \sim p(x) \tag{10}$$

for θ independent of x. Here and in the sequel, we substitute v (or χ) with its (non-random) large-system approximation obtained with the replica ansatz, i.e. $\chi = \langle (x - \eta_v(\theta + x))^2 \rangle_{x,\theta}$. We next extend the "decoupling principle" originally formulated for static solutions to the trajectory of a dynamical algorithm that computes these solutions.

III. The Results of Dynamical Functional Theory

We start with the following system of dynamical equations of the discrete time $t = 0, 1, \ldots, T-1$ as a candidate for solving the TAP equations (9):

$$\boldsymbol{m}(t) = f_t\left(\{\boldsymbol{\gamma}(\tau), \boldsymbol{m}(\tau)\}_{\tau<t}\right) \tag{11a}$$

$$\boldsymbol{\gamma}(t) = \boldsymbol{h} + \boldsymbol{J}\boldsymbol{m}(t) \tag{11b}$$

where $\{f_t\}_{t<T}$ is an appropriate sequence of non-linear scalar functions. We obtain the statistics of the trajectory of this system from its generating functional

$$\begin{aligned} Z(\{\boldsymbol{l}(t)\}) = \int \prod_{t=0}^{T-1} \mathrm{d}\boldsymbol{m}(t)\mathrm{d}\boldsymbol{\gamma}(t)\; e^{\mathrm{i}\boldsymbol{\gamma}(t)^\dagger \boldsymbol{l}(t)} \times \\ \times\, \delta(\boldsymbol{m}(t) - f_t\left(\{\boldsymbol{\gamma}(\tau), \boldsymbol{m}(\tau)\}_{\tau<t}\right)) \times \\ \times\, \delta(\boldsymbol{\gamma}(t) - \boldsymbol{h} - \boldsymbol{J}\boldsymbol{m}(t)). \end{aligned} \tag{12}$$

Note that $\boldsymbol{h}$ is a function of the random elements $\boldsymbol{x}$, $\boldsymbol{n}$ and $\boldsymbol{A}$, specifically $\boldsymbol{h} = \xi \boldsymbol{A}^\dagger(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{n})$. This is what essentially distinguishes the problem from that addressed in [20] and requires the proofs in [20] to be adapted.

We are interested in the thermodynamical properties of $Z(\{\boldsymbol{l}(t)\})$, i.e. its expectation over $\boldsymbol{x}$, $\boldsymbol{n}$ and $\boldsymbol{A}$ that we denote as $\langle Z(\{\boldsymbol{l}(t)\})\rangle$. Here, we note that $Z(\{\boldsymbol{l}(t) = \boldsymbol{0}\}) = 1$ and therefore direct averaging of $Z(\{\boldsymbol{l}(t)\})$ is a proper operation. By making use of the R-transform formulation of the asymptotic Itzykson-Zuber integral [23] and the saddle-point method [24] we obtain that $\langle Z(\{\boldsymbol{l}(t)\})\rangle$ factorizes (or decouples) as

$$\begin{aligned} \prod_{k=1}^{K} \int \mathrm{dP}(x_k)\mathrm{d}\mathcal{N}(\{\phi_k(t)\}; \boldsymbol{0}, \mathcal{C}_\phi) \prod_{t=0}^{T-1} \mathrm{d}m_k(t)\mathrm{d}\gamma_k(t) \times \\ \times e^{O(\frac{1}{K}) + \mathrm{i}\gamma_k(t) l_k(t)} \times \\ \times \delta(m_k(t) - f_t\left(\{\gamma_k(\tau), m_k(\tau)\}_{\tau<t}\right)) \times \\ \times \delta\Big(\gamma_k(t) - \phi_k(t) - \xi x_k - \sum_{\tau=0}^{t-1} \hat{\mathcal{G}}(t,\tau)(m_k(\tau) - x_k)\Big) \end{aligned} \tag{13}$$

where $O(1/K)$ is a constant term that, as indicated, vanishes as $K \to \infty$. The random fields $\{\phi_k(t)\} = \{\phi_k(t)\}_{t<T}$, $k \in [1, K]$ have the $(T \times T)$ covariance matrix

$$\begin{aligned} \mathcal{C}_\phi = \xi \mathbf{1}\mathbf{1}^\dagger - \sum_{n=1}^{\infty} c_n \sum_{k=0}^{n-1} \mathcal{G}^k \mathbf{1}\mathbf{1}^\dagger (\mathcal{G}^\dagger)^{n-1-k} \\ + \sum_{n=1}^{\infty} c_n \sum_{k=0}^{n-2} \mathcal{G}^k \mathcal{C} (\mathcal{G}^\dagger)^{n-2-k}. \end{aligned} \tag{14}$$

Here, $\mathbf{1}$ is the all-ones $T \times 1$ vector and $\mathcal{G}$ and $\mathcal{C}$ are the $T \times T$ response and correlation matrices respectively with

$$\mathcal{G}(t,\tau) = \left\langle \frac{\partial m_k(t)}{\partial \phi_k(\tau)} \right\rangle \tag{15}$$

$$\mathcal{C}(t,\tau) = \langle (m_k(t) - x_k)(m_k(\tau) - x_k) \rangle. \tag{16}$$

The expectations in the above expressions are over random variables ϕ_k and x_k. The $T \times T$ memory matrix $\hat{\mathcal{G}}$ in (13) is defined as

$$\hat{\mathcal{G}} = \mathrm{R}(\mathcal{G}). \tag{17}$$

These results imply that for sufficiently large K, we can compute the marginal statistics (i.e. the statistics of single components of the vector $\boldsymbol{m}(t)$) in (11) from an effective stochastic process with iid components that is given by

$$\boldsymbol{m}(t) = f_t\left(\{\boldsymbol{\gamma}(\tau), \boldsymbol{m}(\tau)\}_{\tau<t}\right) \tag{18a}$$

$$\boldsymbol{\gamma}(t) = \boldsymbol{\phi}(t) + \xi \boldsymbol{x} + \sum_{\tau=0}^{t-1} \hat{\mathcal{G}}(t,\tau)(\boldsymbol{m}(\tau) - \boldsymbol{x}) \tag{18b}$$

where the entries of $\boldsymbol{\phi}(t)$ are iid zero-mean Gaussian processes with $T \times T$ covariance matrix $\mathcal{C}_\phi$. Unfortunately, due to the occurrence of the memory terms in (18b) the field $\boldsymbol{\gamma}(t)$ becomes non–Gaussian which makes an analysis of the algorithm in general intractable. While there are many ways to overcome this problem, we next limit our attention to a particular design in which almost all memory terms disappear.

IV. The Single-Step Memory Design

We consider (18a) to be of the form $\boldsymbol{m}(t+1) = \eta_{v(t)}(\boldsymbol{\psi}(t))$. Here, the scalar $v(t)$ is non-random. As long as $v(t) \to v$ is guaranteed we have the freedom to choose the dynamic evolution of $v(t)$ without referring to the results of DFT. We will use the results of DFT to design the random field $\boldsymbol{\psi}(t)$ in such a way that its components $\{\psi_k(t) : k \in [1, K]\}$ are iid but "temporally" dependent.

In the single-step memory design the dynamical system is built on the basis of the single-step memory condition [20]

$$\hat{\mathcal{G}}(t,\tau) = 0,\ \forall \tau \neq t-1. \tag{19}$$

From (11b) and (18b) this implies that

$$\begin{aligned} \boldsymbol{\phi}(t) &= \boldsymbol{\gamma}(t) - \xi \boldsymbol{x} - \hat{\mathcal{G}}(t,t-1)(\boldsymbol{m}(t-1) - \boldsymbol{x}) \qquad (20) \\ &= \boldsymbol{h} + \boldsymbol{J}\boldsymbol{m}(t) - \hat{\mathcal{G}}(t,t-1)\boldsymbol{m}(t-1) - \\ &\quad - (\xi - \hat{\mathcal{G}}(t,t-1))\boldsymbol{x} \qquad (21) \end{aligned}$$

with $\hat{\mathcal{G}}(0,-1) \triangleq 0$. We design the field $\boldsymbol{\psi}(t)$ to be of the form

$$\begin{aligned} \boldsymbol{\psi}(t) &= \sum_{\tau=0}^{t} \mathcal{A}(t+1,\tau)\left[\boldsymbol{\phi}(\tau) + (\xi - \hat{\mathcal{G}}(\tau,\tau-1))\boldsymbol{x}\right] \qquad (22) \\ &= \sum_{\tau=0}^{t} \mathcal{A}(t+1,\tau)[\boldsymbol{h} + \boldsymbol{J}\boldsymbol{m}(\tau) - \hat{\mathcal{G}}(\tau,\tau-1)\boldsymbol{m}(\tau-1)]. \end{aligned}$$

Conditioned on the static random field $\boldsymbol{x}$, the field $\boldsymbol{\psi}(t)$ is Gaussian, which allows for an explicit analysis. Basically, we have to design the non-random terms $\mathcal{A}(t,\tau)$ in such a way that the construction is consistent with the results of DFT. From (22) we obtain the relationship

$$\mathcal{G}(t,\tau) = \left\langle \frac{\partial \eta_{v(t-1)}(\psi_k(t-1))}{\partial \phi_k(\tau)} \right\rangle \tag{23}$$

$$= \chi(t) v(t-1) \mathcal{A}(t,\tau) \quad \text{with } \chi(t) \triangleq \mathcal{C}(t,t). \tag{24}$$

Furthermore, from (17) we have $\mathcal{G} = \mathrm{R}^{-1}(\hat{\mathcal{G}})$. Combining with the single-step memory property (19) yields

$$\mathcal{G}(t,\tau) = a_{t-\tau} \prod_{s=\tau}^{t-1} \hat{\mathcal{G}}(s+1,s). \tag{25}$$

Finally, we need to specify the memory terms $\hat{\mathcal{G}}(t,t-1)$ in such a way that the resulting algorithm is asymptotically (as $t\to\infty$) consistent with the TAP equations. We choose

$$\hat{\mathcal{G}}(t,t-1)=\frac{\chi(t)}{\chi(t-1)}\mathrm{R}(\chi(t-1)). \tag{26}$$

Combining the above steps, the single-step memory algorithm (for $t\geq 0$) can be written in the form

$$\boldsymbol{m}(t+1)=\eta_{v(t)}(\boldsymbol{\psi}(t)) \tag{27a}$$

$$\boldsymbol{\psi}(t)=\frac{Q(t)}{v(t)}\sum_{\tau=0}^{t}a_{t+1-\tau}\boldsymbol{u}(\tau) \tag{27b}$$

$$\boldsymbol{u}(t)=\frac{\boldsymbol{h}+\boldsymbol{J}\boldsymbol{m}(t)-\hat{\mathcal{G}}(t,t-1)\boldsymbol{m}(t-1)}{\chi(t)Q(t-1)} \tag{27c}$$

where $Q(t)=Q(t-1)\mathrm{R}(\chi(t))$ and $Q(-1)\triangleq 1$.

A. Asymptotic consistency with the TAP equations

Following the arguments [20] one can show that if the single-step memory algorithm (27) converges it solves the TAP fixed-point equations (9) under the so-called weak long-term response assumption [25]

$$\lim_{t\to\infty}\mathcal{G}(t,\tau)=0\quad\forall\text{ finite }\tau. \tag{28}$$

One can also show that (28) holds for matrix ensembles (i) and (ii). Instead, in the sequel we further simplify the form of the single-step memory algorithm for these ensembles in such a way that the consistency of its fixed points with the TAP equations can be trivially checked. We first consider the matrix ensemble (ii).

B. Example-1: Matrix ensemble (ii)

By using the expression of coefficient a_n for matrix ensemble (ii) in (6), we decompose the field $\boldsymbol{\psi}(t)$ as

$$\boldsymbol{\psi}(t)=\underbrace{\frac{Q(t)}{v(t)}\frac{\alpha}{\xi}\sum_{\tau=0}^{t}\frac{\boldsymbol{u}(\tau)}{\xi_1^{t+1-\tau}}}_{\boldsymbol{\psi}_1(t)}-\underbrace{\frac{Q(t)}{v(t)}\frac{\alpha}{\xi}\sum_{\tau=0}^{t}\frac{\boldsymbol{u}(\tau)}{\xi_2^{t+1-\tau}}}_{\boldsymbol{\psi}_2(t)}. \tag{29}$$

It is easy to see that each field admits the recursion

$$\begin{aligned}\boldsymbol{\psi}_i(t)=&\frac{\alpha}{\xi}\frac{\mathrm{R}(\chi(t))}{\chi(t)v(t)}\frac{1}{\xi_i}(\boldsymbol{h}+\boldsymbol{J}\boldsymbol{m}(t)-\hat{\mathcal{G}}(t,t-1)\boldsymbol{m}(t-1))\\&+\frac{\mathrm{R}(\chi(t))v(t-1)}{v(t)}\frac{1}{\xi_i}\boldsymbol{\psi}_i(t-1).\end{aligned} \tag{30}$$

For further convenience, we introduce the auxiliary fields

$$\boldsymbol{z}_i(t)\triangleq\frac{\boldsymbol{\psi}_i(t)}{c(t)}-\boldsymbol{m}(t)\quad\text{with}\quad c(t)\triangleq\frac{\alpha}{\xi}\frac{\mathrm{R}(\chi(t))}{\chi(t)v(t)}. \tag{31}$$

Applying this transformation to the recursion (30), the single-step memory algorithm (19) can be written in the form

$$\begin{aligned}\boldsymbol{m}(t+1)&=\eta_{v(t)}\left(c(t)[\boldsymbol{z}_1(t)-\boldsymbol{z}_2(t)]\right)\\\boldsymbol{z}_i(t)&=\frac{1}{\xi_i}(\boldsymbol{h}+\boldsymbol{J}\boldsymbol{m}(t)-\xi_i\boldsymbol{m}(t)+\hat{\mathcal{G}}(t,t-1)\boldsymbol{z}_i(t-1)).\end{aligned} \tag{32}$$

C. Example-2: AMP algorithm

For matrix ensemble (i), we show that by choosing the dynamic evolution of $v(t)$ as

$$v(t)=\xi-\mathrm{R}(\chi(t))=\frac{\xi\alpha}{\alpha+\xi\chi(t)}, \tag{33}$$

the single-step memory algorithm yields the AMP algorithm [1,2]: We have $a_n=(\alpha\xi^{-1})\xi^{-n}$. From the previous discussion, this implies that $\boldsymbol{\psi}(t)=c(t)(\boldsymbol{z}(t)-\boldsymbol{m}(t))$ where $\boldsymbol{z}(t)$ admits exactly the same form as $\boldsymbol{z}_1(t)$ in (32) and $c(t)$ is as in (29). Moreover, (33) yields that $c(t)=1$. Thereby, the single-step memory algorithm (19) simplifies to

$$\boldsymbol{m}(t+1)=\eta_{v(t)}(\boldsymbol{z}(t)+\boldsymbol{m}(t)) \tag{34}$$

$$\boldsymbol{z}(t)=\boldsymbol{A}^{\dagger}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{m}(t))+\frac{1}{\xi}\hat{\mathcal{G}}(t,t-1)\boldsymbol{z}(t-1). \tag{35}$$

V. The Field Statistic

To describe the trajectory of the single-step memory algorithm we solely need an explicit statistical characterization of the random field $\boldsymbol{\psi}(t)$. We note that its entries are iid and are expressed as linear combinations of the corresponding entries of the Gaussian field $\boldsymbol{\phi}(t)$ (see (22)):

$$\underbrace{\sum_{\tau=0}^{t}\mathcal{A}(t+1,\tau)\phi(\tau)}_{\theta(t)}+x\underbrace{\sum_{\tau=0}^{t}\mathcal{A}(t+1,\tau)(\xi-\hat{\mathcal{G}}(\tau,\tau-1))}_{\sigma_x(t)},$$

i.e. $\psi_k(t)\sim\theta(t)+\sigma_x(t)x$, with $\{\phi(t)\}\sim\mathcal{N}(\mathbf{0},\mathcal{C}_\phi)$, see (14). The expression $\sigma_x(t)$ can be written in the compact form

$$\sigma_x(t)=\frac{1}{v(t)}\left[\zeta(t)\xi-\zeta(t-1)\mathrm{R}(\chi(t))\right]. \tag{36}$$

Here, $\zeta(-1)\triangleq 0$ and for $t\geq 0$ we have

$$\zeta(t)\triangleq Q(t)\sum_{\tau=0}^{t}\frac{a_{t+1-\tau}}{\chi(\tau)Q(\tau-1)}. \tag{37}$$

Thus, we merely need to derive an explicit expression of the covariance matrix of $\{\theta(t)\}$, say $\mathcal{C}_\theta$. To this end, we introduce the bi-variate random matrix transformation

$$\mathrm{B}(\omega,z)\triangleq\left(\frac{1}{\mathrm{R}^{-1}(\omega)}-\frac{1}{\mathrm{R}^{-1}(z)}\right)^{-1}(z-\omega). \tag{38}$$

Moreover, for power series $f(\omega,z)=\sum_{n,k\geq 0}a_{nk}\omega^n z^k$, $\mathrm{Co}_{\omega^n z^k}[f(\omega,z)]\triangleq a_{nk}$. With these definitions, we obtain the explicit expression of $\mathcal{C}_\theta$ given in (39) at the top of the next page.

A. Asymptotic consistency with the replica ansatz

We assume that $\chi(t)$ is convergent as $t\to\infty$. Then, the weak long-term assumption (28) implies that $\zeta(t)\to 1$. Thereby, it is easy to see from the expressions in (36) and (39) that $\sigma_x(t)\to 1$ and $\mathcal{C}_\theta(t,t)\to 1/v$. In other words, the field statistics asymptotically obey the equations of the static solutions of the replica ansatz (10) [8, 14].

$$\mathcal{C}_\theta(t,t') = \frac{\zeta(t')}{v(t')}\sigma_x(t) + \frac{Q(t)Q(t')}{v(t)v(t')}\sum_{\tau=0}^{t}\sum_{s=0}^{t'} \frac{\mathrm{Co}_{\omega^{t+1-\tau}z^{t'+1-s}}[\mathrm{B}(\omega,z)](\mathcal{C}(\tau,s)-\chi(s)\zeta(s-1))}{\chi(\tau)\chi(s)Q(\tau-1)Q(s-1)}. \tag{39}$$

B. State evolution of AMP

For matrix ensemble (i), it follows from $a_n = (\alpha\xi^{-1})\xi^{-n}$ that $\sigma_x(t) = 1$ and $\zeta(t) = 1$. Moreover, from (3) we get $\mathrm{B}(\omega, z) = \alpha\omega z\xi^{-2}$. Plugging these results into (39) yields the state evolution formula of the AMP algorithm [1, 2]:

$$C_\theta(t+1,t+1) = \frac{1}{\xi} + \frac{1}{\alpha}\left\langle \left(\eta_{v(t)}\left(\sqrt{\mathcal{C}_\theta(t,t)}z + x\right) - x\right)^2 \right\rangle$$

where $z \sim \mathcal{N}(0,1)$ is independent of $x \sim p(x)$.

C. Example: Matrix ensemble (ii)

For matrix ensemble (ii), we obtain the expression $\mathrm{Co}_{\omega^n z^k}[\mathrm{B}(\omega,z)] = -(\xi_1\xi_2)^{-n}\delta_{nk}$. Using this result in (39), we recast the variance $C_\theta(t,t)$ in (39) in the compact form

$$C_\theta(t,t) = \frac{\zeta(t)}{v(t)}\sigma_x(t) - \frac{\mathrm{R}(\chi(t))^2}{\chi(t)^2 v(t)^2}\kappa(t) \tag{40}$$

with $\kappa(t) = [\chi(t) - \chi(t)\zeta(t-1) + \hat{\mathcal{G}}(t,t-1)^2\kappa(t-1)]/(\xi_1\xi_2)$.

VI. Summary and Outlook

By making use of DFT, we have introduced a theoretical method to design and analyze iterative algorithms for solving the TAP equations of compressed sensing with general invariant matrix ensembles. We have focused our attention on the single-step memory design as it yields the AMP algorithm and its state evolution formula under the classical iid assumption for the entries of the observation matrix. Yet, there are many interesting way of defining other designs, e.g. the single-step response design: $\mathcal{G}(t,\tau) = 0$, for all $\tau \neq t-1$. We leave the study of such schemes to future work.

We did not include any analysis of the dynamical instability of the algorithm. Our simulation results – not reported here – show sensitivity to initialization, especially for high noise precision ξ. Augmenting the analysis with an investigation of the dynamical instability conditions will potentially allow us to design more powerful algorithms.

Recently, [26, 27] have reported important results on the dynamical analysis of expectation-propagation based algorithms whose fixed points are consistent with the TAP equations (9). An interesting task is to relate these results with those obtained with DFT.

References

[1] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 18 914–18 919, September 2009.

[2] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, pp. 764–785, February 2011.

[3] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," *arXiv preprint arXiv: 1607.02335*, 2016.

[4] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and message passing algorithms," *The Annals of Applied Probability*, vol. 25, no. 2, pp. 753–822, 04 2015.

[5] P. Deift and D. Gioev, *Random Matrix Theory: Invariant Ensembles and Universality*. American Mathematical Society, 2009, vol. 18.

[6] E. Marinari, G. Parisi, and F. Ritort, "Replica field theory for deterministic models. ii. a non-random spin glass with glassy behaviour," *Journal of Physics A: Mathematical and General*, vol. 27.23, p. 7647, 1994.

[7] G. W. Anderson and B. Farrell, "Asymptotically liberating sequences of random unitary matrices," *Advances in Mathematics*, vol. 255, pp. 381 – 413, 2014.

[8] A. M. Tulino, G. Caire, S. Shamai, and S. Verdú, "Support recovery with sparsely sampled free random matrices," *IEEE Transactions on Information Theory*, vol. 59, pp. 4243–4271, 2013.

[9] M. Vehkaperä, Y. Kabashima, and S. Chatterjee, "Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2100–2124, 2016.

[10] A. Bereyhi, R. R. Müller, and H. Schulz-Baldes, "Statistical mechanics of MAP estimation: General replica ansatz," *arXiv preprint arXiv: 1612.01980*, 2016.

[11] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond*. World Scientific, 1987, vol. 9.

[12] M. Opper and O. Winther, "Adaptive and self-averaging Thouless-Anderson-Palmer mean field theory for probabilistic modeling," *Physical Review E*, vol. 64, pp. 056 131–(1–14), October 2001.

[13] ——, "Expectation consistent approximate inference," *Journal of Machine Learning Research*, 6 (2005): 2177-2204.

[14] Y. Kabashima and M. Vehkaperä, "Signal recovery using expectation consistent approximation for linear observations," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 226–230.

[15] B. Çakmak, O. Winther, and B. H. Fleury, "S-AMP: Approximate message passing for general matrix ensembles," in *Proc. IEEE Information Theory Workshop (ITW)*, Hobart, Tasmanina, Australia, November 2014.

[16] B. Çakmak, M. Opper, B. H. Fleury, and O. Winther, "Self-averaging expectation propagation," *arXiv preprint arXiv:1608.06602*, August 2016.

[17] P. C. Martin, E. D. Siggia, and H. A. Rose, "Statistical dynamics of classical systems," *Physical Review A*, vol. 8.1, p. 423, 1973.

[18] H. Eisfeller and M. Opper, "New method for studying the dynamics of disordered spin systems without finite-size effects," *Physical Review Letters*, vol. 68, p. 2094, 1992.

[19] K. Mimura and M. Okada, "Generating functional analysis for iterative CDMA multiuser detectors," *IEEE Transactions on Information Theory*, vol. 60, pp. 3645–3670, June 2014.

[20] M. Opper, B. Çakmak, and O. Winther, "A theory of solving TAP equations for Ising models with general invariant random matrices," *Journal of Physics A: Mathematical and Theoretical*, vol. 49, no. 11, p. 114002, 2016.

[21] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Now Publishers Inc., June 2004, vol. 1, no. 1.

[22] J. R. L. De Almeida and D. J. Thouless, "Stability of the Sherrington-Kirkpatrick solution of a spin glass model," *Journal of Physics A: Mathematical and General*, vol. 11, no. 5, p. 983, 1978.

[23] B. Collins and P. Śniady, "New scaling of Itzykson–Zuber integrals," *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, vol. 43, no. 2, pp. 139 – 146, 2007.

[24] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, no. 12, pp. 1–212, 2010. [Online]. Available: http://dx.doi.org/10.1561/0100000052

[25] L. F. Cugliandolo and J. Kurchan, "On the out-of-equilibrium relaxation of the Sherrington-Kirkpatrick model," *Journal of Physics A: Mathematical and General*, vol. 27.17, p. 5749, 1994.

[26] S. Rangan, P. Schniter, and A. Fletcher, "Vector approximate message passing," *arXiv preprint arXiv: 1610.03082*, 2016.

[27] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *arXiv preprint arXiv: 1701.05284*, 2017.