The dynamics of message passing on dense graphs, with applications to compressed sensing

Mohsen Bayati
Department of Electrical Engineering
Stanford University

Andrea Montanari
Departments of Electrical Engineering and Statistics
Stanford University

***Abstract*—'Approximate message passing' algorithms proved to be extremely effective in reconstructing sparse signals from a small number of incoherent linear measurements. Extensive numerical experiments further showed that their dynamics is accurately tracked by a simple one-dimensional iteration termed *state evolution*. In this paper we provide the first rigorous foundation to state evolution. We prove that indeed it holds asymptotically in the large system limit for sensing matrices with iid gaussian entries.**

While our focus is on message passing algorithms for compressed sensing, the analysis extends beyond this setting, to a general class of algorithms on dense graphs. In this context, state evolution plays the role that density evolution has for sparse graphs.

I. Introduction and main results

Given an $n \times N$ matrix A, the compressed sensing reconstruction problem requires to reconstruct a sparse vector $x_0 \in \mathbb{R}^N$ from a (small) vector of linear observations $y = Ax_0 \in \mathbb{R}^n$. Recently [DMM09] suggested the following first order *approximate message-passing (AMP)* algorithm for reconstructing x_0 given A, y. Start with an initial guess $x^0 = 0$ and proceed by

$$x^{t+1} = \eta_t(A^* z^t + x^t), \tag{I.1}$$
$$z^t = y - Ax^t + \frac{1}{\delta} z^{t-1} \langle \eta'_{t-1}(A^* z^{t-1} + x^{t-1}) \rangle \,,$$

for an appropriate sequence of threshold functions $\{\eta_t\}_{t\geq 0}$. The goal is to show that x^t converges to x_0 (cf. [DMM09] for details). Here we assume that the columns of A have ℓ_2 norm (approximately) equal to 1, and, given a vector $v \in \mathbb{R}^N$ we write $f(x)$ for the vector obtained by applying f componentwise. Further, $\delta = n/N$, $\langle v \rangle \equiv N^{-1} \sum_{i=1}^N v_i$ and A^* is the transpose of matrix A.

Two type of findings were presented in [DMM09]: (1) For random or pseudo-random matrices A, the behavior of AMP algorithms is accurately described by the so called 'state evolution' (SE) formalism; (2) The sparsity-undersampling tradeoff of AMP as derived from SE coincides, for an appropriate choice of the functions η_t, with the one of (much more complex) convex optimization approaches.

These findings were based on heuristic arguments and extensive numerical simulations. In this paper we provide the first rigorous support to finding (1), by proving that SE holds in the large system limit, for a special class of sensing matrices.

Note that AMP is an approximation to the following message-passing algorithm. For all $i, j \in [N]$ and $a, b \in [n]$ (here and below $[N] \equiv \{1, 2, \dots, N\}$) start with messages $x^0_{j\to a} = 0$ and proceed by

$$z^t_{a\to i} = y_a - \sum_{j\in[N]\setminus i} A_{aj} x^t_{j\to a} \,, \tag{I.2}$$
$$x^{t+1}_{i\to a} = \eta_t\Big(\sum_{b\in[n]\setminus a} A_{bi} z^t_{b\to i}\Big).$$

As argued in [DMM10], AMP accurately approximates message passing in the large system limit. An important tool for the analysis of message passing algorithms is provided by density evolution [RU08]. Density evolution is known to hold asymptotically for sequences of sparse graphs that are locally tree-like. The factor graph underlying the algorithm (I.2) is dense: indeed it is the complete bipartite graph. State evolution plays the role of density evolution for dense graphs, and can be regarded (in a very precise sense) as the limit of density evolution for dense graphs.

For the sake of concreteness, we will focus on the algorithm (I.1). Nevertheless our analysis applies to a much larger family of message passing algorithms on dense graphs, for instance the multi-user detection algorithm studied in [Kab03], [NS05], [MT06]. It is important to mention that the algorithms (I.1) and (I.2) are completely different from gaussian belief propagation (BP). More generally, none of the existing rigorous results for BP can be used here.

It is truly remarkable that density evolution (in its special incarnation, SE) holds for dense graphs. This upsets a very popular piece of wisdom: 'density evolution (and message passing) works *because* the graph is locally tree-like, and does not work on graphs with many short loops.'

A. Main result

Given a probability distribution p_{X_0}, let $\tau_0^2 \equiv \mathbb{E}\{X_0^2\}/\delta$, and define recursively for $t \geq 0$,

$$\tau^2_{t+1} = \frac{1}{\delta}\mathbb{E}\left\{[\eta_t(X_0 + \tau_t Z) - X_0]^2\right\} \,, \tag{I.3}$$

with $X_0 \sim p_{X_0}$ and $Z \sim \mathsf{N}(0,1)$ independent. Also a function $\phi : \mathbb{R}^m \to \mathbb{R}$ is called *pseudo-Lipschitz* if there exist a constant L such that for all $x, y \in \mathbb{R}$, $|\phi(x) - \phi(y)| \leq \max(\|x\|, \|y\|, L)\|x - y\|$.

Theorem 1. *Let $\{A(N)\}_N$ be a sequence of sensing matrices $A \in \mathbb{R}^{n\times N}$ indexed by N, with iid entries $A_{ij} \sim \mathsf{N}(0, 1/n)$, and assume $n/N \to \delta \in (0,\infty)$. Consider further a sequence of signals $\{x_0(N)\}_N$, whose empirical distributions converges weakly to a probability measure p_{X_0} on $\mathbb{R}$, and have uniformly bounded fourth moment. Then, for any pseudo-Lipschitz function $\psi : \mathbb{R}^2 \to \mathbb{R}$ and all t, almost surely*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \psi(x_i^t, x_{0,i}) = \mathbb{E}[\psi(\eta_{t-1}(X_0 + \tau_{t-1}Z), X_0)]\,. \quad \text{(I.4)}$$

Up to a trivial change of variables, this is a formalization of the findings of [DMM09] (cf. in particular Eqs. [7], [8] and Finding 2 in that paper).

Note 1. *The empirical distribution of the vector $x_0 \in \mathbb{R}^N$ is the probability distribution that puts a point mass $1/N$ at each of the N entries of the vector.*

B. Alternative representation of AMP

Let $h^t = x_0 - (A^* z^{t-1} + x^{t-1})$, $f(x) = f(x, x_0) = x_0 - \eta(x_0 - x)$, and $g(x) = -x$. Also define $m^t = g(z^t)$, $q^t = f(h^t)$, and $\lambda_t = \frac{1}{\delta}\langle f'(h^t)\rangle$. Therefore, we will obtain the following equivalent version of AMP. Start with $z^0 = Ax_0$ (or $\lambda_0 = 0, q^0 = x_0$) and proceed by

$$\begin{aligned} h^{t+1} &= A^* m^t + q^t \\ z^t &= Aq^t - \lambda_t m^{t-1} \end{aligned} \quad \text{(I.5)}$$

Note 2. *(a) It is simple to see that algorithms* (I.5) *and* (I.1) *are equivalent with a simple change of variable. We only use* (I.5) *to simplify the analysis of* (I.1)*. However, since $q_0 = x_0$ is unknown in practice, one should use* (I.1) *to recover x_0.*

(b) Due to symmetry, for each t, all coordinates of the vector h^t have the same distribution (z^t, q^t and m^t are similar).

(c) Our proof holds for all systems of algorithms of the type (I.5) *and all functions f, g that have bounded derivative. For this general case the coefficient of q^t in the first equation should be changed from 1 to $-\langle g'(z^t)\rangle$.*

(d) Also, the proof applies to any sequence of scalar functions $\{f_t, g_t\}_{t\geq 0}$ with $h^t = f_t(q^t)$ and $m^t = g_t(z^t)$. For simplicity, we shall drop the time dependence of f, g.

II. Analysis

The proof is based on a conditioning technique developed by Erwin Bolthausen for the analysis of the so-called TAP equations in spin glass theory [Bol09]. Related ideas can also be found in [Don06].

First we introduce some new notations and then state and prove a more general result than Theorem 1.

A. Definitions

When the update equation for h^{t+1} in (I.5) is used, all values of $z^0, \dots, z^t$ and also $h^1, \dots, h^t$ have been previously calculated. Additionally any deterministic function of them ($m^0, \dots, m^t$ and $q^1, \dots, q^t$) is known as well. Hence, we can consider the distribution of h^{t+1} when it is conditioned on all these known variables and x_0. In particular, define $\mathfrak{S}_{t_1,t_2}$ to be the σ-algebra generated by $z^0, \dots, z^{t_1-1}$ (thus including $m^0, \dots, m^{t_1-1}$) and $x^0, h^1, \dots, h^{t_2}$ (thus including $q^0, \dots, q^{t_2}$). We are interested in finding the distributions of random variables $z^t|_{\mathfrak{S}_{t,t}}$ and , $h^{t+1}|_{\mathfrak{S}_{t+1,t}}$.

Since h^t, z^t are column vectors, the equations for $z^0, \dots, z^{t-1}$ and $h^1, \dots, h^t$ can be written in matrix form as:

$$\underbrace{\big[h^1 - q^0 | h^2 - q^1 | \cdots | h^t - q^{t-1}\big]}_{X_t} = A^* \underbrace{[m^0 | \dots | m^{t-1}]}_{M_t},$$

$$\underbrace{\big[z^0 | z^1 + \lambda_1 m^0 | \cdots | z^{t-1} + \lambda_{t-1} m^{t-2}\big]}_{Y_t} = A \underbrace{[q^0 | \dots | q^{t-1}]}_{Q_t}.$$

or in short $Y_t = AQ_t$ and $X_t = A^* M_t$. For each matrix M we define $\widehat{M} \equiv M^* M$.

We also introduce the notation $m^t_{\shortparallel}$ for the projection of m^t onto column space of M_t and define $m^t_\perp = m^t - m^t_{\shortparallel}$. Similarly, define $q^t_{\shortparallel}, q^t_\perp$ to be the parallel and orthogonal projections of q^t onto column space of Q_t.

For vectors $u, v \in \mathbb{R}^m$ define $\langle u\rangle = \sum_{i=1}^m u_i/m$ and $\langle u, v\rangle = \sum_{i=1}^m u_i v_i/m$. For random variables X, Y the notion $X \overset{\text{a.s.}}{=} Y$ means that X and Y are equal almost surely, $X \overset{\text{d}}{=} Y$ that they are equal in distribution.

B. Main technical Lemma

We prove the following more general result.

Lemma 1. *Let $\{A(N)\}$ be a sequence of sensing matrices as in Theorem 1, with $n/N = \delta$. Assume x_0 to have i.i.d. entries with distribution p_{X_0}, having finite fourth moment. Then the following hold for all $t \in \mathbb{N} \cup \{0\}$*

(a)

$$h^{t+1}|_{\mathfrak{S}_{t+1,t}} \overset{\text{d}}{=} \sum_{i=0}^{t-1} \alpha_i h^{i+1} + \tilde{A}^* m^t_\perp + Q_t \vec{o}_t(1) \quad \text{(II.1)}$$

$$z^t|_{\mathfrak{S}_{t,t}} \overset{\text{d}}{=} \sum_{i=0}^{t-1} \beta_i z^i + \tilde{A} q^t_\perp + M_t \vec{o}_t(1) \quad \text{(II.2)}$$

where $\tilde{A}$ is an independent copy of A and coefficients α_i, β_j satisfy $m^t_{\shortparallel} = \sum_{i=0}^{t-1} \alpha_i m^i$ and $q^t_{\shortparallel} = \sum_{i=0}^{t-1} \beta_i q^i$. Here $\vec{o}_t(1) \in \mathbb{R}^t$ is a finite dimensional random vector that converges to 0 almost surely as $N \to \infty$.

(b) For any pseudo-Lipschitz function $\phi : \mathbb{R}^{t+1} \to \mathbb{R}$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \phi(h_i^1, \dots, h_i^{t+1}, x_{0,i}) \overset{\text{a.s.}}{=} \mathbb{E}\big[\phi(\tau_0 Z_0, \dots, \tau_t Z_t, X_0)\big] \quad \text{(II.3)}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \phi(z_i^0, \dots, z_i^t) \overset{\text{a.s.}}{=} \mathbb{E}\big[\phi(\tau_0 \hat{Z}_0, \dots, \tau_t \hat{Z}_t)\big], \quad \text{(II.4)}$$

where $Z_0, \dots, Z_t$ ($\hat{Z}_0, \dots, \hat{Z}_t$) have $\mathsf{N}(0,1)$ distribution and are independent of X_0.

(c) For all $0 \le r, s \le t$ the following equations hold and all limits exist, are bounded and non-random.

$$\lim_{N\to\infty} \langle h^{r+1}, h^{s+1}\rangle \stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \langle m^r, m^s\rangle\,, \tag{II.5}$$

$$\lim_{n\to\infty} \langle z^r, z^s\rangle \stackrel{\text{a.s.}}{=} \frac{1}{\delta}\lim_{n\to\infty} \langle q^r, q^s\rangle\,. \tag{II.6}$$

(d) For all $0 \le r, s \le t$, and for any differentiable function φ with bounded first derivative, the following equations hold and all limits exist, are bounded and non-random.

$$\lim_{N\to\infty} \langle h^{r+1}, \varphi(h^{s+1})\rangle \stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \langle h^{r+1}, h^{s+1}\rangle\langle \varphi'(h^{s+1})\rangle, \tag{II.7}$$

$$\lim_{N\to\infty} \langle z^r, \varphi(z^s)\rangle \stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \langle z^r, z^s\rangle\langle \varphi'(z^s)\rangle\,. \tag{II.8}$$

Note 3. (a) *Above and in the following $X|_{\mathfrak{S}} \stackrel{\text{d}}{=} Y$ means that for any integrable function ϕ and for any random variable Z measurable on $\mathfrak{S}$, $\mathbb{E}\{\phi(X)Z\} = \mathbb{E}\{\phi(Y)Z\}$.*
(b) *Eqs. (II.7) and (II.8) have the form of Stein's lemma [Ste72] (Lemma 6 in our Section II-D).*

C. Proof of Theorem 1

Consider first the case in which $n/N = \delta$ and x^0 has iid entries with distribution p_{X_0}. By definition $x^{t+1} = \eta(A^*z^t + x^t) = \eta(x_0 - h^{t+1})$. Therefore, applying Lemma 1(b) to the function $\phi(y_0, \dots, y_t, x_{0,i}) = \psi(\eta(x_{0,i} - y_t), x_{0,i})$ we obtain

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \psi(x_i^t, x_{0,i}) \stackrel{\text{a.s.}}{=} \mathbb{E}\left\{\psi\big[\eta(X_0 - \tau_{t-1}Z), X_0\big]\right\}$$

with $Z \sim \mathsf{N}(0,1)$ independent of X_0, which yields the claim as $Z \stackrel{\text{d}}{=} -Z$.

Let us sketch the generalization to a deterministic sequence of vectors $\{x_0(N)\}_{N\ge 1}$ with converging empirical distribution. Notice that, by symmetry, we can replace $x_0(N)$ by the random vector obtained by uniformly permuting its coordinates. Let us keep the notation $x_0(N)$ for this vector. Let $x_0'(N)$ be the random vector whose entries are iid with distribution p_{X_0}. It is possible to construct a coupling such that $\mathbb{E}\{||x_0 - x_0'||^2\} = o(N)$. The proof is completed by showing that this implies $\mathbb{E}\{||x^t - (x^t)'||^2\} = o(N)$ for all t. The generalization to $n/N \to \delta$ is proved analogously. ∎

D. Useful properties

In order to calculate $z^t|_{\mathfrak{S}_{t,t}}, h^{t+1}|_{\mathfrak{S}_{t+1,t}}$ we find $A|_{\mathfrak{S}_{t,t}}, A|_{\mathfrak{S}_{t+1,t}}$.

Lemma 2. *Let $(t_1, t_2) = (t, t)$ or $(t_1, t_2) = (t+1, t)$. Then the distribution of the conditional random variable $A|_{\mathfrak{S}_{t_1,t_2}}$ satisfies*

$$A|_{\mathfrak{S}_{t_1,t_2}} \stackrel{\text{d}}{=} E_{t_1,t_2} + \mathcal{P}_{V_{t_1,t_2}}(\tilde{A}). \tag{II.9}$$

Here $\tilde{A} \stackrel{\text{d}}{=} A$ is a random matrix independent of $\mathfrak{S}_{t_1,t_2}$. Also, $E_{t_1,t_2} = \mathbb{E}(A|\mathfrak{S}_{t_1,t_2})$ is equal to

$$E_{t_1,t_2} = Y_{t_1}(\widehat{Q}_{t_1})^{-1}Q_{t_1}^* + M_{t_2}(\widehat{M}_{t_2})^{-1}X_{t_2}^* - M_{t_2}(\widehat{M}_{t_2})^{-1}M_{t_2}^*Y_{t_1}(\widehat{Q}_{t_1})^{-1}Q_{t_1}^*\,. \tag{II.10}$$

Further, $\mathcal{P}_{V_{t_1,t_2}}$ is the orthogonal projector onto subspace $\mathrm{V}_{t_1,t_2} = \{\tilde{A}|AQ_{t_1} = 0, A^*M_{t_2} = 0\}$, defined by $\mathcal{P}_{V_{t_1,t_2}}(\tilde{A}) = P_{M_{t_2}}^{\perp}\tilde{A}P_{Q_{t_1}}^{\perp}$. Here $P_{M_{t_2}}^{\perp} = I - P_{M_{t_2}}$, $P_{Q_{t_1}}^{\perp} = I - P_{Q_{t_1}}$, and $P_{Q_{t_1}}$, $P_{M_{t_2}}$ are orthogonal projector onto column spaces of Q_{t_1} and M_{t_2} respectively.

Recall the following well-known formula.

Lemma 3. *Let $z \in \mathbb{R}^n$ be a random vector of iid $\mathsf{N}(0,\alpha)$ variables and let $\mathrm{D} \in \mathbb{R}^{m\times n}$ be a linear operator. Then for any constant vector $b \in \mathbb{R}^m$ the distribution of z conditioned on $\mathrm{D}z = b$ satisfies:*

$$z|_{\mathrm{D}z=b} \stackrel{\text{d}}{=} \mathrm{D}^*(\mathrm{D}\mathrm{D}^*)^{-1}b + \mathrm{P}_{\mathrm{D}z=0}(\tilde{z})$$

where $\mathrm{P}_{\{\mathrm{D}z=0\}}$ is the orthogonal projection onto the subspace $\{\mathrm{D}z = 0\}$ and $\tilde{z}$ is a random vector of iid $\mathsf{N}(0,\alpha)$. Moreover, $\mathrm{D}^*(\mathrm{D}\mathrm{D}^*)^{-1}b = \arg\min_z \left\{\|z\|^2 | \mathrm{D}z = b\right\}$.

Lemma 2 follows from applying Lemma 3 to the operator D that maps A to (AQ, M^*A). Note that for finite values of t as $N \to \infty$ the matrices $\widehat{M}_t$ and $\widehat{Q}_t$ are non-singular almost surely. To the interest of space we leave a detailed proof of Lemma 2 to a longer version of this paper.

Lemma 4. *The following holds*

$$E_{t+1,t}^* m^t \stackrel{\text{a.s.}}{=} X_t(\widehat{M}_t)^{-1}M_t^* m_{\shortparallel}^t + Q_{t+1}(\widehat{Q}_{t+1})^{-1}Y_{t+1}^* m_{\perp}^t, \tag{II.11}$$

$$E_{t,t}q^t \stackrel{\text{a.s.}}{=} Y_t(\widehat{Q}_t)^{-1}Q_t^* q_{\shortparallel}^t + M_t(\widehat{M}_t)^{-1}X_t^* q_{\perp}^t. \tag{II.12}$$

Proof: Writing $m^t = m_{\shortparallel}^t + m_{\perp}^t$ and using (II.10) and the fact that $M_t^* m_{\perp}^t = 0$, we obtain $E_{t+1,t}^* m_{\perp}^t = Q_{t+1}(\widehat{Q}_{t+1})^{-1}Y_{t+1}^* m_{\perp}^t$. On the other hand let $m_{\shortparallel}^t = \sum_{i=0}^{t-1}\alpha_i m^i = M_t\vec{\alpha}$. Then using $A^*M_t = X_t$, (II.9), and $[\mathcal{P}_{V_{t_1,t_2}}(\tilde{A})]^* m_{\shortparallel}^t = 0$ we have, conditionally on $\mathfrak{S}_{t+1,t}$, $E_{t+1,t}^* m_{\shortparallel}^t \stackrel{\text{d}}{=} A^* m_{\shortparallel}^t \stackrel{\text{d}}{=} A^*M_t\vec{\alpha} \stackrel{\text{d}}{=} X_t\vec{\alpha} \stackrel{\text{d}}{=} X_t(M_t^*M_t)^{-1}M_t^*M_t\vec{\alpha} \stackrel{\text{d}}{=} X_t(M_t^*M_t)^{-1}M_t^* m_{\shortparallel}^t$. Since all sides are measurable on $\mathfrak{S}_{t+1,t}$, Eq. (II.11) follows.

Similarly, use $q^t = q_{\shortparallel}^t + q_{\perp}^t$, $q_{\shortparallel}^t = Q_t\vec{\beta}$ and $Q_t^* q_{\perp}^t = 0$ to obtain (II.12). ∎

We will also use the following strong law of large numbers (SLLN) which follows from [HT97][Theorem 2.1].

Theorem 2 (SLLN, [HT97])**.** *Let $\{X_{n,i} : 1 \le i \le n, n \ge 1\}$ be an array of random variables with $(X_{n,1}, \dots, X_{n,n})$ mutually independent with mean equal zero for each n and $\mathbb{E}|X_{n,i}|^{2+\kappa} \le C$ for some $\kappa > 0$, $C < \infty$. Then $\frac{1}{n}\sum_{i=1}^n X_{i,n} \to 0$ a.s. for $n \to \infty$.*

Next, we present a standard property of Gaussian matrices without proof.

Lemma 5. *For any deterministic $u \in \mathbb{R}^N$ and ,$v \in \mathbb{R}^n$ with $\|u\| = \|v\| = 1$ and a gaussian matrix $\tilde{A}$ distributed as A we have $v^*\tilde{A}u \stackrel{\text{d}}{=} Z/\sqrt{n}$ where $Z \sim \mathsf{N}(0,1)$ and $\lim_{n\to\infty}\|\tilde{A}u\|^2 = 1$ almost surely.*

Lemma 6 (Stein's Lemma [Ste72])**.** *For jointly gaussian random variables Z_1, Z_2 and any C^1 function $\varphi : \mathbb{R} \to \mathbb{R}$ the following holds* $\mathbb{E}[Z_1\varphi(Z_2)] = \mathrm{Cov}(Z_1, Z_2)\mathbb{E}[\varphi'(Z_2)]$.

E. Proof of Lemma 1

The proof is by induction on t. Let $\mathcal{H}_{t+1}$ be the property that (II.1), (II.3), (II.5) and (II.7) are correct. Similarly, let $\mathcal{Z}_t$ be the property that (II.2), (II.4), (II.6) and (II.8) hold. The inductive proof consists of the following three main steps. (1) $\mathcal{Z}_0$ holds. (2) If $\mathcal{Z}_r$, $\mathcal{H}_s$ hold for all $r < t$ and $s \le t$ then $\mathcal{Z}_t$ holds. (3) If $\mathcal{Z}_r$, $\mathcal{H}_s$ hold for all $r \le t$ and $s \le t$ then $\mathcal{H}_{t+1}$ holds.

Step 1: $\mathcal{Z}_0$. Note that $z^0 = Ax_0$.

(a) $\mathfrak{S}_{0,0}$ is generated by $q^0 = x_0$. Also $q^0 = q^0_\perp$ since Q_0 is an empty matrix. Hence $z^0|_{\mathfrak{S}_{0,0}} = Ax_0 = Aq^0_\perp$.

(b) Let $\phi : \mathbb{R} \to \mathbb{R}$ be a pseudo-Lipschitz function with constant L, and assume w.l.o.g. $\phi(0) = 0$. Conditioning on $q^0 = x_0$, the random variable $z^0 = \sum_{i=1}^n \phi((Ax_0)_i)/n$ is a sum of iid random variables. By Lemma 5 $(Ax_0)_i \stackrel{\mathrm{d}}{=} Z||x_0||/\sqrt{n}$ for $Z \sim \mathsf{N}(0,1)$. By the SLLN: $\lim_{n\to\infty} ||x_0||^2/n \stackrel{\text{a.s.}}{=} \mathbb{E}(X_0^2)/\delta = \tau_0^2 < \infty$. Hence, for all $p \ge 2$, there exist a constant C_p such that $\mathbb{E}|(Ax_0)_i|^p < C_p$. Therefore $\mathbb{E}|\phi([Ax_0]_i)|^3 \le \max(L^3\mathbb{E}|(Ax_0)_i|^3, \mathbb{E}|(Ax_0)_i|^6) \le C$ for a constant C. Now, we can apply Theorem 2 to get $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n [\phi(z_i^0) - \mathbb{E}_A\phi(z_i^0)] \stackrel{\text{a.s.}}{=} 0$ whence, by the above calculation

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \phi(z_i^0) \stackrel{\text{a.s.}}{=} \mathbb{E}\big[\phi(\tau_0 Z)\big].$$

(c) Using Lemma 5, $\lim_{n\to\infty}\langle z^0, z^0\rangle$=$\lim_{n\to\infty} \|Ax_0\|^2/n \stackrel{\text{a.s.}}{=} \lim_{N\to\infty}\langle q^0, q^0\rangle/\delta \stackrel{\text{a.s.}}{=} \mathbb{E}(X_0^2)/\delta$.

(d) Using part (a) for $t = 0$, and $\phi(x) = xg(x)$ we obtain $\lim_{n\to\infty}\langle z^0, g(z^0)\rangle \stackrel{\text{a.s.}}{=} \mathbb{E}(\tau_0\hat{Z}g(\tau_0\hat{Z}))$, which is equal to $\tau_0^2\mathbb{E}[g'(\tau_0\hat{Z})]$ using Lemma 6. On the other hand, in proof of (b) we showed $\lim_{n\to\infty}\langle z^0, z^0\rangle \stackrel{\text{a.s.}}{=} \tau_0^2$. Now, applying part (b) to $\phi(x) = g'(x)$ we get $\lim_{n\to\infty}\langle g'(z^0)\rangle \stackrel{\text{a.s.}}{=} \mathbb{E}[g'(\tau_0\hat{Z})]$. ∎

Step 2: $\mathcal{Z}_t$. This part is analogous to step 1 albeit more complex.

(a) Note that

$$Y_t = Z_t + [0|M_{t-1}]\Lambda_t, \quad X_t = H_t - Q_t, \tag{II.13}$$

where $Z_t = [z^0|\cdots|z^{t-1}]$, $\Lambda_t = \mathrm{diag}(\lambda_0, \ldots, \lambda_{t-1})$ and $H_t = [h^1|\cdots|h^t]$.

Lemma 7. *The following holds*
(a) $h^{t+1}|_{\mathfrak{S}_{t+1,t}} \stackrel{\mathrm{d}}{=} H_t\widehat{M}_t^{-1}M_t^*m_{\shortparallel}^t + P^\perp_{Q_{t+1}}\tilde{A}^*P^\perp_{M_t}m^t + Q_t\vec{o}_t(1)$.
(b) $z^t|_{\mathfrak{S}_{t,t}} \stackrel{\mathrm{d}}{=} Z_t(\widehat{Q}_t)^{-1}Q_t^*q_{\shortparallel}^t + P^\perp_{M_t}\tilde{A}P^\perp_{Q_t}q^t + M_t\vec{o}_t(1)$.

Proof: In light of Lemmas 2 and 4 we have $h^{t+1}|_{\mathfrak{S}_{t+1,t}} \stackrel{\mathrm{d}}{=} X_t(M_t^*M_t)^{-1}M_t^*m_{\shortparallel}^t + Q_{t+1}(\widehat{Q}_{t+1})^{-1}Y_{t+1}^*m_\perp^t + P^\perp_{Q_{t+1}}\tilde{A}P^\perp_{M_t}m^t + q^t$ and $z^t|_{\mathfrak{S}_{t,t}} \stackrel{\mathrm{d}}{=} Y_t(\widehat{Q}_t)^{-1}Q_t^*q_{\shortparallel}^t + M_t(\widehat{M}_t)^{-1}X_t^*q_\perp^t + P^\perp_{M_t}\tilde{A}P^\perp_{Q_t}q^t - \lambda_t m^{t-1}$. Now using equations (II.13), we only need to show $-Q_t(\widehat{M}_t)^{-1}M_t^*m_{\shortparallel}^t + Q_{t+1}(\widehat{Q}_{t+1})^{-1}Y_{t+1}^*m_\perp^t + q^t = Q_t\vec{o}_t(1)$ and $[0|M_{t-1}]\Lambda_t(\widehat{Q}_t)^{-1}Q_t^*q_{\shortparallel}^t + M_t(\widehat{M}_t)^{-1}X_t^*q_\perp^t - \lambda_t m^{t-1} = M_t\vec{o}_t(1)$. Recall that $m_{\shortparallel}^t = M_t\vec{\alpha}$ and $q_{\shortparallel}^t = Q_t\vec{\beta}$. On the other hand $Y_{t+1}^*m_\perp^t = Z_{t+1}^*m_\perp^t$ because $M_tm_\perp^t = 0$. Similarly, $X_t^*q_\perp^t = H_t^*q_\perp^t$. Hence we need to show

$$-Q_t\vec{\alpha} + Q_{t+1}(\widehat{Q}_{t+1})^{-1}Z_{t+1}^*m_\perp^t + q^t = Q_t\vec{o}_t(1) \tag{II.14}$$
$$[0|M_{t-1}]\Lambda_t\vec{\beta} + M_t(\widehat{M}_t)^{-1}H_t^*q_\perp^t - \lambda_t m^{t-1} = M_t\vec{o}_t(1). \tag{II.15}$$

Here is our strategy to prove (II.15) (proof of (II.14) is similar). The left hand side is a linear combination of vectors $m^0, \ldots, m^{t-1}$. For any $\ell = 1, \ldots, t$ we will prove that the coefficient of each $m^{\ell-1}$ converges to 0. This coefficient in the left hand side is equal to $\left[(\widehat{M}_t)^{-1}H_t^*q_\perp^t\right]_\ell - \lambda_\ell(-\beta_\ell)^{\mathbb{I}_{\ell\neq t}}$ which can be written as $\sum_{r=1}^t \delta^{-1}\left[(\widehat{M}_t/n)^{-1}\right]_{\ell,r}\langle h^r, q^t - \sum_{s=0}^{t-1}\beta_s q^s\rangle - \lambda_\ell(-\beta_\ell)^{\mathbb{I}_{\ell\neq t}}$. To simplify the notation denote the matrix $\widehat{M}_t/n$ by G. Therefore $\lim_{N\to\infty}$ Coefficient of $m^{\ell-1}$ is equal to

$$\lim_{N\to\infty}\Big\{\sum_{r=1}^t (G^{-1})_{\ell,r}\langle h^r, q^t - \sum_{s=0}^{t-1}\beta_s q^s\rangle\frac{1}{\delta} - \lambda_\ell(-\beta_\ell)^{\mathbb{I}_{\ell\neq t}}\Big\}.$$

But using induction hypothesis $\mathcal{H}_t(d)$, the term $\langle h^r, q^t - \sum_{s=0}^{t-1}\beta_s q^s\rangle/\delta$ is almost surely equal to the limit of $\langle h^r, h^t\rangle\lambda_t - \sum_{s=0}^{t-1}\beta_s\langle h^r, h^s\rangle\lambda_s$. This can be modified, using induction hypothesis $\mathcal{H}_t(c)$, to $\langle m^{r-1}, m^{t-1}\rangle\lambda_t - \sum_{s=0}^{t-1}\beta_s\langle m^{r-1}, m^{s-1}\rangle\lambda_s$ almost surely, which can be written as $G_{r,t}\lambda_t - \sum_{s=0}^{t-1}\beta_s G_{r,s}\lambda_s$. Hence

$$\begin{aligned}&\lim_{N\to\infty}\text{ Coefficient of } m^{\ell-1} \stackrel{\text{a.s.}}{=}\\ &\lim_{N\to\infty}\Big\{\sum_{r=1}^t (G^{-1})_{\ell,r}[G_{r,t}\lambda_t - \sum_{s=0}^{t-1}\beta_s G_{r,s}\lambda_s] - \lambda_\ell(-\beta_\ell)^{\mathbb{I}_{\ell\neq t}}\Big\}\\ &\stackrel{\text{a.s.}}{=} \lim_{N\to\infty}\{\lambda_t\mathbb{I}_{t=\ell} - \sum_{s=0}^{t-1}\beta_s\lambda_s\mathbb{I}_{\ell=s} - \lambda_\ell(-\beta_\ell)^{\mathbb{I}_{\ell\neq t}}\} \stackrel{\text{a.s.}}{=} 0\end{aligned}$$

Similarly, using $g'(x) = -1$, (II.14) can be shown by

$$\lim_{N\to\infty}\big\{[(\widehat{Q}_{t+1})^{-1}Z_{t+1}^*m_\perp^t]_\ell + (-\alpha_\ell)^{\mathbb{I}_{\ell\neq t}}\big\} \stackrel{\text{a.s.}}{=} 0.$$

∎

(c) For $r, s < t$ we can use induction hypothesis. For $s = t, r < t$, using (II.2) for t that was just proved,

$$\langle z^t, z^r\rangle|_{\mathfrak{S}_{t,t}} \stackrel{\mathrm{d}}{=} \sum_{i=0}^{t-1}\beta_i\langle z^i, z^r\rangle + \langle\tilde{A}q_\perp^t, z^r\rangle + \sum_{i=0}^{t-1}o(1)\langle m^i, z^r\rangle.$$

Now, by induction hypothesis $\mathcal{Z}_{t-1}(d)$ each term $\langle m^i, z^r\rangle$ has a finite limit. Thus, $\lim_{n\to\infty}\sum_{i=0}^{t-1}o(1)\langle m^i, z^r\rangle \stackrel{\text{a.s.}}{=} 0$. Now we can use induction hypothesis $\mathcal{Z}_r(c)$ or $\mathcal{Z}_i(c)$ for each term of the form $\langle z^i, z^r\rangle$ and use Lemma 5 for $\langle\tilde{A}q_\perp^t, z^i\rangle$ to obtain

$$\begin{aligned}\lim_{n\to\infty}\langle z^t, z^r\rangle &\stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\frac{1}{\delta}\sum_{i=0}^{t-1}\beta_i\langle q^i, q^r\rangle\\ &\stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\frac{1}{\delta}\langle q_{\shortparallel}^t, q^r\rangle \stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\frac{1}{\delta}\langle q^t, q^r\rangle.\end{aligned}$$

Last line uses the definition of β_i and $q_\perp^t \perp q^r$.

For the case of $r = s = t$, we have

$$\langle z^t, z^t\rangle|_{\mathfrak{S}_{t,t}} \stackrel{\mathrm{d}}{=} \sum_{i,j=0}^{t-1} \beta_i\beta_j\langle z^i, z^j\rangle + \langle \tilde{A}q_\perp^t, \tilde{A}q_\perp^t\rangle + o(1).$$

Note that we used similar argument (Lemma 5 and $\mathcal{Z}_{t-1}(c)$) to show the contribution of all products of the form $\langle M_t\vec{o}_t(1), \cdot\rangle$ and $\langle \tilde{A}q_\perp^t, z^i\rangle$ a.s. tend to 0. Now, using induction hypothesis and Lemma 5

$$\begin{aligned}\lim_{n\to\infty}\langle z^t, z^t\rangle|_{\mathfrak{S}_{t,t}} &\stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\sum_{i,j=0}^{t-1}\beta_i\beta_j\frac{\langle q^i, q^j\rangle}{\delta} + \lim_{n\to\infty}\|q_\perp^t\|^2\\ &\stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\frac{\langle q_{\shortparallel}^t, q_{\shortparallel}^t\rangle}{\delta} + \lim_{n\to\infty}\frac{\langle q_\perp^t, q_\perp^t\rangle}{\delta}\\ &\stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\frac{\langle q^t, q^t\rangle}{\delta}.\end{aligned}$$

(b) Using part (a) we can write $\phi(z_i^0, \ldots, z_i^t)|_{\mathfrak{S}_{t,t}}$ as $\phi(z_i^0, \ldots, z_i^{t-1}, [\sum_{r=0}^{t-1}\beta_r z^r + \tilde{A}q_\perp^t + M_t\vec{o}_t(1)]_i)$. First we would like to drop the error term $M_t\vec{o}_t(1)$. For simplicity let $a_{i,n} = (z_i^0, \ldots, z_i^{t-1}, z_i^t|_{\mathfrak{S}_{t,t}})$ and $b_{i,n} = (z_i^0, \ldots, z_i^{t-1}, [\sum_{r=0}^{t-1}\beta_r z^r + \tilde{A}q_\perp^t]_i)$. Since ϕ is pseudo-Lipschitz with constant L we have $|\phi(a_{n,i}) - \phi(b_{n,i})| \le \max(L, \|a_{i,n}\|, \|b_{i,n}\|)|\sum_{r=0}^{t-1} m_i^r|o(1)$. Therefore, the difference $n^{-1}|\sum_{i=1}^n\phi(a_{n,i}) - \sum_{i=1}^n\phi(b_{n,i})|$ is less than

$$\big[\max(L, \sum_{i=1}^n\frac{\|a_{i,n}\|^2}{n}, \frac{\sum_{i=1}^n\|b_{i,n}\|^2}{n})\big]^{\frac{1}{2}}\big[\sum_{r=0}^{t-1}t^{\frac{1}{2}}\langle m^r, m^r\rangle\big]^{\frac{1}{2}}o(1), \tag{II.16}$$

using Cauchy-Schwartz inequality twice. Also note that $n^{-1}\sum_{i=1}^n\|a_{i,n}\|^2 \le \sum_{r=0}^{t-1}\langle z^r, z^r\rangle + \langle z^t, z^t\rangle|_{\mathfrak{S}_{t,t}}$ which is finite almost surely, using induction hypothesis and part (c) that was just proven. Similarly $n^{-1}\sum_{i=1}^n\|a_{i,n}\|^2$ and $\sum_{r=0}^{t-1}\langle m^r, m^r\rangle$ are finite. Hence for any finite t, (II.16) goes to 0 almost surely when n goes to ∞.

Now given, $z^0, \ldots, z^{t-1}$, consider the random variables $\tilde{X}_{i,n} = \phi(z_i^0, \ldots, z_i^{t-1}, \sum_{r=0}^{t-1}\beta_r z_i^r + (\tilde{A}q_\perp^t)_i)$ and $X_{i,n} \equiv \tilde{X}_{i,n} - \mathbb{E}_{\tilde{A}}\tilde{X}_{i,n}$. Proceeding as in Step 1, and using the pseudo-Lipschitz property of ϕ, it is easy to check the conditions of Theorem 2. We therefore get

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\phi(z_i^0, \ldots, [\sum_{r=0}^{t-1}\beta_r z^r + \tilde{A}q_\perp^t]_i)&\\ \stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\mathbb{E}_{\tilde{A}}\phi(z_i^0, \ldots, [\sum_{r=0}^{t-1}\beta_r z^r + \tilde{A}q_\perp^t]_i).&\end{aligned}$$

Note that $[\tilde{A}q_\perp^t]_i$ is a gaussian random variable with variance $\|q_\perp^t\|^2/\delta$. Hence we can use induction hypothesis $\mathcal{Z}_{t-1}(b)$ for $\phi(z_i^0, \ldots, z^{t-1}) = \mathbb{E}_Z\phi(z_i^0, \ldots, \sum_{r=0}^{t-1}\beta_r z_i^r + \delta^{-.5}\|q_\perp^t\|Z)$ where Z is an independent $\mathsf{N}(0,1)$ random variable, to show

$$\begin{aligned}\lim_{n\to\infty}\frac{\sum_{i=1}^n\mathbb{E}_{\tilde{A}}\phi(z_i^0, \ldots, [\sum_{r=0}^{t-1}\beta_r z^r + \tilde{A}q_\perp^t]_i)}{n}&\\ \stackrel{\text{a.s.}}{=} \mathbb{E}\,\mathbb{E}_Z\phi(\tau_0 Z_0, \ldots, \sum_{r=0}^{t-1}\beta_r\tau_r Z_r + \delta^{-.5}\|q_\perp^t\|Z)&\end{aligned}$$

Note that $\sum_{r=0}^{t-1}\beta_r\tau_r Z_r + \delta^{-.5}\|q_\perp^t\|Z$ is gaussian. All that is needed is to show that the variance of this gaussian is τ_t^2. But using what we just proved for the pseudo-Lipschitz function $\phi(y_0, \ldots, y_t) = y_t^2$, we have

$$\mathbb{E}\{(\sum_{r=0}^{t-1}\beta_r\tau_r Z_r + \delta^{-.5}\|q_\perp^t\|Z)^2\} \stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\langle z^t, z^t\rangle \tag{II.17}$$

On the other hand in part (c) we proved $\lim_{n\to\infty}\langle z^t, z^t\rangle \stackrel{\text{a.s.}}{=} \lim_{n\to\infty}\delta^{-1}\langle f(h^t), f(h^t)\rangle$.

By the induction hypothesis $\mathcal{H}_t$(b) for the pseudo-Lipschitz function $\phi(y_0, \ldots, y_t) = f(y_t)^2$ we get $\lim_{n\to\infty}\delta^{-1}\langle f(h^t), f(h^t)\rangle \stackrel{\text{a.s.}}{=} \delta^{-1}\mathbb{E}(f(\tau_{t-1}Z)^2)$. So by the definition (I.3), both sides of (II.17) are equal to τ_t^2.

(d) Very similar to the proof of $\mathcal{Z}_0(d)$, using part (b) for the pseudo-Lipschitz function $\phi : \mathbb{R}^{t+1} \to \mathbb{R}$ that is given by $\phi(y_0, \ldots, y_t) = y_t g(y_s)$ we can obtain $\lim_{n\to\infty}\langle z^t, g(z^s)\rangle \stackrel{\text{a.s.}}{=} \mathbb{E}[\tau_t\hat{Z}_t g(\tau_s\hat{Z}_s)]$ for jointly gaussian $\hat{Z}_t, \hat{Z}_s$ with distribution $\mathsf{N}(0,1)$. Using Lemma 6, this is almost surely equal to $\mathrm{Cov}(\tau_t\hat{Z}_t, \tau_s\hat{Z}_s)\mathbb{E}(g'(\tau_s\hat{Z}_s))$. And another application of part (b) for $\phi(y_0, \ldots, y_t) = y_s y_t$ transforms $\mathrm{Cov}(\tau_t\hat{Z}_t, \tau_s\hat{Z}_s)$ to $\lim_{n\to\infty}\langle z^t, z^s\rangle$. Similarly, $\mathbb{E}(g'(\tau_s\hat{Z}_s))$ can be transformed to $\lim_{n\to\infty}\langle g'(z^t)\rangle$ almost surely. This finishes the proof of (d).

Step 3: $\mathcal{H}_{t+1}$. Due to symmetry, proof of this step is exactly similar to the proof of step 2.

REFERENCES

[Bol09] E. Bolthausen, *On the high-temperature phase of the sherrington-kirkpatrick model*, Seminar at EURANDOM, Eindhoven, September 2009.

[DMM09] D. L. Donoho, A. Maleki, and A. Montanari, *Message passing algorithms for compressed sensing*, Proceedings of the National Academy of Sciences **106** (2009), 18914–18919.

[DMM10] ———, *Message passing algorithms for compressed sensing: I. motivation and construction*, Proceedings of IEEE Inform. Theory Workshop (Cairo), 2010.

[Don06] D. Donoho, *For most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution*, Communications on Pure and Applied Mathematics **59** (2006), 907–934.

[HT97] T. C. Hu and R. L. Taylor, *Strong law for arrays and for the bootstrap mean and variance*, Internat. J. Math. and Math. Sci. **20** (1997), 375–383.

[Kab03] Y. Kabashima, *A cdma multiuser detection algorithm on the basis of belief propagation*, J. Phys. A **36** (2003), 11111–11121.

[MT06] A. Montanari and D. Tse, *Analysis of belief propagation for non-linear problems: the example of cdma (or: how to prove tanaka's formula)*, Proceedings of IEEE Inform. Theory Workshop (Punta de l'Este, Uruguay), 2006.

[NS05] J. Neirotti and D. Saad, *Improved message passing for inference in densely connected systems*, Europhys. Lett. **71** (2005), 866–872.

[RU08] T.J. Richardson and R. Urbanke, *Modern coding theory*, Cambridge University Press, Cambridge, 2008.

[Ste72] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, 1972.