State Evolution for General Approximate Message Passing Algorithms, with Applications to Spatial Coupling

Adel Javanmard* and Andrea Montanari †

December 30, 2012

Abstract

We consider a class of approximated message passing (AMP) algorithms and characterize their high-dimensional behavior in terms of a suitable state evolution recursion. Our proof applies to Gaussian matrices with independent but not necessarily identically distributed entries. It covers – in particular– the analysis of generalized AMP, introduced by Rangan, and of AMP reconstruction in compressed sensing with spatially coupled sensing matrices.

The proof technique builds on the one of [BM11], while simplifying and generalizing several steps.

1 Introduction

Approximate message passing (AMP) algorithms [DMM09] apply ideas from graphical models (belief propagation [Pea88]) and statistical physics (mean field or TAP equations [MPV87, MM09]) to statistical estimation. In particular AMP applies to problems that do not admit a sparse graphical model description. An AMP algorithm takes the form

$$u^t = A\, f(v^t; t) - \mathsf{b}_t\, g(u^{t-1}; t-1)\,, \tag{1}$$

$$v^{t+1} = A^{\mathsf{T}}\, g(u^t; t) - \mathsf{d}_t\, f(v^t; t)\,, \tag{2}$$

with $t \in \mathbb{N}$ being the iteration number. Here $v^t \in \mathbb{R}^n$, $u^t \in \mathbb{R}^m$ are vectors that describe the algorithm's state, $f(\,\cdot\,;t) : \mathbb{R}^n \to \mathbb{R}^n$ and $g(\,\cdot\,;t) : \mathbb{R}^m \to \mathbb{R}^m$ are sequences of functions that can be computed efficiently and b^t, d^t are scalars that also can be computed given the current state. Finally $A \in \mathbb{R}^{m\times n}$ is a matrix that is given as part of the data of the estimation problem.

One domain in which AMP finds application is the ubiquitous problem of estimating an unknown signal $x \in \mathbb{R}^n$ from noisy linear observations:

$$y = A\, x + w\,. \tag{3}$$

Here $A \in \mathbb{R}^{m\times n}$ is a known sensing matrix and $w \in \mathbb{R}^m$ is a noise vector with i.i.d. components with $\mathbb{E}w_i = 0$, $\mathbb{E}\{w_i^2\} = \sigma^2$. In [DMM09] a class of AMP algorithms was developed for this problem in the

*Department of Electrical Engineering, Stanford University

†Department of Electrical Engineering and Department of Statistics, Stanford University

compressed sensing setting in which x is sparse and $m < n$. Several generalization –for instance to signals with small total variation– were developed in [DJM11a], which also provides a more complete list of references. All of these generalizations can be recast on the form of Eqs. (1), (2) for suitable choices of the functions $f(\,\cdot\,;t)$ and $g(\,\cdot\,;t)$.

A striking property of AMP algorithms is that their high-dimensional behavior admits an *exact description*. Simplifying, for a broad range of random matrices A, the vectors u^t, v^t have asymptotically i.i.d. Gaussian entries in the limit $n, m \to \infty$ at t fixed (see next section for a formal statement). The variance of u_i^t, v_i^t can be computed through a one-dimensional recursion termed *state evolution*, because of its analogy with density evolution in coding theory [RU08]. The predictions of state evolution were tested numerically in several papers, see e.g. [DMM09, DMM11, DJM11a, Sch10, KGR11, KMS$^+$12a, SS12, JM12]. In [BM11] it was proved that state evolution does indeed hold if A has i.i.d. Gaussian entries and the functions $f(\,\cdot\,;t)$ and $g(\,\cdot\,;t)$ are Lipschitz continuous and separable[1]. This result was extended in [BLM12] to matrices A that have independent non-Gaussian entries, under the assumption that functions $f(\,\cdot\,;t)$ and $g(\,\cdot\,;t)$ are separable polynomials. On the basis of these results, it is natural to conjecture that state evolution holds for matrices with general independent entries, whenever $f(\,\cdot\,;t)$ and $g(\,\cdot\,;t)$ are separable and locally Lipschitz with polynomial growth. This conjecture is still open.

In this paper we focus on Gaussian matrices and consider a different type of generalization that was motivated by the following recent developments.

Generalized AMP. In [Ran11], Rangan proposed a class of generalized message passing algorithms (G-AMP) which found several interesting applications, see [FRVB11, KBAU12]. In particular, generalized AMP allows to tackle nonlinear estimation problems wherein $x \in \mathbb{R}^n$ is to be estimated from observations $Y = (Y_1, \dots, Y_m)$. Observations are conditionally independent given A and x, with Y_i distributed according to a model $p(\,\cdot\,|\xi_i)$ with $\xi_i = (Ax)_i$. Considering for simplicity the case in which $p(\,\cdot\,|\xi_i)$ has a density (denoted again by p), the joint density of $Y = (Y_1, \dots, Y_m)$ is therefore

$$p_Y(y|A,x) = \prod_{i=1}^{m} p\big(y_i|(Ax)_i\big)\,. \tag{4}$$

In information theory parlance, the vector (Ax) is passed through a memoryless channel with transition probability $p(\,\cdot\,|\,\cdot\,)$. From a statistics point of view, this corresponds to estimation of a generalized linear model [NW72, MN89]. The linear model (3) is recovered as the special case in which the channel is Gaussian or –more generally– the noise is purely additive. Rangan conjectured that suitable state evolution equations hold for G-AMP algorithms as well, without however providing a formal proof.

Spatial coupling. In a separate line of work, Donoho and the present authors [DJM11b] applied AMP to compressed sensing reconstruction with spatially coupled sensing matrices. This type of sensing matrices were developed in [KMS$^+$12b] (see also [KP10] for earlier work in this direction), who demonstrated heuristically the power of this approach. A mathematical analysis requires extending state evolution to matrices with independent centered Gaussian entries, although with non-identical variances (heteroscedastic entries, in the statistics terminology).

[1] Throughout the paper we say that $h : \mathbb{R}^k \to \mathbb{R}^k$ is separable if $h(x_1, x_2, \dots, x_k) = (h_1(x_1), h_2(x_2), \dots, h_k(x_k))$.

More precisely, for $A \in \mathbb{R}^{m\times n}$ we assume that the row index set $[m] = \{1, \dots, m\}$ is partitioned into q groups, and that the same holds for the column index set $[n] = \{1, \dots, n\}$. Then the entries A_{ij} are independent Gaussian with mean $\mathbb{E}\{A_{ij}\} = 0$ and variance $\mathbb{E}\{A_{ij}^2\}$ depending on the group to which i and j belong. Spatially coupled sensing matrices correspond to a special band-diagonal structure of the block variances.

A rigorous analysis of the implications of state evolution for spatially coupled matrices can be found in [DJM11b]. In particular, [DJM11b] studied a class of spatially coupled matrices, and proved that AMP reconstruction achieves the information-theoretic limit stated in [WV10]. More specifically, for sequences of spatially coupled matrices $A \in \mathbb{R}^{m\times n}$ with asymptotic undersampling rate $\delta = \lim_{n\to\infty} m/n$, AMP reconstructs the signal with high probability, provided $\delta > \overline{d}(p_X)$, where $\overline{d}(p_X)$ denotes the (upper) Rényi information dimension of p_X [Rén59]. Further, AMP reconstruction is robust to noise.

Robust regression. Bean, Bickel, El Karoui and Yu [BBEKY12] recently considered the problem of estimating the unknown vector x in the linear model (3) using robust regression. They developed exact asymptotic expressions for the risk that are analogous to the one proved in [BM12] for the Lasso. The results of [BBEKY12] are, on the other hand, based on an heuristic derivation.

The proof in [BM12] was based on the state evolution analysis of a suitable AMP algorithm whose fixed points coincide with the Lasso optima. This is suggestive of a possible approach for proving the results of [BBEKY12]: define a suitable AMP algorithm for solving the robust regression problem, and analyze it through state evolution. Indeed a comparison of the formulae in [BBEKY12] with the state evolution formulae in [Ran11] appears encouraging.

In this paper we establish a rigorous generalization of state evolution that covers all of the above developments. Applications to generalized AMP are already discussed in [Ran11], and applications to spatially coupled sensing matrices can be found in [DJM11b] and Section 3. Finally, applications to robust regression are left for future study.

Remarkably, all of the above applications can be derived by treating the following generalization of the iteration (1), (2). (A formal definition is given in the next section.)

1. The vectors $u^t \in \mathbb{R}^m$, $v^t \in \mathbb{R}^n$ are replaced by matrices $u^t \in \mathbb{R}^{m\times q}$, $v^t \in \mathbb{R}^{n\times q}$, with q kept fixed as $m, n \to \infty$.

2. The functions f, g appearing in Eqs. (1), (2) are now mappings $f(\,\cdot\,;t) : \mathbb{R}^{n\times q} \to \mathbb{R}^{n\times q}$, $g(\,\cdot\,;t) : \mathbb{R}^{m\times q} \to \mathbb{R}^{m\times q}$ that are separable across rows (e.g. the i-th row of $f(v;t)$ only depends on the i-th row on v). Correspondingly, the product $Af(v^t;t)$ has to be interpreted as a matrix multiplication.

3. The memory terms are modified with b_t, d_t replaced by $q \times q$ matrices. More specifically, $\mathsf{b}_t\, g(u^{t-1};t-1)$ and $\mathsf{d}_t\, f(v^t;t)$ are respectively replaced by $g(u^{t-1};t-1)\,\mathbf{B}_t^{\mathsf{T}}$, $f(v^t;t)\,\mathbf{D}_t^{\mathsf{T}}$, with $\mathbf{B}_t, \mathbf{D}_t \in \mathbb{R}^{q\times q}$.

Our proof uses the technique of [BM11], which in turns build on an idea first introduced by Bolthausen [Bol12]. A convenient simplification with respect to [BM11] consists in studying a recursion in which the rectangular matrix A is replaced by a symmetric matrix, and the algorithm state is described by a single vector.

In section 2 we put forward formal definitions and state our main result for the case of symmetric matrices. In section 3 we show how the case of rectangular matrices can be reduced to the symmetric one. We also show how our result applies to the case of compressed sensing reconstruction with spatially coupled matrices. Finally, we prove our main result in Section 4.

2 Main result

We will view AMP as operating on the vector space $\mathcal{V}_{q,N} \equiv (\mathbb{R}^q)^N \simeq \mathbb{R}^{N\times q}$. Given a vector $x \in \mathcal{V}_{q,N}$, we shall most often regard it as an N-vector with entries in $\mathbb{R}^q$, namely $x = (\mathbf{x}_1, \dots, \mathbf{x}_N)$, with $\mathbf{x}_i \in \mathbb{R}^q$. Components of $\mathbf{x}_i \in \mathbb{R}^q$ will be indicated as $(\mathbf{x}_i(1), \dots, \mathbf{x}_i(q)) \equiv \mathbf{x}_i$. For $x \in \mathcal{V}_{q,N}$, we define its norm by $\|x\| = \left(\sum_{i=1}^N \|\mathbf{x}_i\|^2\right)^{1/2}$.

Given a matrix $A \in \mathbb{R}^{N\times N}$, we let it act on $\mathcal{V}_{q,N}$ in the natural way, namely for $v', v \in \mathcal{V}_{q,N}$ we let $v' = Av$ be given by $\mathbf{v}'_i = \sum_{j=1}^N A_{ij}\mathbf{v}_j$ for all $i \in [N]$. Here and below $[N] \equiv \{1, \dots, N\}$ is the set of first N integers. In other words we identify A with the Kronecker product $A \otimes \mathrm{I}_{q\times q}$.

Definition 1. *A symmetric* AMP instance *is a triple $(A, \mathcal{F}, x^0)$ where:*

1. *$A = G + G^{\mathsf{T}}$, where $G \in \mathbb{R}^{N\times N}$ has i.i.d. entries $G_{ij} \sim \mathsf{N}(0, (2N)^{-1})$.*
2. *$\mathcal{F} = \{f^k : k \in [N]\}$ is a collection of mappings $f^k : \mathbb{R}^q \times \mathbb{N} \to \mathbb{R}^q$, $(\mathbf{x}, t) \mapsto f^k(\mathbf{x}, t)$ that are locally Lipschitz in their first argument (and hence almost everywhere differentiable);*
3. *$x^0 \in \mathcal{V}_{q,N}$ is an initial condition.*

Given $\mathcal{F} = \{f^k : k \in [N]\}$, we define $f(\,\cdot\,;t) : \mathcal{V}_{q,N} \to \mathcal{V}_{q,N}$ by letting $v' = f(v;t)$ be given by $\mathbf{v}'_i = f^i(\mathbf{v}_i;t)$ for all $i \in [N]$.

Definition 2. *The* approximate message passing orbit *corresponding to the instance $(A, \mathcal{F}, x^0)$ is the sequence of vectors $\{x^t\}_{t\ge 0}$, $x^t \in \mathcal{V}_{q,N}$ defined as follows, for $t \ge 0$,*

$$x^{t+1} = A\, f(x^t;t) - \mathsf{B}_t\, f(x^{t-1};t-1)\,. \tag{5}$$

Here $\mathsf{B}_t : \mathcal{V}_{q,N} \to \mathcal{V}_{q,N}$ is the linear operator defined by letting, for $v' = \mathsf{B}_t v$,

$$\mathbf{v}'_i = \frac{1}{N}\left(\sum_{j\in[N]} \frac{\partial f^j}{\partial \mathbf{x}}(\mathbf{x}^t_j, t)\right)\mathbf{v}_i\,, \tag{6}$$

with $\frac{\partial f^j}{\partial \mathbf{x}}$ denoting the Jacobian matrix of $f^j(\,\cdot\,;t) : \mathbb{R}^q \to \mathbb{R}^q$.

2.1 State evolution

In order to establish the behavior of the sequence $\{x^t\}_{t\ge 0}$ in the high dimensional limit, we need to consider a sequence of AMP instances $\{A(N), \mathcal{F}_N, x^{0,N}\}_{N\ge 0}$ indexed by the dimension N.

Definition 3. *We say that the sequence of AMP instances $\{(A(N), \mathcal{F}_N, x^{0,N})\}_{N\ge 0}$ is* converging *if there exists: (i) An integer q; (ii) A function $g : \mathbb{R}^q \times \mathbb{R}^q \times [q] \times \mathbb{N} \to \mathbb{R}^q$ with $g(\mathbf{x}, \mathbf{y}, a, t) = (g_1(\mathbf{x}, \mathbf{y}, a, t), \cdots, g_q(\mathbf{x}, \mathbf{y}, a, t))$, such that, for each $r \in [q]$, $a \in [q]$, $t \in \mathbb{N}$, $g_r(\cdots, a, t)$ is Lipschitz*

continuous; (iii) q probability measures $P_1, \dots, P_q$ on $\mathbb{R}^q$; (iv) For each N, a finite partition $C_1^N \cup C_2^N \cup \cdots \cup C_q^N = [N]$; (v) q positive definite matrices $\widehat{\Sigma}_1^1, \dots, \widehat{\Sigma}_q^1 \in \mathbb{R}^{q\times q}$, such that the following happens;

1. *For each $a \in [q]$, we have $\lim_{N\to\infty} |C_a^N|/N = c_a \in (0,1)$.*
2. *For each $N \geq 0$, each $a \in [q]$ and each $i \in C_a^N$, we have $f^i(\mathbf{x}, t) = g(\mathbf{x}, \mathbf{y}_i, a, t)$. Further, the empirical distribution of $\{\mathbf{y}_i\}_{i\in C_a^N}$, denoted by $\hat{P}_a$, converges weakly to P_a.*
3. *For each $a \in [q]$, in probability,*

$$\lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{i\in C_a^N} g\Big(\mathbf{x}_i^0, \mathbf{y}_i, a, 0\Big) g\Big(\mathbf{x}_i^0, \mathbf{y}_i, a, 0\Big)^{\mathsf{T}} = \widehat{\Sigma}_a^0 \,. \tag{7}$$

Remark 1. *An apparent generalization of the above definition would require the partition to be $C_1^N \cup C_2^N \cup \cdots \cup C_{q'}^N = [N]$, while $x^t \in \mathcal{V}_{q,N}$, with $q \neq q'$. It is easy to see that there is no loss of generality in assuming $q = q'$ as we do in our definition. Indeed the case $q' < q$ can be reduced to our setting by refining the partition arbitrarily, and $q' > q$ by adding dummy coordinates to to the variables $\mathbf{x}_i$.*

Remark 2. *The function $f^i(\,\cdot\,,\cdot\,)$ depends implicitly on $\mathbf{y}_i$. However, the $\mathbf{y}_i$'s do not change across iterations and so we do not show this dependence explicitly in our notation.*

Our next result establishes that the low-dimensional marginals of $\{x^t\}$ are asymptotically Gaussian. *State evolution* characterizes the covariance of these marginals. For each $t \geq 1$, state evolution defines a positive semidefinite matrix $\Sigma^t \in \mathbb{R}^{q\times q}$. This is obtained by letting, for each $t \geq 1$

$$\Sigma^t = \sum_{b=1}^{q} c_b \, \widehat{\Sigma}_b^{t-1} \,, \tag{8}$$

$$\widehat{\Sigma}_a^t = \mathbb{E}\left\{ g(Z_a^t, Y_a, a, t) g(Z_a^t, Y_a, a, t)^{\mathsf{T}} \right\} , \tag{9}$$

for all $a \in [q]$. Here $Y_a \sim P_a$, $Z_a^t \sim \mathsf{N}\left(0, \Sigma^t\right)$ and Y_a and Z_a^t are independent.

For $k \geq 1$ we say a function $\phi : \mathbb{R}^m \to \mathbb{R}$ is *pseudo-Lipschitz* of order k and denote it by $\phi \in \mathrm{PL}(k)$ if there exists a constant $L > 0$ such that, for all $x, y \in \mathbb{R}^m$:

$$|\phi(x) - \phi(y)| \leq L(1 + \|x\|^{k-1} + \|y\|^{k-1}) \, \|x - y\| \,. \tag{10}$$

Notice that if $\phi \in \mathrm{PL}(k)$, then there exists a constant L' such that for all $x \in \mathbb{R}^m$: $|\phi(x)| \leq L'(1 + \|x\|^k)$.

Theorem 1. *Let $(A(N), \mathcal{F}_N, x^0)_{N\geq 0}$ be a converging sequence of AMP instances, and denote by $\{x^t\}_{t\geq 0}$ the corresponding AMP sequence. Suppose further that $\mathbb{E}_{P_a}(\|Y_a\|^{2k-2})$ is bounded, and $\mathbb{E}_{\hat{P}_a}(\|Y_a\|^{2k-2}) \to \mathbb{E}_{P_a}(\|Y_a\|^{2k-2})$ as $N \to \infty$, for some $k \geq 2$. Then for all $t \geq 1$, each $a \in [q]$, and any pseudo-Lipschitz function $\psi : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$ of order k, we have, almost surely,*

$$\lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{j\in C_a^N} \psi(\mathbf{x}_j^t, \mathbf{y}_j) = \mathbb{E}\{\psi(Z_a^t, Y_a)\} \,, \tag{11}$$

where $Z_a^t \sim \mathsf{N}(0, \Sigma^t)$ is independent of $Y_a \sim P_a$.

3 AMP for rectangular and spatially-coupled matrices

In this section we develop two applications of our main theorem:

1. We show that AMP iterations with A a rectangular matrix, see e.g. Eqs. (1), (2), can be recast in the form of an iteration with a symmetric matrix A and are therefore covered by Theorem 1. This construction is provided in Section 3.5 (below Proposition 5).

2. We apply the general Theorem 1 to AMP reconstruction in compressed sensing with spatially coupled matrices. In [DJM11b], it was proved that, conditionally to a state evolution lemma, this approach achieves the information-theoretic limits of compressed sensing set forth in [WV10]. Here we show that our main result Theorem 1 implies the state evolution lemma (Lemma 4.1 in [DJM11b]).

3.1 General matrix ensemble

We begin by describing a more general matrix ensemble that encompasses spatially coupled matrices, and will be denoted by $\mathcal{M}(W, m_0, n_0)$. The ensemble depends on two integers $m_0, n_0 \in \mathbb{N}$, and on a matrix with non-negative entries $W \in \mathbb{R}_+^{\mathsf{R}\times\mathsf{C}}$, whose rows and columns are indexed by the finite sets R, C (respectively 'rows' and 'columns'). The matrix is *roughly row-stochastic*, i.e.

$$\frac{1}{2} \le \sum_{c\in\mathsf{C}} W_{r,c} \le 2\,, \qquad \text{for all } r \in \mathsf{R}\,. \tag{12}$$

We will let $|\mathsf{R}| \equiv L_r$ and $|\mathsf{C}| \equiv L_c$ denote the matrix dimensions. The ensemble parameters are related to the sensing matrix dimensions by $n = n_0 L_c$ and $m = m_0 L_r$.

In order to describe a random matrix $A \sim \mathcal{M}(W, m_0, n_0)$ from this ensemble, partition the column and row indices of A in –respectively– L_c and L_r groups of equal size. Explicitly

$$\begin{aligned} [n] &= \cup_{s\in\mathsf{C}} C_s\,, \quad |C_s| = n_0\,,\\ [m] &= \cup_{r\in\mathsf{R}} R_r\,, \quad |R_r| = m_0\,. \end{aligned}$$

Further, if $i \in R_r$ or $j \in C_s$ we will write, respectively, $r = \mathsf{g}(i)$ or $s = \mathsf{g}(j)$. In other words $\mathsf{g}(\,\cdot\,)$ is the operator determining the group index of a given row or column.

With this notation we have the following concise definition of the ensemble.

Definition 4. *A random sensing matrix A is distributed according to the ensemble $\mathcal{M}(W, m_0, n_0)$ (and we write $A \sim \mathcal{M}(W, m_0, n_0)$) if the entries $\{A_{ij},\ \ i \in [m], j \in [n]\}$ are independent Gaussian random variables with*

$$A_{ij} \sim \mathsf{N}\Big(0, \frac{1}{m_0}\, W_{\mathsf{g}(i),\mathsf{g}(j)}\Big)\,. \tag{13}$$

See Fig. 1 for a schematic of matrix A. Note that the ensemble $\mathcal{M}(W, m_0, n_0)$ includes, as special case, rectangular non-symmetric matrices with i.i.d. entries.

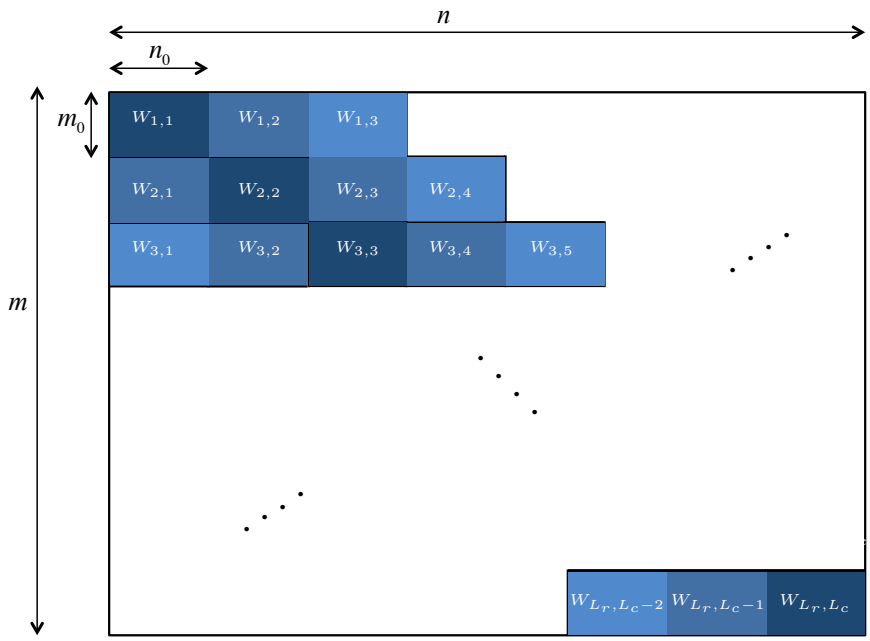

Figure 1: Construction of the spatially coupled measurement matrix A for compressive sensing as described in Section 3.1. The matrix is divided into blocks with size m_0 by n_0. (Number of blocks in each row and each column are respectively L_c and L_r, hence $m = m_0 L_r$, $n = n_0 L_c$). The matrix elements A_{ij} are chosen as $\mathsf{N}(0, \frac{1}{m_0} W_{\mathsf{g}(i),\mathsf{g}(j)})$. In this figure, $W_{i,j}$ depends only on $|i-j|$ and thus blocks on each diagonal have the same variance.

3.2 AMP for compressed sensing reconstruction

AMP algorithms were applied in [DJM11b] to compressed sensing reconstruction with spatially coupled sensing matrices [KMS+12b]. Here we follow the scheme and notations of [DJM11b]. In particular, we assume that the unknown vector x to be reconstructed has entries whose empirical distribution converges weakly to a probability measure p_X over $\mathbb{R}$. The AMP algorithm takes the following form (initialized with $x_i^1 = \mathbb{E}_{p_X}(X)$ for all $i \in [n]$):

$$x^{t+1} = \eta_t(x^t + (Q^t \odot A)^{\mathsf{T}} r^t)\,, \tag{14}$$

$$r^t = y - Ax^t + \mathsf{b}^t \odot r^{t-1}\,. \tag{15}$$

Here, for each t, $\eta_t : \mathbb{R}^n \to \mathbb{R}^n$ is a differentiable non-linear function that depends on the input distribution p_X. Further, for $v \in \mathbb{R}^n$, we have $\eta_t(v) = (\eta_{t,1}(v_1), \dots, \eta_{t,n}(v_n))$ for some functions $\eta_{i,t} : \mathbb{R} \to \mathbb{R}$. The symbol $\odot$ indicates Hadamard (entrywise) product. The specific choices for $\eta_t, Q^t, \mathsf{b}^t$ are given in Section 3.4 below.

3.3 State evolution

Given $W \in \mathbb{R}_+^{\mathsf{R}\times\mathsf{C}}$ roughly row-stochastic, and undersampling rate $\delta \in (0,1)$, the corresponding state evolution is defined as follows. Start with initial condition

$$\psi_i(0) = \infty \text{ for all } i \in \mathsf{C}\,. \tag{16}$$

For all $t \geq 0$, $a \in \mathsf{R}$, and $i \in \mathsf{C}$, let

$$\begin{aligned} \phi_a(t) = &\qquad \sigma^2 + \frac{1}{\delta}\sum_{i\in\mathsf{C}} W_{a,i}\,\psi_i(t)\,, \\ \psi_i(t+1) = &\quad \mathsf{mmse}\Big(\sum_{b\in\mathsf{R}} W_{b,i}\phi_b(t)^{-1}\Big)\,. \end{aligned} \tag{17}$$

Here and below, $\mathsf{mmse}(s)$ denotes the minimum mean square error in estimating $X \sim p_X$ from a noisy observation in Gaussian noise, at signal-to-noise ratio s. Formally,

$$\mathsf{mmse}(s) = \mathbb{E}\{[X - \mathbb{E}[X|Y]]^2\}, \qquad Y = \sqrt{s}X + Z\,.$$

3.4 Construction of $\eta_t, \mathsf{b}^t, Q^t$

In the constructions for the matrix Q^t, the nonlinearities η_t, and the vector b^t, we use the fact that the state evolution sequence can be precomputed.

Define Q^t by

$$Q^t_{ij} \equiv \frac{\phi_{\mathsf{g}(i)}(t)^{-1}}{\sum_{k=1}^{L_r} W_{k,\mathsf{g}(j)}\phi_k(t)^{-1}}\,. \tag{18}$$

The nonlinearity η_t is chosen as follows:

$$\eta_t(v) = \big(\eta_{t,1}(v_1), \eta_{t,2}(v_2),\ \dots\ , \eta_{t,N}(v_N)\big)\,, \tag{19}$$

where $\eta_{t,i}$ is the conditional expectation estimator for $X \sim p_X$ in gaussian noise:

$$\eta_{t,i}(v_i) = \mathbb{E}\big\{X \,\big|\, X + s_{\mathsf{g}(i)}(t)^{-1/2}Z = v_i\,\big\}\,, \quad s_r(t) \equiv \sum_{u\in\mathsf{R}} W_{u,r}\phi_u(t)^{-1}\,. \tag{20}$$

Notice that the function $\eta_{t,i}(\,\cdot\,)$ depends on i only through the group index $\mathsf{g}(i)$, and in fact parametrically through $s_{\mathsf{g}(i)}(t)$. We define $\tilde{\eta}_{t,i} = \eta_{t,u}$ for $i \in C_u$.

Finally, in order to define the vector b^t_i, let us introduce the quantity (with $\eta'_{t,i}$ denoting the derivative of $v_i \mapsto \eta_{t,i}(v_i)$)

$$\langle \eta'_t\rangle_u = \frac{1}{n_0}\sum_{i\in C_u} \eta'_{t,i}\big(x^t_i + ((Q^t \odot A)^{\mathsf{T}} r^t)_i\big)\,. \tag{21}$$

The vector b^t is then defined by

$$\mathsf{b}^t_i \equiv \frac{1}{\delta}\sum_{u\in\mathsf{C}} W_{\mathsf{g}(i),u}\widetilde{Q}^{t-1}_{\mathsf{g}(i),u}\,\langle \eta'_{t-1}\rangle_u\,, \tag{22}$$

where we defined $Q^t_{i,j} = \widetilde{Q}^t_{r,u}$ for $i \in R_r$, $j \in C_u$.

The following Lemma (Lemma 4.1 in [DJM11b]) claims that the state evolution (17) allows an exact asymptotic analysis of AMP algorithm (14)- (15) in the limit of a large number of dimensions.

Lemma 1. *Let $W \in \mathbb{R}_+^{\mathsf{R}\times\mathsf{C}}$ be a roughly row-stochastic matrix and $\phi(t)$, Q^t, b^t be defined as in Section 3.4. Let $m_0 = m_0(n_0)$ be such that $m_0/n_0 \to \delta$, as $n_0 \to \infty$, and let $A(n) \sim \mathcal{M}(W, m_0, n_0)$. Further suppose that the empirical distribution of the entries of $x(n)$ converges weakly to a probability measure p_X on $\mathbb{R}$ with bounded second moment and the empirical second moment of $x(n)$ also converges to $\mathbb{E}_{p_X}(X^2)$. Similarly, suppose that the empirical distribution of the entries of $w(n)$ converges weakly to a probability measure p_W on $\mathbb{R}$ with bounded second moment and the empirical second moment of $w(n)$ also converges to $\mathbb{E}_{p_W}(W^2) \equiv \sigma^2$. Then, for all $t \geq 1$, almost surely we have*

$$\limsup_{n_0\to\infty} \frac{1}{n_0} \|x^t_{C_a}(A(n); y(n)) - x_{C_a}\|_2^2 = \mathsf{mmse}\Big(\sum_{i\in\mathsf{R}} W_{i,a}\phi_i(t-1)^{-1}\Big)\,, \tag{23}$$

for all $a \in \mathsf{C}$, where $x^t_{C_a}, x_{C_a} \in \mathbb{R}^{n_0}$ respectively denote the restrictions of x^t, x to indices in C_a.

3.5 Proof of Lemma 1

We show that Lemma 1 follows from Theorem 1. Consider the following change of variables:

$$\tilde{x}^{t+1} = x - (Q^t \odot A)^{\mathsf{T}} r^t - x^t, \tag{24}$$
$$\tilde{r}^t = w - r^t. \tag{25}$$

Rewriting Eqs (14) and (15) in terms of $\tilde{x}$ and $\tilde{r}$, we obtain

$$\tilde{x}^{t+1} = (Q^t \odot A)^{\mathsf{T}}(\tilde{r}^t - w) - \{\eta_{t-1}(x - \tilde{x}^t) - x\}, \tag{26}$$
$$\tilde{r}^t = A\{\eta_{t-1}(x - \tilde{x}^t) - x\} + \mathsf{b}^t \odot (\tilde{r}^{t-1} - w). \tag{27}$$

Let $q = L_r + L_c$ and define functions $e(\cdot,\cdot,\cdot;t), h(\cdot,\cdot,\cdot;t) : \mathbb{R}^q \times \mathbb{R}^q \times [q] \to \mathbb{R}^q$ as follows:

$$h(\mathbf{u},\mathbf{w},a;t) = \sqrt{L_r}\,(\mathbf{u}(a) - \mathbf{w}(a))\,[\sqrt{W_{a,1}}\widetilde{Q}^t_{a,1}, \ldots, \sqrt{W_{a,L_c}}\widetilde{Q}^t_{a,L_c}, *, \ldots, *] \qquad \text{for } a \in [L_r]\,,$$
$$e(\mathbf{v},\mathbf{y},a;t) = \sqrt{L_r}\,\{\tilde{\eta}_{t-1,a}(\mathbf{y}(a) - \mathbf{v}(a)) - \mathbf{y}(a)\}\,[\sqrt{W_{1,a}}, \ldots, \sqrt{W_{L_r,a}}, *, \ldots, *] \quad \text{for } a \in [L_c]\,.$$

In our definition, we do not care about the values of entries represented by $*$, since they are irrelevant for our purposes. Values of $h(\mathbf{u},\mathbf{w},a;t)$ for $a \in \{L_r+1, \ldots, L_r+L_c\}$ and $e(\mathbf{v},\mathbf{y},a;t)$ for $a \in \{L_c+1, \ldots, L_r+L_c\}$ are also irrelevant for our purposes and can be defined arbitrarily. Note that $h, e \in \mathrm{PL}(2)$. We also define function $\widehat{e}(\cdot,\cdot;t) : \mathcal{V}_{q,n} \times \mathcal{V}_{q,n} \to \mathcal{V}_{q,n}$ by letting $v' = \widehat{e}(v,y;t)$ be given by $\mathbf{v}'_j = e(\mathbf{v}_j, \mathbf{y}_j, \mathbf{g}(j); t)$ for all $j \in [n]$. Similarly, $\widehat{h}(\cdot,\cdot;t) : \mathcal{V}_{q,m} \times \mathcal{V}_{q,m} \to \mathcal{V}_{q,m}$ is defined by letting $u' = \widehat{h}(u,w;t)$ be given by $\mathbf{u}'_i = h(\mathbf{u}_i, \mathbf{w}_i, \mathbf{g}(i); t)$ for all $i \in [m]$.

Let $\tilde{A} \in \mathbb{R}^{m\times n}$ be a normalized version of A obtained as in the following:

$$\tilde{A}_{ij} = \sqrt{\frac{1}{L_r\, W_{\mathbf{g}(i),\mathbf{g}(j)}}}\, A_{ij}.$$

Therefore, $\tilde{A}$ has i.i.d. entries $\mathsf{N}(0, 1/m)$.

Proposition 5. *Consider the following approximate message passing orbit with vectors $\{v^t, u^t\}_{t\geq 0}$, $v^t \in \mathcal{V}_{q,n}$, $u^t \in \mathcal{V}_{q,m}$:*

$$u^t = \tilde{A}\,\widehat{e}(v^t, y; t) - \mathsf{B}_t\,\widehat{h}(u^{t-1}, w; t-1)\,, \tag{28}$$
$$v^{t+1} = \tilde{A}^{\mathsf{T}}\,\widehat{h}(u^t, w; t) - \mathsf{D}_t\,\widehat{e}(v^t, y; t)\,, \tag{29}$$

for given $y \in \mathcal{V}_{q,n}$ and $w \in \mathcal{V}_{q,m}$. Here $\mathsf{B}_t : \mathcal{V}_{q,m} \to \mathcal{V}_{q,m}$ is the linear operator defined by letting, for $z' = \mathsf{B}_t z$, and any $i \in [m]$,

$$\mathbf{z}'_i = \frac{1}{m} \left(\sum_{k \in [n]} \frac{\partial e}{\partial \mathbf{v}} (\mathbf{v}_k^t, \mathbf{y}_k, \mathrm{g}(k); t) \right) \mathbf{z}_i \,. \tag{30}$$

Analogously $\mathsf{D}_t : \mathcal{V}_{q,n} \to \mathcal{V}_{q,n}$ is the linear operator defined by letting, for $z' = \mathsf{D}_t z$, and any $j \in [n]$,

$$\mathbf{z}'_j = \frac{1}{m} \left(\sum_{l \in [m]} \frac{\partial h}{\partial \mathbf{u}} (\mathbf{u}_l^t, \mathbf{w}_l, \mathrm{g}(l); t) \right) \mathbf{z}_j \,. \tag{31}$$

Assume that $y = (\mathbf{y}_1, \dots, \mathbf{y}_n)$, $w = (\mathbf{w}_1, \dots, \mathbf{w}_m)$, and $v^1 = (\mathbf{v}_1^1, \dots, \mathbf{v}_n^1)$ are given by

$$\begin{aligned}
\mathbf{y}_k &= (*, \cdots, *, \underbrace{x_k}_{\text{position g(k)}}, *, \cdots, *) \in \mathbb{R}^q, \qquad \forall k \in [n] \\
\mathbf{w}_k &= (*, \cdots, *, \underbrace{w_k}_{\text{position g(k)}}, *, \cdots, *) \in \mathbb{R}^q, \qquad \forall k \in [m] \\
\mathbf{v}_k^1 &= (*, \cdots, *, \underbrace{\tilde{x}_k^1}_{\text{position g(k)}}, *, \cdots, *) \in \mathbb{R}^q, \qquad \forall k \in [n].
\end{aligned}$$

Then, we have $\mathbf{u}_i^t(\mathrm{g}(i)) = \tilde{r}_i^t$ and $\mathbf{v}_j^{t+1}(\mathrm{g}(j)) = \tilde{x}_j^{t+1}$, for all $i \in [m], j \in [n]$, and $t \geq 0$.

We refer to Section 3.5.1 for the proof of Proposition 5.

We proceed by constructing a suitable converging sequence of symmetric AMP instances, recognizing that a subset of the resulting orbit corresponds to the orbit $\{v^t, u^t\}$ of interest. The converging symmetric AMP instances $(A_s(N), g, x_s^0)$ are defined as:

- The instances has dimensions $N = m + n$ and $q = L_r + L_c$.
- Let $B_1 = C_1 + C_1^{\mathsf{T}}$ and $B_2 = C_2 + C_2^{\mathsf{T}}$, where $C_1 \in \mathbb{R}^{m \times m}$ and $C_2 \in \mathbb{R}^{n \times n}$ have i.i.d. entries distributed as $\mathsf{N}(0, (2m)^{-1})$. The symmetric matrix A_s is given by
$$A_s = \sqrt{\frac{\delta}{\delta + 1}} \begin{pmatrix} B_1 & \tilde{A} \\ \tilde{A}^{\mathsf{T}} & B_2 \end{pmatrix}.$$
- Let $\mathbf{y}_{s,i} = \mathbf{w}_i \in \mathbb{R}^q$ for $i \leq m$ and $\mathbf{y}_{s,i} = \mathbf{y}_{i-m} \in \mathbb{R}^q$ for $i > m$.
- The initial condition is given by $x_s^0 = (\mathbf{x}_{s,1}^0, \cdots, \mathbf{x}_{s,N}^0) \in \mathcal{V}_{q,N}$, where $\mathbf{x}_{s,i}^0 = 0$ for $i \leq m$ and $\mathbf{x}_{s,i}^0 = \mathbf{v}_{i-m}^1$ for $m < i \leq m + n$.
- Finally, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^q$, $t \geq 0$, we let

$$\begin{aligned}
g(\mathbf{x}, \mathbf{y}, a, 2t) &= 0 && \text{for } a \in \{1, \cdots, L_r\}, && (32) \\
g(\mathbf{x}, \mathbf{y}, a, 2t) &= \sqrt{\tfrac{\delta+1}{\delta}}\, e(\mathbf{x}, \mathbf{y}, a - L_r; t) && \text{for } a \in \{L_r + 1, \cdots, L_r + L_c\}, && (33) \\
g(\mathbf{x}, \mathbf{y}, a, 2t+1) &= \sqrt{\tfrac{\delta+1}{\delta}}\, h(\mathbf{x}, \mathbf{y}, a; t+1) && \text{for } a \in \{1, \cdots, L_r\}, && (34) \\
g(\mathbf{x}, \mathbf{y}, a, 2t+1) &= 0 && \text{for } a \in \{L_r + 1, \cdots, L_r + L_c\}. && (35)
\end{aligned}$$

Now, it is easy to see that, for all $t \geq 0$,

$$\mathbf{x}_{s,i}^{2t+1} = \mathbf{u}_i^t, \qquad \text{for } i \leq m, \tag{36}$$

$$\mathbf{x}_{s,i}^{2t} = \mathbf{v}_{i-m}^{t+1}, \qquad \text{for } m+1 \leq i \leq m+n. \tag{37}$$

Now we are ready to prove Lemma 1 by applying Theorem 1.

Fix $a' \in \{L_r+1, \ldots, L_r+L_c\}$ and $t \geq 1$. Let $a = a' - L_r$ and choose function $\psi(\mathbf{x}, \mathbf{y}) = \{\tilde{\eta}_{t,a}(\mathbf{y}(a) - \mathbf{x}(a)) - \mathbf{y}(a)\}^2$. Then,

$$\begin{aligned}
\lim_{n_0\to\infty} \frac{1}{n_0} \sum_{j'\in C_{a'}} \psi(\mathbf{x}_{s,j}^{2t}, \mathbf{y}_{s,j}) &= \lim_{n_0\to\infty} \frac{1}{n_0} \sum_{j'\in C_{a'}} [\tilde{\eta}_{t,a}(\mathbf{y}_{s,j'}(a) - \mathbf{x}_{s,j'}^{2t}(a)) - \mathbf{y}_{s,j'}(a)]^2 \\
&\overset{(a)}{=} \lim_{n_0\to\infty} \frac{1}{n_0} \sum_{j\in C_a} [\tilde{\eta}_{t,a}(\mathbf{y}_j(a) - \mathbf{v}_j^{t+1}(a)) - \mathbf{y}_j(a)]^2 \\
&\overset{(b)}{=} \lim_{n_0\to\infty} \frac{1}{n_0} \sum_{j\in C_a} [\eta_{t,j}(x_j - \tilde{x}_j^{t+1}) - x_j]^2 \\
&= \lim_{n_0\to\infty} \frac{1}{n_0} \sum_{j\in C_a} (x_j^{t+1} - x_j)^2 = \lim_{n_0\to\infty} \frac{1}{n_0} \|x_{C_a}^{t+1} - x_{C_a}\|^2.
\end{aligned} \tag{38}$$

Here (a) follows from Eq. (37) and the definition of $\mathbf{y}_{s,j}$ (note that $j' = j - m$); (b) follows from the fact $a = \mathsf{g}(j)$ and Proposition 5.

Applying Theorem 1, we have almost surely

$$\lim_{n_0\to\infty} \frac{1}{n_0} \sum_{j'\in C_{a'}} \psi(\mathbf{x}_{s,j}^{2t}, \mathbf{y}_{s,j}) = \mathbb{E}[\eta_{t,a}(X+Z) - X]^2, \tag{39}$$

with $X \sim p_X$ and $Z \sim \mathsf{N}(0, \Sigma_{aa}^{2t})$. Therefore, to complete the proof we need to show that

$$(\Sigma_{aa}^{2t})^{-1} = \sum_{i\in\mathsf{R}} W_{i,a}\phi_i(t)^{-1}. \tag{40}$$

Note that Eq. (8) reduces to:

$$\Sigma^t = \frac{m_0}{m+n} \sum_{b'=1}^{L_r} \widehat{\Sigma}_{b'}^{t-1} + \frac{n_0}{m+n} \sum_{b'=L_r+1}^{L_r+L_c} \widehat{\Sigma}_{b'}^{t-1}. \tag{41}$$

By definition of function g (see Eq.s (32)- (35)), it is easy to see that Eq. (9) reduces to:

$$(\widehat{\Sigma}_{a'}^{2t})_{ij} = \begin{cases} 0, & \text{for } a' \in [L_r], \\ \frac{\delta+1}{\delta} L_r \sqrt{W_{i,a}W_{j,a}}\, \mathbb{E}\{\eta_{t-1,a}(X - Z_a^t) - X\}^2, & \text{for } a' \in \{L_r+1, \cdots, L_r+L_c\}, i,j \in [L_r], \\ *, & \text{otherwise.} \end{cases} \tag{42}$$

Here $a = a' - L_r$, $X \sim p_X$ and $Z_a^t \sim \mathsf{N}(0, \Sigma_{aa}^{2t})$. Also,

$$(\widehat{\Sigma}_{a'}^{2t-1})_{ij} = \begin{cases} \frac{\delta+1}{\delta} L_r \sqrt{W_{a',i}W_{a',j}}\, \widetilde{Q}_{a',i}^t \widetilde{Q}_{a',j}^t \{\sigma^2 + \Sigma_{a'a'}^{2t-1}\}, & \text{for } a' \in [L_r], i,j \in [L_c], \\ 0, & \text{for } a' \in \{L_r+1, \cdots, L_r+L_c\}, \\ *, & \text{otherwise.} \end{cases} \tag{43}$$

Consequently, we obtain

$$\begin{aligned}
\Sigma_{aa}^{2t} &= \frac{m_0}{m+n}\sum_{b=1}^{L_r}(\widehat{\Sigma}_b^{2t-1})_{aa} \\
&= \frac{m_0 L_r}{m+n}\cdot\frac{\delta+1}{\delta}\sum_{b=1}^{L_r} W_{b,a}(\widetilde{Q}_{b,a}^t)^2\{\sigma^2+\Sigma_{bb}^{2t-1}\} \\
&= \frac{m_0 L_r}{m+n}\cdot\frac{\delta+1}{\delta}\sum_{b=1}^{L_r} W_{b,a}(\widetilde{Q}_{b,a}^t)^2\{\sigma^2+\frac{n_0}{m+n}\sum_{c'=L_r+1}^{L_r+L_c}(\widehat{\Sigma}_{c'}^{2t-2})_{bb}\} \\
&= \frac{m_0 L_r}{m+n}\cdot\frac{\delta+1}{\delta}\sum_{b=1}^{L_r} W_{b,a}(\widetilde{Q}_{b,a}^t)^2\left\{\sigma^2+\frac{n_0 L_r}{m+n}\cdot\frac{\delta+1}{\delta}\sum_{c=1}^{L_c} W_{b,c}\mathsf{mmse}((\Sigma_{cc}^{2t-2})^{-1})\right\} \\
&= \sum_{b=1}^{L_r} W_{b,a}(\widetilde{Q}_{b,a}^t)^2\left\{\sigma^2+\frac{1}{\delta}\sum_{c=1}^{L_c} W_{b,c}\ \mathsf{mmse}((\Sigma_{cc}^{2t-2})^{-1})\right\}.
\end{aligned}$$

We prove relation (40) using induction on t. The induction basis ($t=0$) is trivial. Suppose that the claim holds for $t-1$. Then,

$$\begin{aligned}
\Sigma_{aa}^{2t} &= \sum_{b=1}^{L_r} W_{b,a}(\widetilde{Q}_{b,a}^t)^2\left\{\sigma^2+\frac{1}{\delta}\sum_{c=1}^{L_c} W_{b,c}\ \mathsf{mmse}(\sum_{i\in\mathsf{R}} W_{i,c}\phi_i(t-1)^{-1})\right\} \\
&= \sum_{b=1}^{L_r} W_{b,a}(\widetilde{Q}_{b,a}^t)^2\phi_b(t) \\
&= \sum_{b=1}^{L_r} W_{b,a}\frac{\phi_b(t)^{-2}}{\left(\sum_{k=1}^{L_r} W_{k,a}\phi_k(t)^{-1}\right)^2}\phi_b(t) \\
&= \left(\sum_{b=1}^{L_r} W_{b,a}\phi_b(t)^{-1}\right)^{-1}.
\end{aligned}$$

This proves the induction claim for t. Combining (38),(39) and (40), Lemma 1 follows.

3.5.1 Proof of Proposition 5

We prove the result by induction on t. For $t=0$, the claim follows from our definition. Suppose that the claim holds for $t-1$, we prove that for t.

Writing Eq. (28) for coordinate i, we have

$$\mathbf{u}_i^t = \sum_{k\in[n]}\tilde{A}_{ik}e(\mathbf{v}_k^t,\mathbf{y}_k,\mathbf{g}(k);t) - \frac{1}{m}\left(\sum_{k\in[n]}\frac{\partial e}{\partial\mathbf{v}}(\mathbf{v}_k^t,\mathbf{y}_k,\mathbf{g}(k);t)\right)h(\mathbf{u}_i^{t-1},\mathbf{w}_i,\mathbf{g}(i);t-1) \tag{44}$$

Restricting to coordinate $\mathsf{g}(i)$, we get

$$\begin{aligned}\mathbf{u}_i^t(\mathsf{g}(i)) = &\sum_{k\in[n]} \tilde{A}_{ik}[e(\mathbf{v}_k^t, \mathbf{y}_k, \mathsf{g}(k); t)]_{\mathsf{g}(i)} \\ &- \frac{1}{m}\sum_{k\in[n]} [\frac{\partial e}{\partial v}(\mathbf{v}_k^t(\mathsf{g}(k)), \mathbf{y}_k(\mathsf{g}(k)), \mathsf{g}(k); t)]_{\mathsf{g}(i)}\ [h(\mathbf{u}_i^{t-1}, \mathbf{w}_i, \mathsf{g}(i); t-1)]_{\mathsf{g}(k)}.\end{aligned} \tag{45}$$

Here, we have used the fact that $e(\mathbf{v}_k^t, \mathbf{y}_k, \mathsf{g}(k), t)$ does not depend on $\mathbf{v}_{k,l}^t$ for $l \neq \mathsf{g}(k)$.

Substituting for e and h, we have

$$\begin{aligned}\sum_{k\in[n]} \tilde{A}_{ik}[e(\mathbf{v}_k^t, \mathbf{y}_k, \mathsf{g}(k); t)]_{\mathsf{g}(i)} &= \sum_{k\in[n]} \tilde{A}_{ik}\sqrt{L_r\, W_{\mathsf{g}(i),\mathsf{g}(k)}}\{\tilde{\eta}_{t-1,\mathsf{g}(k)}(\mathbf{y}_k(\mathsf{g}(k)) - \mathbf{v}_k^t(\mathsf{g}(k))) - \mathbf{y}_k(\mathsf{g}(k))\} \\ &= \sum_{k\in[n]} A_{ik}\{\eta_{t-1,k}(x_k - \tilde{x}_k^t) - x_k\},\end{aligned} \tag{46}$$

where we used the induction hypothesis in the last step. Furthermore,

$$\begin{aligned}&\frac{1}{m}\sum_{k\in[n]} [\frac{\partial e}{\partial v}(\mathbf{v}_k^t(\mathsf{g}(k)), \mathbf{y}_k(\mathsf{g}(k)), \mathsf{g}(k); t)]_{\mathsf{g}(i)}\ [h(\mathbf{u}_i^{t-1}, \mathbf{w}_i, \mathsf{g}(i); t-1)]_{\mathsf{g}(k)} \\ &= -\frac{1}{m}\sum_{k\in[n]} \tilde{\eta}'_{t-1,\mathsf{g}(k)}\left(\mathbf{y}_k(\mathsf{g}(k)) - \mathbf{v}_k^t(\mathsf{g}(k))\right)\sqrt{L_r\, W_{\mathsf{g}(i),\mathsf{g}(k)}}\left(\mathbf{u}_i^{t-1}(\mathsf{g}(i)) - \mathbf{w}_i(\mathsf{g}(i))\right)\sqrt{L_r\, W_{\mathsf{g}(i),\mathsf{g}(k)}}\widetilde{Q}^{t-1}_{\mathsf{g}(i),\mathsf{g}(k)} \\ &= -\frac{1}{m}\sum_{k\in[n]} L_r W_{\mathsf{g}(i),\mathsf{g}(k)}\widetilde{Q}^{t-1}_{\mathsf{g}(i),\mathsf{g}(k)}\,\eta'_{t-1,k}(x_k - \tilde{x}_k^t)\,(\tilde{r}_i^{t-1} - w_i) \\ &= -\mathsf{b}_i^t(\tilde{r}_i^{t-1} - w_i),\end{aligned} \tag{47}$$

where we used the induction hypothesis in the second equality. The last equality follows from the definition of b_i^t (see Eq. (22));

Using (46) and (47) in (45), we obtain

$$\mathbf{u}_i^t(\mathsf{g}(i)) = \sum_{k\in[n]} A_{ik}\{\eta_{t-1,k}(x_k - \tilde{x}_k^t) - x_k\} + \mathsf{b}_i^t(\tilde{r}_i^{t-1} - w_i) = \tilde{r}_i^t, \tag{48}$$

where the second equality follows from (27). This proves the induction claim for $\mathbf{u}_i^t(\mathsf{g}(i))$.

Next we prove the claim for $\mathbf{v}_j^{t+1}(\mathsf{g}(j))$. Writing Eq. (29) for coordinate j, we have

$$\mathbf{v}_j^{t+1} = \sum_{l\in[m]} \tilde{A}_{lj} h(\mathbf{u}_l^t, \mathbf{w}_l, \mathsf{g}(l); t) - \frac{1}{m}\left(\sum_{l\in[m]} \frac{\partial h}{\partial \mathbf{u}}(\mathbf{u}_l^t, \mathbf{w}_l, \mathsf{g}(l); t)\right) e(\mathbf{v}_j^t, \mathbf{y}_j, \mathsf{g}(j); t) \tag{49}$$

Restricting to coordinate $\mathsf{g}(j)$, we get

$$\begin{aligned}\mathbf{v}_j^{t+1}(\mathsf{g}(j)) = &\sum_{l\in[m]} \tilde{A}_{lj}[h(\mathbf{u}_l^t, \mathbf{w}_l, \mathsf{g}(l); t)]_{\mathsf{g}(j)} \\ &- \frac{1}{m}\sum_{l\in[m]} [\frac{\partial h}{\partial u}(\mathbf{u}_l^t(\mathsf{g}(l)), \mathbf{w}_l(\mathsf{g}(l)), \mathsf{g}(l); t)]_{\mathsf{g}(j)}[e(\mathbf{v}_j^t, \mathbf{y}_j, \mathsf{g}(j); t)]_{\mathsf{g}(l)}.\end{aligned} \tag{50}$$

Here, we have used the fact that $h(\mathbf{u}_l^t, \mathbf{w}_l, \mathsf{g}(l), t)$ does not depend on $\mathbf{u}_{l,k}^t$ for $k \neq \mathsf{g}(l)$.

Substituting for e and h, we have

$$\begin{aligned}\sum_{l\in[m]} \tilde{A}_{lj}[h(\mathbf{u}_l^t, \mathbf{w}_l, \mathsf{g}(l); t)]_{\mathsf{g}(j)} &= \sum_{l\in[m]} \tilde{A}_{lj}\sqrt{L_r\, W_{\mathsf{g}(l),\mathsf{g}(j)}}\, \widetilde{Q}^t_{\mathsf{g}(l),\mathsf{g}(j)}\, (\mathbf{u}_l^t(\mathsf{g}(l)) - \mathbf{w}_l(\mathsf{g}(l))) \\ &= \sum_{l\in[m]} A_{lj} Q^t_{l,j}\, (\tilde{r}_l^t - w_l), \end{aligned} \tag{51}$$

where in the last step we used the result $\mathbf{u}_l^t(\mathsf{g}(l)) = \tilde{r}_l^t$, proved above. Moreover,

$$\begin{aligned}&\frac{1}{m}\sum_{l\in[m]} [\frac{\partial h}{\partial u}(\mathbf{u}_l^t(\mathsf{g}(l)), \mathbf{w}_l(\mathsf{g}(l)), \mathsf{g}(l); t)]_{\mathsf{g}(j)} [e(\mathbf{v}_j^t, \mathbf{y}_j, \mathsf{g}(j); t)]_{\mathsf{g}(l)} \\ &= \frac{1}{m}\sum_{l\in[m]} \sqrt{L_r W_{\mathsf{g}(l),\mathsf{g}(j)}}\, \widetilde{Q}^t_{\mathsf{g}(l),\mathsf{g}(j)} \{\tilde{\eta}_{t-1,\mathsf{g}(j)}(\mathbf{y}_j(\mathsf{g}(j)) - \mathbf{v}_j^t(\mathsf{g}(j))) - \mathbf{y}_j(\mathsf{g}(j))\} \sqrt{L_r W_{\mathsf{g}(l),\mathsf{g}(j)}} \\ &= \frac{1}{m}\left(\sum_{l\in[m]} L_r W_{\mathsf{g}(l),\mathsf{g}(j)} Q^t_{l,j}\right) \{\eta_{t-1,j}(x_j - \tilde{x}_j^t) - x_j\} \\ &= \eta_{t-1,j}(x_j - \tilde{x}_j^t) - x_j. \end{aligned} \tag{52}$$

Using (51) and (52) in (50), we obtain

$$\mathbf{v}_j^{t+1}(\mathsf{g}(j)) = \sum_{l\in[m]} A_{lj} Q^t_{l,j}\, (\tilde{r}_l^t - w_l) - \{\eta_{t-1,j}(x_j - \tilde{x}_j^t) - x_j\} = \tilde{x}_j^{t+1}, \tag{53}$$

where the second equality follows from (26). This proves the induction claim for $\mathbf{v}_i^{t+1}(\mathsf{g}(i))$.

4 Proof of Theorem 1

4.1 Definitions and notations

Letting $m^t = f(x^t; t)$ for $t \geq 0$, Eq. (5), becomes

$$x^{t+1} = A\, m^t - \mathsf{B}_t\, m^{t-1}\,. \tag{54}$$

This is initialized with $m^{-1} = 0$ and $m^0 = m^{0,N} \in \mathcal{V}_{q,N}$, a sequence of deterministic vectors in $\mathcal{V}_{q,N}$, with $\limsup_{N\to\infty} N^{-1}\sum_{i=1}^N \|\mathbf{m}_i^0\|^{2k-2} < \infty$. Also recall that the vectors $y = (\mathbf{y}_1, \dots, \mathbf{y}_N) \in \mathcal{V}_{q,N}$ are a fixed sequence indexed by N, with converging empirical distributions.

The idea of the proof is to study the stochastic process $\{x^0, x^1, \dots, x^t, \dots\}$ taking values in $\mathcal{V}_{q,N}$ without conditioning on the matrix A. Instead, for each t, we will compute the conditional distribution of x^{t+1} given $x^0, \dots, x^t$, and hence $m^0, \dots, m^t$. More precisely, let $\mathfrak{S}_t$ be the σ-algebra generated by these variables. We will compute the conditional distributions $x^{t+1}|_{\mathfrak{S}_t}$, by characterizing the conditional distribution of the matrix A given this filtration.

Throughout the proof, we identify $\mathcal{V}_{q,N}$ with the set of matrices $\mathbb{R}^{N\times q}$. Adopting this convention, the linear operator B_t can be more conveniently identified with the $q \times q$ matrix

$$\mathbf{B}_t = \frac{1}{N}\left(\sum_{j\in[N]} \frac{\partial f^j}{\partial \mathbf{x}}(\mathbf{x}_j^t, t)\right)\,. \tag{55}$$

We therefore have $\mathsf{B}_t m_{t-1} = m_{t-1}\mathbf{B}_t^{\mathsf{T}}$ and the equations for $x^1, \dots, x^t$ can be written in matrix form as:

$$\underbrace{\left[x^1 | x^2 + m^0\mathbf{B}_1^{\mathsf{T}} | \dots | x^t + m^{t-2}\mathbf{B}_{t-1}^{\mathsf{T}}\right]}_{Y_{t-1}} = A \underbrace{[m^0 | \dots | m^{t-1}]}_{M_{t-1}}. \tag{56}$$

In short $Y_{t-1} = AM_{t-1}$. Here and below we use $[Q|P]$ to denote the matrix obtained by concatenating Q and P horizontally.

We also introduce the notation $m_{\|}^t$ for the projection of m^t onto the column space of M_{t-1}. More precisely, $m_{\|}^t \in \mathbb{R}^{N\times q}$ is the matrix whose columns are the projections of the columns of m^t. This can be written as

$$m_{\|}^t = \sum_{i=0}^{t-1} m^i \alpha_i \,, \tag{57}$$

where $\alpha_i \in \mathbb{R}^{q\times q}$, $0 \le i \le t-1$ contain the coefficients of these projections. Defining by $m_{\perp}^t = m^t - m_{\|}^t$ the perpendicular component , we have $M_{t-1}^{\mathsf{T}} m_{\perp}^t = 0$. We further denote by $\alpha \in \mathbb{R}^{tq\times q}$ the matrix obtained by concatenating α_i's vertically. Using this notation, we have

$$m_{\|}^t = M_{t-1}\alpha \,. \tag{58}$$

For an integer $\ell \ge 1$, let $(\ell) = \{(\ell-1)q+1, \dots, \ell q\}$. For a matrix u and set of indices I, J, we let $u_{I,J}$ denote the submatrix formed by the rows in I and columns in J. We further let u_I denote the submatrix containing just the rows in I. For $v = (\mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathcal{V}_{q,N}$ and a set of indices $I = \{i_1, \dots, i_r\}$, let $v_I = (\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_r})$.

Given $v \in \mathcal{V}_{q,m}$ and $\varphi : \mathbb{R}^q \to \mathbb{R}^q$, we write $\varphi(v) = (\varphi(\mathbf{v}_1), \dots, \varphi(\mathbf{v}_m))$. We also define $\nabla\varphi(v) = [\frac{\partial\varphi}{\partial\mathbf{v}}(\mathbf{v}_1), \dots, \frac{\partial\varphi}{\partial\mathbf{v}}(\mathbf{v}_m)]^{\mathsf{T}}$ with $\frac{\partial\varphi}{\partial\mathbf{v}} \in \mathbb{R}^{q\times q}$ denoting the Jacobian matrix of φ. Note that $\nabla\varphi(v) \in \mathbb{R}^{mq\times q}$.

For $u \in \mathbb{R}^{mq\times q}$, let $\langle u \rangle = (1/m)\sum_{i=1}^m u_{(i)} \in \mathbb{R}^{q\times q}$. Also, for $u, v \in \mathcal{V}_{q,N}$ we define

$$\langle u, v\rangle = \frac{1}{N}\sum_{i=1}^N \mathbf{u}_i \mathbf{v}_i^{\mathsf{T}} \in \mathbb{R}^{q\times q}.$$

Note that $\langle u, v\rangle = (1/N)u^{\mathsf{T}}v$, as we regard $\mathcal{V}_{q,N} \equiv \mathbb{R}^{N\times q}$.

Given two random variables X, Y, and a σ-algebra $\mathfrak{S}$, the notation $X|_{\mathfrak{S}} \stackrel{\mathrm{d}}{=} Y$ means that for any integrable function ϕ and for any random variable Z measurable on $\mathfrak{S}$, $\mathbb{E}\{\phi(X)Z\} = \mathbb{E}\{\phi(Y)Z\}$. In words we will say that X is distributed as (or is equal in distribution to) Y *conditional on* $\mathfrak{S}$. In case $\mathfrak{S}$ is the trivial σ-algebra we simply write $X \stackrel{\mathrm{d}}{=} Y$ (i.e. X and Y are equal in distribution). For random variables X, Y the notation $X \stackrel{\mathrm{a.s.}}{=} Y$ means that X and Y are equal almost surely.

The large system limit will be denoted as $\lim_{N\to\infty}$. In the large system limit, we use the notation $\vec{o}_t(1)$ to represent a matrix in $\mathbb{R}^{tq\times q}$ (with t fixed) such that all of its coordinates converge to 0 almost surely as $N \to \infty$.

The indicator function of property $\mathcal{A}$ is denoted by $\mathbb{I}(\mathcal{A})$ or $\mathbb{I}_{\mathcal{A}}$. The normal distribution with mean μ and variance v^2 is represented as $\mathsf{N}(\mu, v^2)$.

4.2 Main technical Lemma

We will say that a convergent sequence of mappings $(\mathcal{F}_N)_{N\in\mathbb{N}}$ is non-trivial if there exists $\varepsilon_0 > 0$ such that, for each N, $t \geq 0$, $a \in [q]$, $i \in [N]$, $\gamma \in \mathbb{R}^q$ with $\|\gamma\|_2 = 1$, $b \in \mathbb{R}$, we have

$$\int \big(\gamma^{\mathsf{T}} g(\mathbf{x}, \mathbf{y}_i, a, t) - b\big)^2 \mathrm{d}\mathbf{x} \geq \varepsilon_0 \,.$$

This condition is useful to rule out trivial degeneracies.

Lemma 2. *Let $\{(A(N), \mathcal{F}_N, x^{0,N})\}_N$ be a converging sequence of AMP instances as in Theorem 1 with $\mathcal{F}_N$ non-trivial. Then the following hold for all $t \in \mathbb{N}$*

(*a*)

$$x^{t+1}|_{\mathfrak{S}_t} \stackrel{\mathrm{d}}{=} \sum_{i=0}^{t-1} x^{i+1}\alpha_i + \tilde{A}m_\perp^t + \tilde{M}_{t-1}\vec{o}_{t-1}(1)\,, \tag{59}$$

where $\tilde{A}$ is an independent copy of A. The matrix $\tilde{M}_t$ is such that its columns form an orthogonal basis for the column space of M_t and $\tilde{M}_t^{\mathsf{T}}\tilde{M}_t = N\,\mathrm{I}_{tq\times tq}$. Recall that, $\vec{o}_{t-1}(1) \in \mathbb{R}^{(t-1)q\times q}$ is a random vector that converges to 0 almost surely as $N \to \infty$.

(*b*) *For any pseudo-Lipschitz function $\phi : (\mathbb{R}^q)^{t+2} \to \mathbb{R}$ of order k,*

$$\lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{i\in C_a^N} \phi(\mathbf{x}_i^1, \dots, \mathbf{x}_i^{t+1}, \mathbf{y}_i) \stackrel{\text{a.s.}}{=} \mathbb{E}\big[\phi(Z_a^1, \dots, Z_a^{t+1}, Y_a)\big]\,. \tag{60}$$

where $(Z_a^1, \dots, Z_a^{t+1})$ is a Gaussian vector independent of $Y_a \sim P_a$ and, for each i, $Z_a^i \sim \mathsf{N}(0, \Sigma^i)$

(*c*) *For all $1 \leq r, s \leq t, a \in [q]$ the following equations hold and all limits exist, are bounded and have degenerate distribution (i.e. they are constant random variables):*

$$\lim_{N\to\infty} \langle x^{r+1}_{C_a^N}, x^{s+1}_{C_a^N}\rangle \stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \langle m^r, m^s\rangle\,. \tag{61}$$

(*d*) *Consider any set of q Lipschitz continuous functions $\varphi^a : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}^q$. For all $1 \leq r, s \leq t$, the following equations hold and all limits exist, are bounded and have degenerate distribution (i.e. they are constant random matrices):*

$$\lim_{N\to\infty} \langle x^{r+1}_{C_a^N}, \varphi(x^{s+1}_{C_a^N}, y_{C_a^N})\rangle \stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \langle x^{r+1}_{C_a^N}, x^{s+1}_{C_a^N}\rangle \langle \nabla\varphi^a(x^{s+1}_{C_a^N}, y_{C_a^N})\rangle\,. \tag{62}$$

The Jacobians here are computed according to the first component. Define $\varphi : \mathcal{V}_{q,N} \times \mathcal{V}_{q,N} \to \mathcal{V}_{q,N}$ by letting $v' = \varphi(u, v)$ be given by $\mathbf{v}'_i = \varphi^a(\mathbf{u}_i, \mathbf{v}_i)$ for $i \in C_a^N$. Let $\nabla\varphi \in \mathbb{R}^{Nq\times q}$ be the matrix obtained by concatenating the matrices $\nabla\varphi^a \in \mathbb{R}^{|C_a^N|q\times q}$, for $a \in [q]$. Then, Eq. (62) implies that for all $1 \leq r, s \leq t$, the following equations hold:

$$\lim_{N\to\infty} \langle x^{r+1}, \varphi(x^{s+1}, y)\rangle \stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \langle x^{r+1}, x^{s+1}\rangle \langle \nabla\varphi(x^{s+1}, y)\rangle\,. \tag{63}$$

(*e*) *For $\ell = k-1$ and $a \in [q]$, the following holds almost surely*

$$\lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{i\in C_a^N} \|\mathbf{x}_i^{t+1}\|^{2\ell} < \infty\,. \tag{64}$$

(*f*) *For all $0 \le r \le t$ the following limit exists and there are positive constants ρ_r (independent of N) such that almost surely*

$$\lim_{N\to\infty} \langle m_\perp^r, m_\perp^r\rangle - \rho_r\, \mathrm{I}_{q\times q} \succeq 0\,. \tag{65}$$

4.2.1 Proof of Theorem 1

First assume that the sequence of functions $\mathcal{F}_N$ is non-trivial. Theorem 1 follows readily from Lemma 2. More specifically, Theorem 1 is obtained by applying Lemma 2(*b*) to functions $\phi(\mathbf{x}_i^1, \dots, \mathbf{x}_i^t) = \psi(\mathbf{x}_i^t, \mathbf{y}_i)$.

Consider then the case in which $\mathcal{F}_N$ is not non-trivial. In this case we perturb the functions $g(\mathbf{x}, \mathbf{y}, a, t)$ as follows. Let $\varphi(\mathbf{x}) : \mathbb{R}^q \to \mathbb{R}^q$ be a bounded smooth function. Define

$$g^\epsilon(\mathbf{x}, \mathbf{y}, a, t) = g(\mathbf{x}, \mathbf{y}, a, t) + \epsilon\, \varphi(\mathbf{x}).$$

The resulting sequence of instances is then non-trivial and state evolution applies. Call $\Sigma^t(\varepsilon)$ the resulting state evolution sequence, and denote by $x^t(\epsilon)$ the corresponding orbit. Applying Theorem 1, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \psi(\mathbf{x}_j^t(\epsilon), \mathbf{y}_i) = \mathbb{E}\{\psi(Z_a^t(\epsilon), Y_a)\}, \tag{66}$$

with $Z_a^t(\epsilon) \sim \mathsf{N}(0, \Sigma^t(\epsilon))$. In order to prove the same theorem for the orbit $\{x^t\}_{t\ge 0}$, we need to show the following two facts:

(*i*) $\lim_{\epsilon\to 0} \mathbb{E}\{\psi(Z_a^t(\epsilon), Y_a)\} = \mathbb{E}\{\psi(Z_a^t, Y_a)\}$, with $Z_a^t \sim \mathsf{N}(0, \Sigma^t)$.

(*ii*) Let $a_N(\epsilon) = \frac{1}{N}\sum_{i=1}^N \psi(\mathbf{x}_i^t(\epsilon), \mathbf{y}_i)$. Then $|a_N(\epsilon) - a_N(0)| \le C\epsilon$, with constant C being independent of N.

Given (*i*)-(*ii*), we have

$$\begin{aligned}\lim_{N\to\infty} \big|a_N(0) - \mathbb{E}\{\psi(Z_a^t, Y_a)\}\big| &\le \limsup_{N\to\infty} \Big\{ \big|a_N(0) - a_N(\epsilon)\big| + \big|a_N(\epsilon) - \mathbb{E}\{\psi(Z_a^t(\epsilon), Y_a)\}\big| \Big\} \\ &\quad + \big|\mathbb{E}\{\psi(Z_a^t(\epsilon), Y_a)\} - \mathbb{E}\{\psi(Z_a^t, Y_a)\}\big| \\ &\le C\varepsilon + 0 + \big|\mathbb{E}\{\psi(Z_a^t(\epsilon), Y_a)\} - \mathbb{E}\{\psi(Z_a^t, Y_a)\}\big|,\end{aligned}$$

where the last step follows from (*ii*) and Eq. (66). Therefore, taking the limit of both sides as $\epsilon \to 0$,

$$\lim_{N\to\infty} \big|a_N(0) - \mathbb{E}\{\psi(Z_a^t, Y_a)\}\big| \le \lim_{\epsilon\to 0} C\varepsilon + \lim_{\epsilon\to 0} \big|\mathbb{E}\{\psi(Z_a^t(\epsilon), Y_a)\} - \mathbb{E}\{\psi(Z_a^t, Y_a)\}\big| = 0\,,$$

where the last step follows from (*i*). This proves Theorem 1 for $\{x^t\}_{t\ge 0}$.

It remains to prove facts (i)-(ii). The claim in (i) follows readily by applying dominated convergence theorem and noting that $\psi(\cdot,\cdot)$ is Lipschitz continuous.

To prove (ii), write

$$\begin{aligned}
&|a_N(\epsilon) - a_N(0)| \\
&\le \frac{1}{N}\sum_{i=1}^{N} \left|\psi(\mathbf{x}_i^t(\epsilon), \mathbf{y}_i) - \psi(\mathbf{x}_i^t, \mathbf{y}_i)\right| \\
&\le \frac{L'}{N}\sum_{i=1}^{N} (1 + \|\mathbf{x}_i^t(\epsilon)\|^{k-1} + \|\mathbf{x}_i^t\|^{k-1} + \|\mathbf{y}_i\|^{k-1})\, \|\mathbf{x}_i^t(\epsilon) - \mathbf{x}_i^t\| \\
&\le \frac{L'}{N}\Big\{\sum_{i=1}^{N} (1 + \|\mathbf{x}_i^t(\epsilon)\|^{k-1} + \|\mathbf{x}_i^t\|^{k-1} + \|\mathbf{y}_i\|^{k-1})^2\Big\}^{\frac{1}{2}} \Big\{\sum_{i=1}^{N} \|\mathbf{x}_i^t(\epsilon) - \mathbf{x}_i^t\|^2\Big\}^{\frac{1}{2}} \\
&\le 3L'\Big\{1 + \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i^t(\epsilon)\|^{2k-2} + \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i^t\|^{2k-2} + \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{y}_i\|^{2k-2}\Big\}^{\frac{1}{2}} \Big\{\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i^t(\epsilon) - \mathbf{x}_i^t\|^2\Big\}^{\frac{1}{2}},
\end{aligned}$$

where second inequality holds since $\psi \in \mathrm{PL}(k)$ and third inequality follows by using Cauchy-Schwartz inequality. In the last expression, the term in the first braces is bounded using the assumption on the second moment of y and using part (e) of Lemma 2 for orbit $\{x^t(\epsilon)\}$. To bound the second braces, note that both A and B_t in the AMP iteration (5) have bounded operator norm (the former with probability $1 - e^{-\Theta(N)}$). Since $g(\,\cdot\,, t)$ is Lipschitz continuous and $\varphi(\mathbf{x})$ is bounded by assumption, we conclude that $\|x^t(\varepsilon) - x^t\|^2 \le c^t N\varepsilon^2$ for some absolute constant c. This completes the proof of fact (ii).

4.3 Proof of Lemma 2

The proof is by induction on t. Let $\mathcal{B}_t$ be the property that (59), (60), (61), (63), (64), and (65) hold.

4.3.1 Induction basis: $\mathcal{B}_0$

Note that $x^1 = Am^0$.

(a) $\mathfrak{S}_0$ is generated by y, x^0 and m^0. Also $m^0 = m_\perp^0$ since M_{-1} is an empty matrix. Hence

$$x^1|_{\mathfrak{S}_1} = Am_\perp^0.$$

(b) Let $\phi : \mathcal{V}_{q,2} \to \mathbb{R}$ be a pseudo-Lipschitz function of order k. Hence, $|\phi(x)| \le L(1 + \|x\|^k)$. Given m^0, y, the random variable $\sum_{i\in C_a^N} \phi([Am^0]_i, \mathbf{y}_i)/|C_a^N|$ is a sum of independent random variables. By Lemma 4(a) $[Am^0]_i \overset{\mathrm{d}}{=} Z$ for $Z \sim \mathsf{N}(0, \langle m^0, m^0\rangle)$. Using Eq. (7),

$$\begin{aligned}
\lim_{N\to\infty} \langle m^0, m^0\rangle &= \sum_{a\in[q]} c_a \left(\lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{i\in C_a^N} (\mathbf{m}_i^0)^{\mathsf{T}} \mathbf{m}_i^0\right) \\
&= \sum_{a\in[q]} c_a \widehat{\Sigma}_a^0 = \Sigma^1.
\end{aligned}$$

Hence for all $p \geq 1$, there exists a constant c_p such that $\mathbb{E}\{\|[Am^0]_i\|^p\} \leq \|\langle m^0_\perp, m^0_\perp\rangle\|_2^{\frac{p}{2}}\, \mathbb{E}_Z\|Z\|^p < c_p$, with $Z \sim \mathsf{N}(0, \mathrm{I}_q)$. Next, we check conditions of Theorem 2 for $X_{N,i} \equiv \phi(\mathbf{x}^1_i, \mathbf{y}_i) - \mathbb{E}_A\{\phi(\mathbf{x}^1_i, \mathbf{y}_i)\}$ for $\kappa > 0$,

$$
\begin{aligned}
&\frac{1}{|C^N_a|}\sum_{i\in C^N_a}\mathbb{E}|X_{N,i}|^{2+\kappa} \qquad (67)\\
&= \frac{1}{|C^N_a|}\sum_{i\in C^N_a}\mathbb{E}\big|\phi(\mathbf{x}^1_i, \mathbf{y}_i) - \mathbb{E}_A\{\phi(\mathbf{x}^1_i, \mathbf{y}_i)\}\big|^{2+\kappa}\\
&= \frac{1}{|C^N_a|}\sum_{i\in C^N_a}\Big|\mathbb{E}_{A,\tilde{A}}\Big\{\phi([\tilde{A}m^0]_i, \mathbf{y}_i) - \phi([Am^0]_i, \mathbf{y}_i)\Big\}\Big|^{2+\kappa}\\
&\leq \frac{1}{|C^N_a|}\sum_{i\in C^N_a}\Big|\mathbb{E}_{A,\tilde{A}}\Big\{\phi([\tilde{A}m^0]_i, \mathbf{y}_i) - \phi([Am^0]_i, \mathbf{y}_i)\Big\}\Big|^{2+\kappa}\\
&\leq \frac{1}{|C^N_a|}\sum_{i\in C^N_a}\Big|\mathbb{E}_{A,\tilde{A}}\Big\{\|[\tilde{A}m^0]_i - [Am^0]_i\|(1 + \|\mathbf{y}_i\|^{k-1} + \|[\tilde{A}m^0]_i\|^{k-1} + \|[Am^0]_i\|^{k-1})\Big\}\Big|^{2+\kappa}\\
&\leq c + \frac{L'c'}{|C^N_a|}\sum_{i\in C^N_a}\|\mathbf{y}_i\|^{(k-1)(2+\kappa)}\\
&\leq c + L'c'|C^N_a|^{\kappa/2}\left(\frac{1}{|C^N_a|}\sum_{i\in C^N_a}\|\mathbf{y}_i\|^{2(k-1)}\right)^{1+\kappa/2} \leq c''|C^N_a|^{\kappa/2}.
\end{aligned}
$$

Here $\tilde{A}$ is an independent copy of A, and the last inequality uses assumption on empirical moments of $\{\mathbf{y}_i\}_{i\in C^N_a}$. By applying Theorem 2, we get

$$
\lim_{N\to\infty}\frac{1}{|C^N_a|}\sum_{i\in C^N_a}\big[\phi(\mathbf{x}^1_i, \mathbf{y}_i) - \mathbb{E}_A\{\phi(\mathbf{x}^1_i, \mathbf{y}_i)\}\big] \overset{\text{a.s.}}{=} 0.
$$

Hence, using Lemma 6 for $v = w$ and $\psi(\mathbf{y}_i) = \mathbb{E}_Z\{\phi(Z, \mathbf{y}_i)\}$ we get

$$
\lim_{N\to\infty}\frac{1}{C^N_a}\sum_{i\in C^N_a}\mathbb{E}_Z[\phi(\mathbf{x}^1_i, \mathbf{y}_i)] \overset{\text{a.s.}}{=} \mathbb{E}\big\{\phi(Z_a, Y_a)\big\},
$$

with $Z_a \sim \mathsf{N}(0, \Sigma^1)$ independent of $Y_a \sim P_a$. Note that ψ belongs to $\mathrm{PL}(k)$ since ϕ belongs to $\mathrm{PL}(k)$.

(*c*) Let $\hat{A} = A_{C^N_a}$ be the submatrix formed by the rows in C^N_A. Using Lemma 4(c), conditioned on m^0,

$$
\lim_{N\to\infty}\langle x^1_{C^N_a}, x^1_{C^N_a}\rangle = \lim_{N\to\infty}\langle \hat{A}m^0, \hat{A}m^0\rangle \overset{\text{a.s.}}{=} \lim_{N\to\infty}\langle m^0, m^0\rangle = \Sigma^1\,.
$$

(*d*) Write

$$
\lim_{N\to\infty}\langle x^1_{C^N_a}, \varphi(x^1_{C^N_a}, y_{C^N_a})\rangle = \lim_{N\to\infty}\frac{1}{|C^N_a|}\sum_{i\in C^N_a}\mathbf{x}^1_i[\varphi^a(\mathbf{x}^1_i, \mathbf{y}_i)]^{\mathsf{T}} \overset{\text{a.s.}}{=} \mathbb{E}(Z_a[\varphi^a(Z_a, Y_a)]^{\mathsf{T}}),
$$

where the last step follows by applying $\mathcal{B}_0(b)$ to the functions $\phi(\mathbf{x}_i^1, \mathbf{y}_i) = \mathbf{x}_i^1(l)[\varphi^a(\mathbf{x}_i^1, \mathbf{y}_i)]_k$, for all $l, k \in [q]$. Furthermore, using Lemma 5,

$$\mathbb{E}(Z_a[\varphi^a(Z_a, Y_a)]^{\mathsf{T}}) = \Sigma^1 \, \mathbb{E}([\frac{\partial \varphi^a}{\partial \mathbf{z}}(Z_a, Y_a)]^{\mathsf{T}}) \, .$$

As proved in part (c), $\lim_{N\to\infty} \langle x^1_{C_a^N}, x^1_{C_a^N} \rangle = \Sigma^1$. Also, by part (b), the empirical distribution of $\{(\mathbf{x}_i^1, \mathbf{y}_i)\}_{i \in C_a^N}$ converges weakly to the distribution of (Z_a, Y_a), and consequently we get

$$\lim_{N\to\infty} \langle \nabla \varphi^a(x^1_{C_a^N}, y_{C_a^N}) \rangle = \lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} [\frac{\partial \varphi^a}{\partial \mathbf{x}}(\mathbf{x}_i^1, \mathbf{y}_i)]^{\mathsf{T}} \stackrel{\text{a.s.}}{=} \mathbb{E}([\frac{\partial \varphi^a}{\partial \mathbf{z}}(Z_a, Y_a)]^{\mathsf{T}}) \, .$$

This proves Eq. (62). To prove Eq. (63), notice that

$$\langle x^1, \varphi(x^1, y) \rangle = \sum_{a \in [q]} c_a \langle x^1_{C_a^N}, \varphi(x^1_{C_a^N}, y_{C_a^N}) \rangle \, . \tag{68}$$

Also,

$$\lim_{N\to\infty} \langle x^1, x^1 \rangle = \sum_{a \in [q]} c_a \lim_{N\to\infty} \langle x^1_{C_a^N}, x^1_{C_a^N} \rangle = \sum_{a \in [q]} c_a \Sigma^1 = \Sigma^1, \tag{69}$$

where the last step holds since $\sum_{a\in[q]} c_a = 1$. Further,

$$\langle \nabla \varphi(x^1, y) \rangle = \sum_{a \in [q]} c_a \langle \nabla \varphi^a(x^1_{C_a^N}, y_{C_a^N}) \rangle \tag{70}$$

Combining Eqs. (68), (69), (70) and Eq. (62), we get the desired result.

(e) Similar to (b), conditioning on m^0, the term $\sum_{i \in C_a^N} \|[Am^0]_i\|^{2\ell} / |C_a^N|$ is sum of independent random variables and

$$\mathbb{E}\{\|[Am^0]_i\|^p\} \le \|\langle m^0_\perp, m^0_\perp \rangle^{\frac{1}{2}}\|_2^p \, \mathbb{E}\{\|Z\|^p\} < c_p \, ,$$

for a constant c_p. Therefore, by Theorem 2, we get

$$\lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \Big[\|[Am^0]_i\|^{2\ell} - \mathbb{E}_A\{\|[Am^0]_i\|^{2\ell}\} \Big] \stackrel{\text{a.s.}}{=} 0.$$

But, $\frac{1}{|C_a^N|} \sum_{i \in C_a^N} \mathbb{E}_A\{\|[Am^0]_i\|^{2\ell}\} \le \|\langle m^0, m^0 \rangle^{\frac{1}{2}}\|_2^\ell \, \mathbb{E}_Z\{\|Z\|^{2\ell}\} < \infty$.

(f) Since $t = 0$ and $m^0 = m^0_\perp$, the result follows from $\lim_{N\to\infty} \langle m^0, m^0 \rangle = \Sigma^1$ and that $\Sigma^1 = \sum_{b\in[q]} c_b \widehat{\Sigma}_b^0 \succ 0$.

4.3.2 Proof of $\mathcal{B}_t$:

Suppose that $\mathcal{B}_{t-1}$ holds. We prove $\mathcal{B}_t$.

(f) It is sufficient to consider $r = t$. Write $m^t_\perp = m^t - m^t_\parallel$ and recall that $m^t_\parallel = \sum_{s=0}^{t-1} m^s \alpha_s$. Hence, for any $\gamma_0 \in \mathbb{R}^q$, with $\|\gamma_0\| = 1$, we have

$$\gamma_0^\mathsf{T} \langle m^t_\perp, m^t_\perp \rangle \gamma_0 = \frac{1}{N} \sum_{i=1}^N \left(\gamma_0^\mathsf{T} \mathbf{m}_i^t - \sum_{s=0}^{t-1} \gamma_0^\mathsf{T} \alpha_s^\mathsf{T} \mathbf{m}_i^s \right) \left(\gamma_0^\mathsf{T} \mathbf{m}_i^t - \sum_{s=0}^{t-1} \gamma_0^\mathsf{T} \alpha_s^\mathsf{T} \mathbf{m}_i^s \right)^\mathsf{T} .$$

Note that the matrix

$$\alpha = (\alpha_0, \dots, \alpha_{t-1}) = \Big[\frac{M_{t-1}^\mathsf{T} M_{t-1}}{N} \Big]^{-1} \frac{M_{t-1}^\mathsf{T} m^t}{N} ,$$

has a finite limit as $N \to \infty$ by the induction hypothesis $\mathcal{B}_{t-1}(b)$. Furthermore, $\mathbf{m}_i^t = g(\mathbf{x}_i^t, \mathbf{y}_i, a, t)$, for $i \in C_a^N$. By induction hypothesis $\mathcal{B}_{t-1}(a)$, it is sufficient to show that there exists $\rho > 0$ depending on t such that,

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \left(\gamma_0^\mathsf{T} g(\mathbf{Z} + \sum_{r=1}^{t-1} \alpha_{r-1}^\mathsf{T} \mathbf{x}_i^r, \mathbf{y}_i, a, t) - \sum_{s=0}^{t-1} \gamma_0^\mathsf{T} \alpha_s^\mathsf{T} \mathbf{m}_i^s \right) \cdot \\ \left(\gamma_0^\mathsf{T} g(\mathbf{Z} + \sum_{r=1}^{t-1} \alpha_{r-1}^\mathsf{T} \mathbf{x}_i^r, \mathbf{y}_i, a, t) - \sum_{s=0}^{t-1} \gamma_0^\mathsf{T} \alpha_s^\mathsf{T} \mathbf{m}_i^s \right)^\mathsf{T} \geq 2\rho , \end{aligned} \tag{71}$$

where $\mathbf{Z} = (\tilde{A} m_\perp^{t-1})^\mathsf{T} e_i \in \mathbb{R}^q$ (e_i being the i-the element of the canonical basis). By the strong law of large numbers for triangular arrays, the above is lower bounded by

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\tilde{A}} \Big[\gamma_0^\mathsf{T} g(\mathbf{Z} + \sum_{r=1}^{t-1} \alpha_{r-1}^\mathsf{T} \mathbf{x}_i^r, \mathbf{y}_i, a, t) - \sum_{s=0}^{t-1} \gamma_0^\mathsf{T} \alpha_s^\mathsf{T} \mathbf{m}_i^s \Big] \cdot \\ \mathbb{E}_{\tilde{A}} \Big[\gamma_0^\mathsf{T} g(\mathbf{Z} + \sum_{r=1}^{t-1} \alpha_{r-1}^\mathsf{T} \mathbf{x}_i^r, \mathbf{y}_i, a, t) - \sum_{s=0}^{t-1} \gamma_0^\mathsf{T} \alpha_s^\mathsf{T} \mathbf{m}_i^s \Big]^\mathsf{T} \\ \geq \liminf_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \mathrm{Var}_{\mathbf{Z}} \Big(\gamma_0^\mathsf{T} g(\mathbf{Z} + \sum_{r=1}^{t-1} \alpha_{r-1}^\mathsf{T} \mathbf{x}_i^r, \mathbf{y}_i, a, t) \Big) . \end{aligned}$$

The variance in the last expression is taken only with respect to $\tilde{A}$ or, equivalently, with respect to $\mathbf{Z} \sim \mathsf{N}(0, \langle m_\perp^{t-1}, m_\perp^{t-1} \rangle)$. Notice that the covariance of $\mathbf{Z}$ is lower bounded by $\rho' \, \mathrm{I}_{q\times q}$ for some $\rho' > 0$, by the induction hypothesis $\mathcal{B}_{t-1}(f)$. It is a straightforward probability exercise to show that, for any non-constant continuous function $G : \mathbb{R}^q \to \mathbb{R}$, and any $U > 0$ there exists $\varepsilon > 0$ such that

$$\inf_{\|\vec{\alpha}\|_2 \leq U} \mathrm{Var}_{\mathbf{Z}} \big(G(\vec{\alpha} + \mathbf{Z}) \big) \geq \varepsilon .$$

Using $\mathcal{B}_{t-1}(e)$, we can choose U large enough to ensure that there exists at least $N/2$ values of the indices $i \in [N]$ such that $\| \sum_{r=0}^{t-1} \alpha_{r-1}^\mathsf{T} \mathbf{x}_i^r \| \leq U$. Note that U and therefore ε depend on t but do not depend on N. The lower bound (71) follows then by taking $\rho = \varepsilon/4$.

(*a*) Let $\mathbf{B}_t \in \mathbb{R}^{q\times q}$ be given by $\mathbf{B}_t = \frac{1}{N}\left(\sum_{j\in[N]} \frac{\partial f^j}{\partial \mathbf{x}}(\mathbf{x}_j^t, t)\right)$. Further let $\mathscr{B}$ be a square block-diagonal matrix of size tq with matrices $\mathbf{B}_0^\mathsf{T}, \dots, \mathbf{B}_{t-1}^\mathsf{T}$ on its diagonal. Define $X_{t-1} = [x^1|x^2|\dots|x^t]$. Recalling the definition of Y_{t-1} and M_{t-1} from Section 4.1,

$$Y_{t-1} = X_{t-1} + [0_{N\times q}|M_{t-2}]\mathscr{B}\,.$$

Lemma 3. *The following holds*

$$x^{t+1}|_{\mathfrak{S}_t} \stackrel{\mathrm{d}}{=} X_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}M_{t-1}^\mathsf{T}m_\parallel^t + P^\perp_{M_{t-1}}\tilde{A}m_\perp^t + M_{t-1}\vec{o}_{t-1}(1)\,.$$

Proof. Lemma 10 in [BM11] implies that $A|_{\mathfrak{S}_t} \stackrel{\mathrm{d}}{=} E(A|_{\mathfrak{S}_t}) + \mathcal{P}_t(\tilde{A})$, where $\tilde{A} \stackrel{\mathrm{d}}{=} A$ is a random matrix independent of $\mathfrak{S}_t$ and $\mathcal{P}_t$ is the orthogonal projector onto subspace $V_t = \{A|AM_{t-1} = 0, A = A^\mathsf{T}\}$. Following the same argument as in [BM11], we have

$$\begin{aligned} E(A|_{\mathfrak{S}_t}) &= Y_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}M_{t-1}^\mathsf{T} + M_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}Y_{t-1}^\mathsf{T} \\ &\quad -M_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}M_{t-1}^\mathsf{T}Y_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}M_{t-1}^\mathsf{T}\,. \\ \mathcal{P}_t(\tilde{A}) &= P^\perp_{M_{t-1}}\tilde{A}P^\perp_{M_{t-1}}\,. \end{aligned}$$

Using $M_{t-1}^\mathsf{T}m_\perp^t = 0$ and $Y_{t-1} = AM_{t-1}$, it is immediate to see that

$$A|_{\mathfrak{S}_t}m^t \stackrel{\mathrm{d}}{=} Y_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}M_{t-1}^\mathsf{T}m_\parallel^t + M_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}Y_{t-1}^\mathsf{T}m_\perp^t + P^\perp_{M_{t-1}}\tilde{A}m_\perp^t\,.$$

Moreover, $Y_{t-1}^\mathsf{T}m_\perp^t = X_{t-1}^\mathsf{T}m_\perp^t$ because $M_{t-2}^\mathsf{T}m_\perp^t = 0$. Recalling $m_\parallel^t = M_{t-1}\alpha$ we need to show

$$[0|M_{t-2}]\mathscr{B}\alpha + M_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}X_{t-1}^\mathsf{T}m_\perp^t - m^{t-1}\mathbf{B}_t^\mathsf{T} = M_{t-1}\vec{o}_{t-1}(1)\,. \tag{72}$$

Note that we used the fact $\mathsf{B}_t m^{t-1} = m^{t-1}\mathbf{B}_t^\mathsf{T}$ which follows from our convention $\mathcal{V}_{q,N} \equiv \mathbb{R}^{N\times q}$.

Here is our strategy to prove (72). The left hand side is a linear combination of $m^0, \dots, m^{t-1}$. For any $\ell = 1, \dots, t$ we will prove that the coefficient of $m^{\ell-1}$ converges to 0. Note that the coefficients are matrices of size q. The coefficient of $m^{\ell-1}$ in the left hand side is equal to

$$\left[(M_{t-1}^\mathsf{T}M_{t-1})^{-1}X_{t-1}^\mathsf{T}m_\perp^t\right]_{(\ell)} - \mathbf{B}_\ell^\mathsf{T}(-\alpha_\ell)^{\mathbb{I}_{\ell\neq t}} = \sum_{r=1}^t \left[(\frac{M_{t-1}^\mathsf{T}M_{t-1}}{N})^{-1}\right]_{(\ell),(r)} \langle x^r, m^t - \sum_{s=0}^{t-1} m^s\alpha_s\rangle - \mathbf{B}_\ell^\mathsf{T}(-\alpha_\ell)^{\mathbb{I}_{\ell\neq t}}\,.$$

To simplify the notation denote the matrix $M_{t-1}^\mathsf{T}M_{t-1}/N$ by G. Therefore,

$$\lim_{N\to\infty} \text{Coefficient of } m^{\ell-1} = \lim_{N\to\infty}\left\{\sum_{r=1}^t (G^{-1})_{(\ell),(r)}\langle x^r, m^t - \sum_{s=0}^{t-1} m^s\alpha_s\rangle - \mathbf{B}_\ell^\mathsf{T}(-\alpha_\ell)^{\mathbb{I}_{\ell\neq t}}\right\}.$$

But using the induction hypothesis $\mathcal{B}_{t-1}(d)$ for $\varphi = f(\cdot;1), \dots, f(\cdot;t)$, the term $\langle x^r, m^t - \sum_{s=0}^{t-1} m^s\alpha_s\rangle$ is almost surely equal to the limit of $\langle x^r, x^t\rangle\mathbf{B}_t^\mathsf{T} - \sum_{s=0}^{t-1}\langle x^r, x^s\rangle\mathbf{B}_s^\mathsf{T}\alpha_s$. This can be

modified, using the induction hypothesis $\mathcal{B}_{t-1}(c)$, to $\langle m^{r-1}, m^{t-1}\rangle \mathbf{B}_t^\mathsf{T} - \sum_{s=0}^{t-1}\langle m^{r-1}, m^{s-1}\rangle \mathbf{B}_s^\mathsf{T}\alpha_s$ almost surely, which can be written as $G_{(r),(t)}\mathbf{B}_t^\mathsf{T} - \sum_{s=0}^{t-1} G_{(r),(s)}\mathbf{B}_s^\mathsf{T}\alpha_s$. Hence,

$$\begin{aligned}\lim_{N\to\infty} \text{Coefficient of } m^{\ell-1} &\overset{\text{a.s.}}{=} \lim_{N\to\infty}\left\{\sum_{r=1}^{t}(G^{-1})_{(\ell),(r)}[G_{(r),(t)}\mathbf{B}_t^\mathsf{T} - \sum_{s=0}^{t-1} G_{(r),(s)}\mathbf{B}_s^\mathsf{T}\alpha_s] - \mathbf{B}_\ell^\mathsf{T}(-\alpha_\ell)^{\mathbb{I}_{\ell\neq t}}\right\}\\ &\overset{\text{a.s.}}{=} \lim_{N\to\infty}\left\{\mathbf{B}_t^\mathsf{T}\mathbb{I}_{t=\ell} - \sum_{s=0}^{t-1}\mathbf{B}_s^\mathsf{T}\alpha_s\mathbb{I}_{\ell=s} - \mathbf{B}_\ell^\mathsf{T}(-\alpha_\ell)^{\mathbb{I}_{\ell\neq t}}\right\}\\ &\overset{\text{a.s.}}{=} 0\,.\end{aligned}$$

Notice that in the above equalities we used the fact that G has, almost surely, a non-singular limit as $N\to\infty$ which was discussed in part (f). □

The proof of Eq. (59) follows immediately since the last lemma yields

$$x^{t+1}|_{\mathfrak{S}_t} \overset{\text{d}}{=} \sum_{i=0}^{t-1} x^{i+1}\alpha_i + \tilde{A}m_\perp^t - M_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}M_{t-1}^\mathsf{T}\tilde{A}m_\perp^t + M_{t-1}\vec{o}_{t-1}(1)\,.$$

Note that, using Lemma 4(d), as $N\to\infty$,

$$M_{t-1}(M_{t-1}^\mathsf{T}M_{t-1})^{-1}M_{t-1}^\mathsf{T}\tilde{A}m_\perp^t \overset{\text{d}}{=} \tilde{M}_{t-1}\vec{o}_{t-1}(1)\,,$$

which finishes the proof since $\tilde{M}_{t-1}\vec{o}_{t-1}(1) + M_{t-1}\vec{o}_{t-1}(1) = \tilde{M}_{t-1}\vec{o}_{t-1}(1)$.

(c) For $r, s < t$ we can use induction hypothesis. For $r = t, s < t$,

$$\langle x_{C_a^N}^{t+1}, x_{C_a^N}^{s+1}\rangle|_{\mathfrak{S}_t} \overset{\text{d}}{=} \sum_{i=0}^{t-1}\alpha_i^\mathsf{T}\langle x_{C_a^N}^{i+1}, x_{C_a^N}^{s+1}\rangle + \langle [P_{M_{t-1}}^\perp \tilde{A}m_\perp^t]_{C_a^N}, x_{C_a^N}^{s+1}\rangle + \sum_{i=0}^{t-1}\vec{o}_1(1)\langle m_{C_a^N}^i, x_{C_a^N}^{s+1}\rangle.$$

Now, by induction hypothesis $\mathcal{B}_{t-1}(d)$, for $\varphi(\mathbf{v},\mathbf{u}) = g(\mathbf{v},\mathbf{u},a,i)$, each term $\langle m_{C_a^N}^i, x_{C_a^N}^{s+1}\rangle$ has a finite limit. Thus,

$$\lim_{N\to\infty}\sum_{i=0}^{t-1}\vec{o}_1(1)\langle m_{C_a^N}^i, x_{C_a^N}^{s+1}\rangle \overset{\text{a.s.}}{=} 0.$$

We can use Lemma 4 (b)-(d) for $\langle [P_{M_{t-1}}^\perp \tilde{A}m_\perp^t]_{C_a^N}, x_{C_a^N}^{s+1}\rangle$ to obtain $\langle [P_{M_{t-1}}^\perp \tilde{A}m_\perp^t]_{C_a^N}, x_{C_a^N}^{s+1}\rangle \to 0$, almost surely. Finally, using induction hypothesis $\mathcal{B}_s(c)$ or $\mathcal{B}_i(c)$ for each term of the form $\langle x_{C_a^N}^i, x_{C_a^N}^{s+1}\rangle$

$$\begin{aligned}\lim_{N\to\infty}\langle x_{C_a^N}^{t+1}, x_{C_a^N}^{s+1}\rangle &\overset{\text{a.s.}}{=} \lim_{N\to\infty}\sum_{i=0}^{t-1}\alpha_i^\mathsf{T}\langle m^i, m^s\rangle\\ &\overset{\text{a.s.}}{=} \lim_{N\to\infty}\langle m_\parallel^t, m^s\rangle \overset{\text{a.s.}}{=} \lim_{N\to\infty}\langle m^t, m^s\rangle\,,\end{aligned}$$

where the last line uses the definition of α_i and $m_\perp^t \perp m^s$.

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

$$\langle x^{t+1}_{C^N_a}, x^{t+1}_{C^N_a}\rangle|_{\mathfrak{S}_t} \stackrel{\mathrm{d}}{=} \sum_{i,j=0}^{t-1} \alpha_i^{\mathsf{T}} \langle x^{i+1}_{C^N_a}, x^{j+1}_{C^N_a}\rangle \alpha_j + \langle [P^{\perp}_{M_{t-1}} \tilde{A} m^t_{\perp}]_{C^N_a}, [P^{\perp}_{M_{t-1}} \tilde{A} m^t_{\perp}]_{C^N_a}\rangle + \vec{o}_1(1).$$

Note that the contribution of all products of the form$\langle [P^{\perp}_{M_{t-1}} \tilde{A} m^t_{\perp}]_{C^N_a}, x^{i+1}_{C^N_a}\rangle$ almost surely tend to 0. Now, using induction hypothesis $\mathcal{B}_i(c)$ and Lemma 4 (c), we obtain

$$\begin{aligned}
\lim_{N\to\infty} \langle x^{t+1}_{C^N_a}, x^{t+1}_{C^N_a}\rangle|_{\mathfrak{S}_t} &\stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \sum_{i,j=0}^{t-1} \alpha_i^{\mathsf{T}} \langle m^i, m^j\rangle \alpha_j + \lim_{N\to\infty} \langle m^t_{\perp}, m^t_{\perp}\rangle \\
&\stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \langle m^t_{\|}, m^t_{\|}\rangle + \lim_{N\to\infty} \langle m^t_{\perp}, m^t_{\perp}\rangle \\
&\stackrel{\text{a.s.}}{=} \lim_{N\to\infty} \langle m^t, m^t\rangle.
\end{aligned}$$

(e) This part follows by a very similar argument to the one in the proof of Lemma 1 (Step $\mathcal{B}_t(e)$) in [BM11].

(b) Using part (a) we can write

$$\phi(\mathbf{x}^1_i, \ldots, \mathbf{x}^{t+1}_i, \mathbf{y}_i)|_{\mathfrak{S}_{t,t}} \stackrel{\mathrm{d}}{=} \phi\left(\mathbf{x}^1_i, \ldots, \mathbf{x}^t_i, \left[\sum_{r=0}^{t-1} x^{r+1}\alpha_r + \tilde{A} m^t_{\perp} + \tilde{M}_{t-1}\vec{o}_{t-1}(1)\right]_i, \mathbf{y}_i\right).$$

We show that we can drop the error term $\tilde{M}_{t-1}\vec{o}_{t-1}(1)$. Indeed, defining

$$\begin{aligned}
a_i &= \left(\mathbf{x}^1_i, \ldots, \mathbf{x}^t_i, \left[\sum_{r=0}^{t-1} x^{r+1}\alpha_r + \tilde{A} m^t_{\perp} + \tilde{M}_{t-1}\vec{o}_{t-1}(1)\right]_i, \mathbf{y}_i\right), \\
b_i &= \left(\mathbf{x}^1_i, \ldots, \mathbf{x}^t_i, \left[\sum_{r=0}^{t-1} x^{r+1}\alpha_r + \tilde{A} m^t_{\perp}\right]_i, \mathbf{y}_i\right),
\end{aligned}$$

by the pseudo-Lipschitz assumption

$$|\phi(a_i) - \phi(b_i)| \le L\left(1 + \|a_i\|^{k-1} + \|b_i\|^{k-1}\right)\left(\sum_{r=0}^{t-1} \|\tilde{m}^r_i\|\right) o(1).$$

Therefore, using Cauchy-Schwartz inequality twice, we have

$$\begin{aligned}
&\frac{1}{|C^N_a|}\Big|\sum_{i\in C^N_a} \phi(a_i) - \sum_{i\in C^N_a} \phi(b_i)\Big| \\
&\le L'\left\{1 + \frac{1}{|C^N_a|}\sum_{i\in C^N_a} \|a_i\|^{2k-2} + \frac{1}{|C^N_a|}\sum_{i\in C^N_a} \|b_i\|^{2k-2}\right\}^{\frac{1}{2}} \left\{\frac{1}{|C^N_a|}\sum_{r=0}^{t-1} \|\tilde{m}^r\|^2\right\}^{\frac{1}{2}} t^{\frac{1}{2}} o(1). \qquad (73)
\end{aligned}$$

Also note that

$$\frac{1}{|C^N_a|}\sum_{i\in C^N_a} \|a_i\|^{2\ell} \le (t+1)^{\ell}\Big\{\sum_{r=0}^{t} \frac{1}{|C^N_a|}\sum_{i\in C^N_a} \|\mathbf{x}^{r+1}_i\|^{2\ell} + \frac{1}{|C^N_a|}\sum_{i\in C^N_a} \|\mathbf{y}_i\|^{2\ell}\Big\},$$

which is finite almost surely (for $\ell = k-1$) using $\mathcal{B}_r(e)$ for $r \in [t]$ and the assumption on (the moment of) y. The term $|C_a^N|^{-1} \sum_{i \in C_a^N} \|b_i\|^{2\ell}$ is bounded almost surely since

$$\begin{aligned}\frac{1}{|C_a^N|} \sum_{i \in C_a^N} \|b_i\|^{2\ell} &\le \frac{C}{|C_a^N|} \sum_{i \in C_a^N} \|a_i\|^{2\ell} + C \sum_{r=0}^{t-1} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \|\tilde{\mathbf{m}}_i^r\|^{2\ell} o(1) \\ &\le \frac{C}{|C_a^N|} \sum_{i \in C_a^N} \|a_i\|^{2\ell} + C' \sum_{r=0}^{t-1} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \|\mathbf{m}_i^r\|^{2\ell} o(1)\,,\end{aligned}$$

where the last inequality follows from the fact that $[M_{t-1}^{\mathsf{T}} M_{t-1}/N]$ has almost surely a non-singular limit as $N \to \infty$, as discussed in point (f) above. Finally, for $r \le t-1$, each term $(1/|C_a^N|) \sum_{i \in C_a^N} \|\mathbf{m}_i^r\|^{2\ell}$ can be easily proved to be bounded using the induction hypothesis $\mathcal{B}_{t-1}(e)$.

Hence for any fixed t, (73) vanishes almost surely when N goes to ∞.

Now given, $\mathbf{x}^1, \dots, \mathbf{x}^t$, consider the random variables

$$\tilde{X}_i = \phi\left(\mathbf{x}_i^1, \dots, \mathbf{x}_i^t, \sum_{r=0}^{t-1} \alpha_r^{\mathsf{T}} \mathbf{x}_i^{r+1} + (\tilde{A} m_\perp^t)_i, \mathbf{y}_i\right)$$

and $X_i \equiv \tilde{X}_i - \mathbb{E}_{\tilde{A}}\{\tilde{X}_i\}$. Proceeding as in $\mathcal{B}_0$, and using the pseudo-Lipschitz property of ϕ, it is easy to check the conditions of Theorem 2. We therefore get

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \Big[&\phi\Big(\mathbf{x}_i^1, \dots, \mathbf{x}_i^t, \big[\sum_{r=0}^{t-1} x^{r+1} \alpha_r + \tilde{A} m_\perp^t\big]_i, \mathbf{y}_i\Big) \\ &- \mathbb{E}_{\tilde{A}}\Big\{\phi\Big(\mathbf{x}_i^1, \dots, \mathbf{x}_i^t, \big[\sum_{r=0}^{t-1} x^{r+1} \alpha_r + \tilde{A} m_\perp^t\big]_i, \mathbf{y}_i\Big)\Big\}\Big] \overset{\text{a.s.}}{=} 0. \quad (74)\end{aligned}$$

Note that $[\tilde{A} m_\perp^t]_i$ is a gaussian random vector with covariance $\langle m_\perp^t, m_\perp^t \rangle$. Further $\langle m_\perp^t, m_\perp^t \rangle$ converges to a finite limit Γ_t^2 almost surely as $N \to \infty$. Indeed $\langle m_\perp^t, m_\perp^t \rangle = \langle m^t, m^t \rangle - \langle m_\parallel^t, m_\parallel^t \rangle$. By $\mathcal{B}_t(c)$, $\langle m^t, m^t \rangle$ converges to a finite limit. Further, $\langle m_\parallel^t, m_\parallel^t \rangle = \sum_{r,s=0}^{t-1} \alpha_r^{\mathsf{T}} \langle x^r, x^s \rangle \alpha_s$ also converges since the products $\langle x^r, x^s \rangle$ do and the coefficients α_r, $r \le t-1$ as discussed in $\mathcal{B}_t(f)$.

Hence we can use induction hypothesis $\mathcal{B}_{t-1}(b)$ for

$$\widehat{\phi}(\mathbf{x}_i^1, \dots, \mathbf{x}_i^t, \mathbf{y}_i) = \mathbb{E}_Z\Big\{\phi\Big(\mathbf{x}_i^1, \dots, \mathbf{x}_i^t, \sum_{r=0}^{t-1} \alpha_r^{\mathsf{T}} \mathbf{x}_i^{r+1} + \langle m_\perp^t, m_\perp^t \rangle^{\frac{1}{2}} Z, \mathbf{y}_i\Big)\Big\},$$

with $Z \sim \mathsf{N}(0, \mathrm{I}_{q\times q})$ independent of $\mathbf{x}_i^{r+1}$, $r \le t-1$, to show

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \mathbb{E}_{\tilde{A}} \left\{ \phi\left(\mathbf{x}_i^1, \dots, \mathbf{x}_i^t, \left[\sum_{r=0}^{t-1} \alpha_r^{\mathsf{T}} \mathbf{x}_i^{r+1} + \tilde{A} m_\perp^t\right]_i, \mathbf{y}_i\right)\right\} \\ \overset{\text{a.s.}}{=} \mathbb{E}\, \mathbb{E}_Z\Big\{\phi\Big(Z_a^1, \dots Z_a^t, \sum_{r=0}^{t-1} \alpha_r^{\mathsf{T}} Z_a^{r+1} + \Gamma_t\, Z, Y_a\Big)\Big\}. \quad (75)\end{aligned}$$

Note that $\sum_{r=0}^{t-1} \alpha_r^{\mathsf{T}} Z_a^{r+1} + \Gamma_t\, Z$ is a gaussian vector. All that we need, is to show that the covariance matrix of this gaussian vector is Σ^{t+1}. But using a combination of (74) and (75) for the pseudo-Lipschitz functions $\phi(\mathbf{v}_1, \ldots, \mathbf{v}_{t+1}, \mathbf{y}_i) = \mathbf{v}_{t+1}(\ell)\mathbf{v}_{t+1}(k)$, for all $\ell, k \in [q]$,

$$\lim_{N\to\infty} \langle x_{C_a^N}^{t+1}, x_{C_a^N}^{t+1} \rangle \overset{\text{a.s.}}{=} \mathbb{E}\Big\{ \Big(\sum_{r=0}^{t-1} \alpha_r^{\mathsf{T}} Z_a^{r+1} + \Gamma_t\, Z \Big) \Big(\sum_{r=0}^{t-1} \alpha_r^{\mathsf{T}} Z_a^{r+1} + \Gamma_t\, Z \Big)^{\mathsf{T}} \Big\}. \tag{76}$$

On the other hand as proved in part (c),

$$\lim_{N\to\infty} \langle x_{C_a^N}^{t+1}, x_{C_a^N}^{t+1} \rangle \overset{\text{a.s.}}{=} \lim_{N\to\infty} \langle m^t, m^t \rangle = \lim_{N\to\infty} \langle f(x^t, t), f(x^t, t) \rangle \,.$$

Hence,

$$\begin{aligned} \lim_{N\to\infty} \langle x_{C_a^N}^{t+1}, x_{C_a^N}^{t+1} \rangle &\overset{\text{a.s.}}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f^i(\mathbf{x}_i^t, t)[f^i(\mathbf{x}_i^t, t)]^{\mathsf{T}} \\ &= \sum_{a\in[q]} c_a \frac{1}{|C_a^N|} \sum_{i\in C_a^N} g(\mathbf{x}_i^t, \mathbf{y}_i, a, t) g(\mathbf{x}_i^t, \mathbf{y}_i, a, t)^{\mathsf{T}} \,. \end{aligned}$$

By induction hypothesis $\mathcal{B}_{t-1}(b)$ for the pseudo-Lipschitz functions

$$\phi(\mathbf{v}_1, \ldots, \mathbf{v}_t, \mathbf{y}_i) = [g(\mathbf{v}_t, \mathbf{y}_i, a, t)]_\ell [g(\mathbf{v}_t, \mathbf{y}_i, a, t)]_k \,,$$

for all $\ell, k \in [q]$, we get

$$\frac{1}{|C_a^N|} \sum_{i\in C_a^N} g(\mathbf{x}_i^t, \mathbf{y}_i, a, t) g(\mathbf{x}_i^t, \mathbf{y}_i, a, t)^{\mathsf{T}} \overset{\text{a.s.}}{=} \mathbb{E}\Big\{ g(Z_a^t, Y_a, a, t) g(Z_a^t, Y_a, a, t)^{\mathsf{T}} \Big\} = \widehat{\Sigma}_a^t \,.$$

Consequently,

$$\lim_{N\to\infty} \langle x_{C_a^N}^{t+1}, x_{C_a^N}^{t+1} \rangle \overset{\text{a.s.}}{=} \sum_{a\in[q]} c_a \widehat{\Sigma}_a^t = \Sigma^{t+1} \,.$$

which proves the claim.

(d) In a very similar manner to the proof of $\mathcal{B}_0(d)$, using part (b) for the pseudo-Lipschitz function $\phi : \mathcal{V}_{q,t+2} \to \mathbb{R}$ given by $\phi(\mathbf{x}_i^1, \ldots, \mathbf{x}_i^{t+1}, \mathbf{y}_i) = \mathbf{x}_i^{r+1}(l)[\varphi(\mathbf{x}_i^{s+1}, \mathbf{y}_i)]_k$, for all $l, k \in [q]$, we can obtain

$$\lim_{N\to\infty} \langle x_{C_a^N}^{r+1}, \varphi(x_{C_a^N}^{s+1}, y_{C_a^N}) \rangle \overset{\text{a.s.}}{=} \mathbb{E}(Z_a^{r+1} [\varphi(Z_a^{s+1}, Y_a)]^{\mathsf{T}}) \,,$$

for gaussian vectors $Z_a^{r+1} \sim \mathsf{N}(0, \Sigma^{r+1})$, $Z_a^{s+1} \sim \mathsf{N}(0, \Sigma^{s+1})$. Using Lemma 5, we have almost surely,

$$\mathbb{E}(Z_a^{r+1} [\varphi(Z_a^{s+1}, Y_a)]^{\mathsf{T}}) = \mathrm{Cov}(Z_a^{r+1}, Z_a^{s+1})\ \mathbb{E}([\frac{\partial \varphi^a}{\partial \mathbf{z}}(Z_a^{s+1}, Y_a)]^{\mathsf{T}}) \,.$$

By another application of part (b) for $\phi(\mathbf{x}_i^1, \ldots, \mathbf{x}_i^{t+1}, \mathbf{y}_i) = \mathbf{x}_i^{r+1}(l)\mathbf{x}_i^{s+1}(k)$ for all $l, k \in [q]$,

$$\lim_{N\to\infty} \langle x_{C_a^N}^{r+1}, x_{C_a^N}^{s+1} \rangle = \mathrm{Cov}(Z_a^{r+1}, Z_a^{s+1}) \,.$$

Similar to $\mathcal{B}_0(d)$ we also have $\lim_{N\to\infty} \langle \nabla \varphi^a(x_{C_a^N}^{s+1}, y_{C_a^N}) \rangle = \mathbb{E}([\frac{\partial \varphi^a}{\partial \mathbf{z}}(Z_a^{s+1}, Y_a)]^{\mathsf{T}})$, almost surely, as the empirical distribution of $\{(\mathbf{x}_i^{s+1}, \mathbf{y}_i)\}_{i\in C_a^N}$ converges weakly to the distribution of (Z_a^{s+1}, Y_a). This finishes the proof of Eq. (62).

Eq. (63) follows from Eq. (62) exactly by the same argument as in $\mathcal{B}_0(d)$.

A Reference probability results

In this appendix, we summarize a few probability facts that are repeatedly used in the proof of Lemma 2. We start by the following strong law of large numbers (SLLN) for triangular arrays of independent but not identically distributed random variables. The form stated below follows immediately from [HT97, Theorem 2.1].

Theorem 2 (SLLN, [HT97])**.** *Let $\{X_{n,i} : 1 \leq i \leq n, n \geq 1\}$ be a triangular array of random variables with $(X_{n,1}, \ldots, X_{n,n})$ mutually independent with mean equal to zero for each n and $n^{-1}\sum_{i=1}^{n} \mathbb{E}|X_{n,i}|^{2+\kappa} \leq cn^{\kappa/2}$ for some $0 < \kappa < 1$, $c < \infty$. Then $\frac{1}{n}\sum_{i=1}^{n} X_{i,n} \to 0$ almost surely for $n \to \infty$.*

Next, we present a standard property of Gaussian matrices without proof. This is a generalization of [BM11, Lemma 2].

Lemma 4. *For any deterministic $u \in \mathcal{V}_{q,N}$, $v \in \mathcal{V}_{q,n}$ and a gaussian matrix $\tilde{A} \in \mathbb{R}^{n\times N}$ with i.i.d. entries $\mathsf{N}(0, 1/N)$, we have*

(a) $[\tilde{A}u]_i \overset{\mathrm{d}}{=} \langle u, u\rangle^{\frac{1}{2}}\mathbf{z}$, *where* $\mathbf{z} \sim \mathsf{N}(0, \mathrm{I}_{q\times q})$.

(b) $\langle \tilde{A}u, v\rangle \overset{\mathrm{d}}{=} \langle u, u\rangle^{\frac{1}{2}}\langle v, z\rangle$, *where* $z \in \mathcal{V}_{q,n}$, $\mathbf{z}_i \sim \mathsf{N}(0, \mathrm{I}_{q\times q})$.

(c) $\lim_{n\to\infty}\langle \tilde{A}u, \tilde{A}u\rangle = \langle u, u\rangle$ *almost surely.*

(d) Consider, for $d \leq n$, a d-dimensional subspace W of $\mathbb{R}^n$, an orthogonal basis $w_1, \ldots, w_d$ of W with $\|w_i\|^2 = n$ for $i = 1, \ldots, d$, and the orthogonal projection P_W onto W. Then for $D = [w_1|\ldots|w_d]$, and $u \in \mathcal{V}_{q,N}$ with $\langle u, u\rangle = \mathrm{I}_{q\times q}$, we have $P_W\tilde{A}u \overset{\mathrm{d}}{=} Dx$ where $x \in \mathcal{V}_{q,d}$ satisfies: $\lim_{n\to\infty}\|x\| \overset{\text{a.s.}}{=} 0$. (the limit being taken with d fixed).

Lemma 5 (Stein's Lemma [Ste72])**.** *For jointly gaussian random vectors $Z_1, Z_2 \in \mathbb{R}^q$ with zero mean, and any function $\varphi : \mathbb{R}^q \to \mathbb{R}^{\tilde{q}}$ where $\mathbb{E}\{\frac{\partial\varphi}{\partial\mathbf{z}}(Z_1)\}$ and $\mathbb{E}\{Z_1[\varphi(Z_2)]^{\mathsf{T}}\}$ exist, the following holds*

$$\mathbb{E}\{Z_1[\varphi(Z_2)]^{\mathsf{T}}\} = \mathrm{Cov}(Z_1, Z_2)\,\mathbb{E}\{[\frac{\partial\varphi}{\partial\mathbf{z}}(Z_2)]^{\mathsf{T}}\}\,.$$

The following law of large numbers is a generalization of [BM11, Lemma 4] and can be proved in a very similar manner.

Lemma 6. *Let $k \geq 2$ and consider a sequence of vectors $\{v(N)\}_{N\geq 0}$ in $\mathcal{V}_{q,N}$, whose empirical distribution, denoted by $\hat{p}_{v(N)}$, converges weakly to a probability measure p_V on $\mathbb{R}^q$, such that $\mathbb{E}_{p_V}(\|V\|^k) < \infty$. Further assume $\mathbb{E}_{\hat{p}_{v(N)}}(\|V\|^k) \to \mathbb{E}_{p_V}(\|V\|^k)$ as $N \to \infty$. Then, for any pseudo-Lipschitz function $\psi : \mathbb{R}^q \to \mathbb{R}$ of order k:*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\psi(\mathbf{v}_i) \overset{\text{a.s.}}{=} \mathbb{E}\big[\psi(V)\big]\,. \tag{77}$$

References

[BBEKY12] D. Bean, P. Bickel, N. El Karoui, and B. Yu, *Optimal objective function in high-dimensional regression*, Submitted to PNAS (2012).

[BLM12] M. Bayati, M. Lelarge, and A. Montanari, *Universality in Polytope Phase Transitions and Message Passing Algorithms*, arXiv:1207.7321v1, 2012.

[BM11] M. Bayati and A. Montanari, *The dynamics of message passing on dense graphs, with applications to compressed sensing*, IEEE Trans. on Inform. Theory **57** (2011), 764–785.

[BM12] ———, *The LASSO risk for gaussian matrices*, IEEE Trans. on Inform. Theory **58** (2012), 1997–2017.

[Bol12] E. Bolthausen, *An iterative construction of solutions of the tap equations for the sherrington-kirkpatrick model*, arXiv:1201.2891, 2012.

[DJM11a] D. Donoho, I. Johnstone, and A. Montanari, *Accurate Prediction of Phase Transitions in Compressed Sensing via a Connection to Minimax Denoising*, arXiv:1111.1041, 2011.

[DJM11b] D. L. Donoho, A. Javanmard, and A. Montanari, *Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing*, arXiv:1112.0708v1, 2011.

[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.

[DMM11] D.L. Donoho, A. Maleki, and A. Montanari, *The Noise Sensitivity Phase Transition in Compressed Sensing*, IEEE Trans. on Inform. Theory **57** (2011), 6920–6941.

[FRVB11] A.K. Fletcher, S. Rangan, L.R. Varshney, and A. Bhargava, *Neural reconstruction with approximate message passing (neuramp)*, Proc. 25th Ann. Conf. Neural Information Processing Systems, NIPS, 2011.

[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.

[JM12] A. Javanmard and A. Montanari, *Subsampling at information theoretically optimal rates*, IEEE Intl. Symp. on Inform. Theory (Cambridge), July 2012.

[KBAU12] U.S. Kamilov, A. Bourquard, A. Amini, and M. Unser, *One-bit measurements with adaptive thresholds*, Signal Processing Letters, IEEE **19** (2012), no. 10, 607–610.

[KGR11] U.S. Kamilov, V.K. Goyal, and S. Rangan, *Message-Passing Estimation from Quantized Samples*, arXiv:1105.6368, 2011.

[KMS+12a] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, *Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices*, Journal of Statistical Mechanics: Theory and Experiment **2012** (2012), no. 08, P08009.

[KMS+12b] F. Krzakala, M. Mézard, F. Sausset, YF Sun, and L. Zdeborová, *Statistical-physics-based reconstruction in compressed sensing*, Physical Review X **2** (2012), no. 2, 021005.

[KP10] S. Kudekar and H.D. Pfister, *The effect of spatial coupling on compressive sensing*, 48th Annual Allerton Conference, 2010, pp. 347 –353.

[MM09] M. Mézard and A. Montanari, *Information, Physics and Computation*, Oxford University Press, Oxford, 2009.

[MN89] P. McCullagh and J. A. Nelder, *Generalized linear models (Second edition)*, London: Chapman & Hall, 1989.

[MPV87] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond*, World Scientific, Singapore, 1987.

[NW72] J. A. Nelder and R. W. M. Wedderburn, *Generalized linear models*, Journal of the Royal Statistical Society, Series A, General **135** (1972), 370–384.

[Pea88] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann, San Francisco, 1988.

[Ran11] S. Rangan, *Generalized Approximate Message Passing for Estimation with Random Linear Mixing*, IEEE Intl. Symp. on Inform. Theory (St. Petersbourg), August 2011, pp. 2168 – 2172.

[Rén59] A. Rényi, *On the dimension and entropy of probability distributions*, Acta Mathematica Hungarica **10** (1959), 193–215.

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

[Sch10] P. Schniter, *Turbo Reconstruction of Structured Sparse Signals*, Proceedings of the Conference on Information Sciences and Systems (Princeton), 2010.

[SS12] S. Som and P. Schniter, *Compressive imaging using approximate message passing and a markov-tree prior*, Signal Processing, IEEE Transactions on **60** (2012), no. 7, 3439–3448.

[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.

[WV10] Y. Wu and S. Verdú, *Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression*, IEEE Trans. on Inform. Theory **56** (2010), 3721–3748.