Universality in Polytope Phase Transitions and Message Passing Algorithms

Mohsen Bayati,[*] Marc Lelarge[†] and Andrea Montanari [‡]

July 31, 2012

Abstract

We consider a class of nonlinear mappings $\mathsf{F}_{A,N}$ in $\mathbb{R}^N$ indexed by symmetric random matrices $A \in \mathbb{R}^{N \times N}$ with independent entries. Within spin glass theory, special cases of these mappings correspond to iterating the TAP equations and were studied by Erwin Bolthausen. Within information theory, they are known as 'approximate message passing' algorithms.

We study the high-dimensional (large N) behavior of the iterates of F for polynomial functions F, and prove that it is universal, i.e. it depends only on the first two moments of the entries of A, under a subgaussian tail condition. As an application, we prove the universality of a certain phase transition arising in polytope geometry and compressed sensing. This solves –for a broad class of random projections– a conjecture by David Donoho and Jared Tanner.

1   Introduction and main results

Let $A \in \mathbb{R}^{N \times N}$ be a random Wigner matrix, i.e. a random matrix with i.i.d. entries A_{ij} satisfying $\mathbb{E}\{A_{ij}\} = 0$ and $\mathbb{E}\{A_{ij}^2\} = 1/N$. Considerable effort has been devoted to studying the distribution of the eigenvalues of such a matrix [AGZ09, BS05, TV12]. The *universality phenomenon* is a striking recurring theme in these studies. Roughly speaking, many asymptotic properties of the joint eigenvalues distribution are independent of the entries distribution as long as the latter has the prescribed first two moments, and satisfies certain tail conditions. We refer to [AGZ09, BS05, TV12] and references therein for a selection of such results. Universality is extremely useful because it allows to compute asymptotics for one entries distribution (typically, for Gaussian entries) and then export the results to a broad class of distributions.

In this paper we are concerned with random matrix universality, albeit we do not focus on eigenvalues properties. Given $A \in \mathbb{R}^{N \times N}$, and an initial condition $x^0 \in \mathbb{R}^N$ independent of A, we consider the sequence $(x^t)_{t \geq 0}$ $t \in \mathbb{N}$ defined by letting, for $t \geq 0$,

$$x^{t+1} = A \, f(x^t; t) - \mathsf{b}_t \, f(x^{t-1}; t-1), \qquad \mathsf{b}_t \equiv \frac{1}{N} \mathrm{div}(f(x; t))\big|_{x=x^t}. \tag{1.1}$$

[*]Graduate School of Business, Stanford University

[†]INRIA and ENS, Paris

[‡]Department of Electrical Engineering and Department of Statistics, Stanford University

Here, div denotes the divergence operator and, for each $t \geq 0$, $f(\,\cdot\,;t) : \mathbb{R}^N \to \mathbb{R}^N$ is a separable function, i.e. $f(z;t) = (f_1(z_1;t), \ldots, f_2(z_N;t))$ where the functions $f_i(\,\cdot\,;t) : \mathbb{R} \to \mathbb{R}$ are polynomials of bounded degree. In particular $\mathsf{b}_t = N^{-1} \sum_{i=1}^N f_i'(x_i^t;t)$.

The present paper is concerned with the asymptotic distribution of x^t as $N \to \infty$ with t fixed, and establishes the following results:

Universality. As $N \to \infty$, the finite-dimensional marginals of the distribution of x^t are asymptotically insensitive to the distribution of the entries of A_{ij}.

State evolution. The entries of x^t are asymptotically Gaussian with zero mean, and variance that can be explicitly computed through a one-dimensional recursion, that we will refer to as *state evolution*

Phase transitions in polytope geometry. As an application, we use state evolution to prove universality of a phase transition on polytope geometry, with connections to compressed sensing. This solves –for a broad class of random matrices with independent entries– a conjecture put forward by David Donoho and Jared Tanner in [Don05a, DT11].

In order to illustrate the usefulness of the first two technical results, we start the presentation of our results from the third one.

1.1   Universality of polytope neighborliness

A polytope Q is said to be *centrosymmetric* if $x \in Q$ implies $-x \in Q$. Following [Don05b, Don05a] we say that such a polytope is *k-neighborly* if the condition below holds:

(I) Every subset of k vertices of Q which does not contain an antipodal pair, spans a $(k-1)$ dimensional face.

The *neighborliness* of Q is the largest value of k for which this condition holds. The prototype of neighborly polytope is the ℓ_1 ball $C^n \equiv \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$, whose neighborliness is indeed equal to n.

It was shown in a series of papers [Don05b, Don05a, DT05b, DT05a, DT09] that polytope neighborliness has tight connections with the geometric properties of random point clouds, and with sparsity-seeking methods to solve underdetermined systems of linear equations. The latter are in turn central in a number of applied domains, including model selection for data analysis and compressed sensing. For the reader's convenience, these connections will be briefly reviewed in Section 5.

Intuitive images of low-dimensional polytopes suggest that 'typical' polytopes are not neighborly: already selecting $k = 2$ vertices, does lead to a segment that connects them and passes through the interior of Q. This conclusion is spectacularly wrong in high dimension. Natural random constructions lead to polytopes whose neighborliness scales *linearly* in the dimension. Motivated by the above applications, and following [Don05b, Don05a, DT05b, DT05a], we focus here on a weaker notion of neighborliness. Roughly speaking, this corresponds to the largest k such that *most* subsets of k vertices of Q span a $(k-1)$-dimensional face. In order to formalize this notion, we denote by $\mathfrak{F}(Q;\ell)$ the number of $\lfloor \ell \rfloor$-dimensional faces of Q.

Definition 1. *Let $\mathcal{Q} = \{Q^n\}_{n \geq 0}$ be a sequence of centrosymmetric polytopes indexed by n where Q_n has $2n$ vertices and has dimension $m = m(n)$: $Q^n \subseteq \mathbb{R}^m$. We say that $\mathcal{Q}$ has* weak neighborliness *$\rho \in (0,1)$ if for any $\xi > 0$,*

$$\lim_{n \to \infty} \frac{\mathfrak{F}(Q^n; m(n)\rho(1-\xi))}{\mathfrak{F}(C^n; m(n)\rho(1-\xi))} = 1,$$

$$\lim_{n \to \infty} \frac{\mathfrak{F}(Q^n; m(n)\rho(1+\xi))}{\mathfrak{F}(C^n; m(n)\rho(1+\xi))} = 0.$$

If the sequence $\mathcal{Q}$ is random, we say that $\mathcal{Q}$ has weak neighborliness *ρ (in probability) if the above limits hold* in probability.

In other words, a sequence of polytopes $\{Q^n\}_{n \geq 0}$ has weak neighborliness ρ, if for large n the m dimensional polytope Q^n has close to the maximum possible number of k faces, for all $k < m\rho(1-\xi)$.

Note 1. *Note that previously the neighborliness of a polytope was defined to be the largest integer k satisfying condition (I). However, in our definition weak neighborliness refers to the fraction k/n. This is due to the fact that weak neighborliness is defined in the limit $n \to \infty$.*

The existence of weakly neighborly polytope sequences is clear when $m(n) = n$ since in this case we can take $Q^n = C^n$ with $\rho = 1$, but the existence is highly non-trivial when m is only a fraction of n.

It comes indeed as a surprise that this is a generic situation as demonstrated by the following construction. For a matrix $A \in \mathbb{R}^{m \times n}$, and $S \subseteq \mathbb{R}^n$, let $AS \equiv \{Ax \in \mathbb{R}^m : x \in S\}$. In particular, AC^n is the centrosymmetric m-dimensional polytope obtained by projecting the n-dimensional ℓ_1 ball to m dimensions. The following result was proved in [Don05a].

Theorem 1 (Donoho, 2005). *There exists a function $\rho_* : (0,1) \to (0,1)$ such that the following holds. Fix $\delta \in (0,1)$. For each $n \in \mathbb{N}$, let $m(n) = \lfloor n\delta \rfloor$ and define $A(n) \in \mathbb{R}^{m(n) \times n}$ to be a random matrix with i.i.d. Gaussian entries.*

Then, the sequence of polytopes $\{A(n)C^n\}_{n \geq 0}$ has weak neighborliness $\rho_*(\delta)$ in probability.

A characterization of the curve $\delta \mapsto \rho_*(\delta)$ was provided in [Don05a], but we omit it here since a more explicit expression will be given below.

The proof of Theorem 1 is based on exact expressions for the number of faces $\mathfrak{F}(A(n)C^n; \ell)$. These are in turn derived from earlier works in polytope geometry by Affentranger and Schneider [AS92] and by Vershik and Sporyshev [VS92]. This approach relies in a fundamental way on the invariance of the distribution of $A(n)$ under rotations.

Motivated by applications to data analysis and signal processing, Donoho and Tanner [DT11] carried out extensive numerical simulations for random polytopes of the form $A(n)C^n$ for several choices of the distribution of $A(n)$. They formulated a *universality hypothesis* according to which the conclusion of Theorem 1 holds for a far broader class of random matrices. The results of their numerical simulations were consistent with this hypothesis.

Here we establish the first rigorous result indicating universality of polytope neighborliness for a broad class of random matrices. Define the curve $(\delta, \rho_*(\delta))$, $\delta \in (0,1)$, parametrically by letting, for

3

$\alpha \in (0, \infty)$:

$$\delta = \frac{2\phi(\alpha)}{\alpha + 2(\phi(\alpha) - \alpha\Phi(-\alpha))}, \tag{1.2}$$

$$\rho = 1 - \frac{\alpha\Phi(-\alpha)}{\phi(\alpha)}, \tag{1.3}$$

where $\phi(z) = e^{-z^2/2}/\sqrt{2\pi}$ is the Gaussian density and $\Phi(x) \equiv \int_{-\infty}^{x} \phi(z)\,\mathrm{d}z$ is the Gaussian distribution. Explicitly, if the above functions on the right-hand side of Eqs. (1.2), (1.3) are denoted by $f_\delta(\alpha)$, $f_\rho(\alpha)$, then[1] $\rho_*(\delta) \equiv f_\rho(f_\delta^{-1}(\delta))$.

Here we extend the scope of Theorem 1 from Gaussian matrices to matrices with independent subgaussian[2] entries (not necessarily identically distributed).

Theorem 2. *Fix $\delta \in (0, 1)$. For each $n \in \mathbb{N}$, let $m(n) = \lfloor n\delta \rfloor$ and define $A(n) \in \mathbb{R}^{m(n) \times n}$ to be an random matrix with independent subgaussian entries, with zero mean, unit variance, and common scale factor s independent of n. Further assume $A_{ij}(n) = \tilde{A}_{ij}(n) + \nu_0\,G_{ij}(n)$ where $\nu_0 > 0$ is independent of n and $\{G_{ij}(n)\}_{i \in [m], j \in [n]}$ is a collection of i.i.d. $\mathsf{N}(0,1)$ random variables independent of $\tilde{A}(n)$.*

Then the sequence of polytopes $\{A(n)C^n\}_{n \geq 0}$ has weak neighborliness $\rho_*(\delta)$ in probability.

It is likely that this theorem can be improved in two directions. First, a milder tail condition than subgaussianity is probably sufficient. Second, we are assuming that the distribution of A_{ij} has an arbitrarily small Gaussian component. This is not necessary for the upper bound on neighborliness, and appears to be an artifact of the proof of the lower bound.

The proof of Theorem is provided in Section 5. By comparison, the most closely related result towards universality is by Adamczak, Litvak, Pajor, and Tomczak-Jaegermann [ALPTJ11]. For a class of matrices $A(n)$ with i.i.d. columns, these authors prove that $A(n)C^n$ has neighborliness scaling linearly with n. This however does not suggest that a limit weak neighborliness exists, and is universal, as established instead in Theorem 2.

At the other extreme, universality of compressed sensing phase transitions can be conjectured from the results of the non-rigorous replica method [KWT09, RFG09].

1.2 Universality of iterative algorithms

We will consider here and below a setting that is somewhat more general than the one described by Eq. (1.1). Following the terminology of [DMM09], we will refer to such an iteration as to the approximate message passing (AMP) iteration/algorithm.

We generalize the iteration (1.1) to take place in 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 $x = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, with $\mathbf{x}_i \in \mathbb{R}^q$. Components of $\mathbf{x}_i \in \mathbb{R}^q$ will be indicated as $(x_i(1), \ldots, x_i(q)) \equiv \mathbf{x}_i$.

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 $v', v \in \mathcal{V}_{q,N}$ letting $v' = Av$ be given by $\mathbf{v}_i' = \sum_{j=1}^{N} A_{ij}\mathbf{v}_j$ for all $i \in [N]$. Here and below $[N] \equiv \{1, \ldots, N\}$ is the set of first N integers. In other words we identify A with the Kronecker product $A \otimes \mathrm{I}_{q \times q}$.

---

[1]It is easy to show that $f_\delta(\alpha)$ is strictly decreasing in $\alpha \in [0, \infty)$, with $f_\delta(0) = 1$, $\lim_{\alpha \to \infty} f_\delta(\alpha) = 0$, and hence f_δ^{-1} is well defined on $[0, 1]$. Further properties of this curve can be found in [DMM09, DMM11].

[2]See Eq. (1.7) for the definition of subgaussian random variables.

Definition 2. *An* AMP instance *is a triple* $(A, \mathcal{F}, x^0)$ *where:*

1. $A \in \mathbb{R}^{N \times N}$ *is a symmetric matrix with* $A_{i,i} = 0$ *for all* $i \in [N]$.

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;*

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

$$x^{t+1} = A\, f(x^t; t) - \mathsf{B}_t\, f(x^{t-1}; t-1). \tag{1.4}$$

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

$$\mathbf{v}'_i = \left(\sum_{j \in [N]} A_{ij}^2 \frac{\partial f^j}{\partial \mathbf{x}}(\mathbf{x}_j^t, t) \right) \mathbf{v}_i\,, \tag{1.5}$$

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

The above definition can also be summarized by the following expression for the evolution of a single coordinate under AMP

$$\mathbf{x}_i^{t+1} = \sum_{j \in [N]} A_{ij} f^j(\mathbf{x}_j^t, t) - \sum_{j \in [N]} A_{ij}^2 \frac{\partial f^j}{\partial \mathbf{x}}(\mathbf{x}_j^t, t) f^i(\mathbf{x}_i^{t-1}, t-1). \tag{1.6}$$

Notice that Eq. (1.1) corresponds to the special case $q = 1$, in which we replaced A_{ij}^2 by $\mathbb{E}\{A_{ij}^2\} = 1/N$ for simplicity of exposition.

Recall that a centered random variable X is subgaussian with scale factor σ^2 if, for all $\lambda > 0$, we have

$$\mathbb{E}\left(e^{\lambda X}\right) \leq e^{\frac{\sigma^2 \lambda^2}{2}}. \tag{1.7}$$

Definition 4. *Let* $\{(A(N), \mathcal{F}_N, x^{0,N})\}_{N \geq 1}$ *be a sequence of AMP instances indexed by the dimension* N, *with* $A(N)$ *a random matrix and* $x^{0,N}$ *a random vector. We say that the sequence is* (C, d)-regular *(or, for short,* regular*) polynomial sequence if*

1. *For each* N, *the entries* $(A_{ij}(N))_{1 \leq i < j \leq N}$ *are independent centered random variables. Further they are subgaussian with common scale factor* C/N.

2. *For each* N, *the functions* $f^i(\cdot; t)$ *in* $\mathcal{F}_N$ *(possibly random, as long as they are independent from* $A(N)$, $x^{0,N}$*) are polynomials with maximum degree* d *and coefficients bounded by* C.

3. *For each* N, $A(N)$ *and* $x^{0,N}$ *are independent. Further, we have* $\sum_{i=1}^{N} \exp\{\|\mathbf{x}_i^{0,N}\|_2^2 / C\} \leq NC$ *with probability converging to one as* $N \to \infty$.

We state now our universality result for the algorithm (1.4).

Theorem 3. *Let $(A(N), \mathcal{F}_N, x^{0,N})_{N \geq 1}$ and $(\tilde{A}(N), \mathcal{F}_N, x^{0,N})_{N \geq 1}$ be any two (C, d)-regular polynomial sequences of instances, that differ only in the distribution of the random matrices $A(N)$ and $\tilde{A}(N)$.*

Denote by $\{x^t\}_{t \geq 0}$, $\{\tilde{x}^t\}_{t \geq 0}$ the corresponding AMP orbits. Assume further that for all N and all $i < j$, $\mathbb{E}\{A_{ij}^2\} = \mathbb{E}\{\tilde{A}_{ij}^2\}$. Then, for any set of polynomials $\{p_{N,i}\}_{N \geq 0, 1 \leq i \leq N}$ $p_{N,i} : \mathbb{R}^q \to \mathbb{R}$, with degree bounded by D and coefficients bounded by B for all N and $i \in [N]$, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \left\{ \mathbb{E}p_{N,i}(\mathbf{x}_i^t) - \mathbb{E}p_{N,i}(\tilde{\mathbf{x}}_i^t) \right\} = 0 \,. \tag{1.8}$$

1.3 State evolution

Theorem 3 establishes that the behavior of the sequence $\{x^t\}_{t \geq 0}$ is, in the high dimensional limit, insensitive to the distribution of the entries of the random matrix A. In order to characterize this limit, we need to make some assumption on the collection of functions $\mathcal{F}_N$.

Definition 5. *We say that the sequence of AMP instances $\{(A(N), \mathcal{F}_N, x^{0,N})\}_{N \geq 0}$ is* polynomial and converging *(or simply* converging*) if it is (C, d)-regular and there exists: (i) An integer k; (ii) A symmetric matrix $W \in \mathbb{R}^{k \times k}$ with non-negative entries; (iii) A function $g : \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \times [k] \times \mathbb{N} \to \mathbb{R}^q$, with $g(\mathbf{x}, Y, a, t) = (g_1(\mathbf{x}, Y, a, t), \ldots, g_q(\mathbf{x}, Y, a, t))$ and, for each $r \in [q]$, $a \in [k]$, $t \in \mathbb{N}$, $g_r(\cdot, a, t)$ a polynomial with degree d and coefficients bounded by C; (iv) k probability measures $P_1, \ldots, P_k$ on $\mathbb{R}^{\tilde{q}}$, with P_a a finite mixture of (possibly degenerate) Gaussians for each $a \in [k]$; (v) For each N, a finite partition $C_1^N \cup C_2^N \cup \cdots \cup C_k^N = [N]$; (vi) k positive semidefinite matrices $\widehat{\Sigma}_1^0, \ldots \widehat{\Sigma}_k^0 \in \mathbb{R}^{q \times q}$, such that the following happens.*

1. *For each $a \in [k]$, we have $\lim_{N \to \infty} |C_a^N|/N = c_a \in (0, 1)$.*

2. *For each $N \geq 0$, each $a \in [k]$ and each $i \in C_a^N$, we have $f^i(\mathbf{x}, t) = g(\mathbf{x}, Y(i), a, t)$ where $Y(1), \ldots, Y(N)$ are independent random variables with $Y(i) \sim P_a$ whenever $i \in C_a^N$ for some $a \in [k]$.*

3. *For each N, the entries $\{A_{ij}(N)\}_{1 \leq i < j \leq N}$ are independent subgaussian random variables with scale factor C/N, $\mathbb{E}A_{ij} = 0$, and, for $i \in C_a^N$ and $j \in C_b^N$, $\mathbb{E}\{A_{ij}^2\} = W_{ab}/N$.*

4. *For each $a \in [k]$, in probability,*

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} g\left(\mathbf{x}_i^0, Y(i), a, 0\right) g\left(\mathbf{x}_i^0, Y(i), a, 0\right)^\mathsf{T} = \widehat{\Sigma}_a^0 \,. \tag{1.9}$$

With a slight abuse of notation, we will sometime denote a converging sequence by $\{(A(N), g, x^{0,N})\}_{N \geq 0}$. We use capital letters to denote the $Y(i)$'s to emphasize that they are random and do not change across iterations.

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 set of k positive semidefinite matrices $\Sigma^t = (\Sigma_1^t, \Sigma_2^t, \ldots, \Sigma_k^t)$, with $\Sigma_a^t \in \mathbb{R}^{q \times q}$. These are obtained by letting, for each $t \geq 1$

$$\Sigma_a^t = \sum_{b=1}^k c_b \, W_{ab} \, \widehat{\Sigma}_b^{t-1} \tag{1.10}$$

$$\widehat{\Sigma}_a^t = \mathbb{E}\left\{ g(Z_a^t, Y_a, a, t) g(Z_a^t, Y_a, a, t)^{\mathsf{T}} \right\}, \tag{1.11}$$

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

Theorem 4. *Let $(A(N), \mathcal{F}_N, x^0)_{N \geq 0}$ be a polynomial and converging sequence of AMP instances, and denote by $\{x^t\}_{t \geq 0}$ the corresponding AMP sequence. Then for each $t \geq 1$, each $a \in [k]$, and each locally Lipschitz function $\psi : \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \to \mathbb{R}$ such that $|\psi(\mathbf{x}, y)| \leq K(1 + \|y\|_2^2 + \|\mathbf{x}\|_2^2)^K$, we have, in probability,*

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{j \in C_a^N} \psi(\mathbf{x}_j^t, Y(i)) = \mathbb{E}\{\psi(Z_a, Y_a)\}, \tag{1.12}$$

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

We conclude by mentioning that, following [DMM09], generalizations of the algorithm (1.4) were studied by several groups [Sch10, Ran11, MAYB11], for a number of applications. Universality results analogous to the one proved here are expected to hold for such generalizations as well.

1.4 Outline of the paper

The paper is organized as follows. After some preliminary facts and notations in Section 2, Section 3 considers the AMP iteration (1.4) and proves Theorems 3 and 4. In order to achieve our goal, we introduce two different iterations whose analysis provides useful intermediate steps. We also prove a generalization of Theorem 4 to estimate functions of messages at two distinct times $\psi(\mathbf{x}_i^t, \mathbf{x}_i^s, Y(i))$.

Section 4 proves a generalization of Theorem 4 to the case of rectangular (non-symmetric) matrices A. This is achieved by effectively embedding the rectangular matrix, into a larger symmetric matrix and applying our results for symmetric matrices.

The generalization to rectangular matrices is finally used in Section 5 to prove our result on the universality of polytope neighborliness, Theorem 2. This is done via a correspondence with compressed sensing reconstruction established in [Don05a], and a sharp analysis of an AMP iteration that solves this reconstruction problem.

2 Notations and basic simplifications

We will always view vectors as column vectors. The transpose of vector v is the row vector indicated by v^{T}. Analogously, the transpose of a matrix (or vector) M is denoted by M^{T}. For a vector $v \in \mathbb{R}^m$, we denote its ℓ_p norm, $p \geq 1$ by $\|v\|_p \equiv (\sum_{i=1}^m |v_i|^p)^{1/p}$. This is extended in the usual way to $p = \infty$. We will often omit the subscript if $p = 2$. For a matrix M, we denote by $\|M\|_p$ the corresponding ℓ_p operator norm. The standard scalar product of $u, v \in \mathbb{R}^m$ is denoted by $\langle u, v \rangle = \sum_{i=1}^m u_i v_i$. Given $v \in \mathbb{R}^m$, $w \in \mathbb{R}^n$, we denote by $[v, w] \in \mathbb{R}^{m+n}$ the (column) vector obtained by concatenating v

and w. The identity matrix is denoted by I, or $\mathrm{I}_{m \times m}$ if the dimensions need to be specified. The indicator function is $\mathbf{1}(\,\cdot\,)$. The set of first m integers is indicated by $[m] = \{1, \ldots, m\}$. Finally, given $\mathbf{x} = (x(1), x(2), \ldots, x(q)) \in \mathbb{R}^q$ and $\mathbf{m} = (m(1), \ldots, m(q)) \in \mathbb{N}^q$, we write

$$\mathbf{x^m} \equiv \prod_{r=1}^{q} x(r)^{m(r)}. \tag{2.1}$$

Following the common practice, degenerate Gaussian distributions will be considered Gaussian, without further qualification. In particular, any distribution with finite support in $\mathbb{R}^k$ is a finite mixture of Gaussians.

In our proof of Theorem 4 we will make use of the following simplification, that lightens somewhat the notation.

Remark 1. *For proving Theorem 4, it is sufficient to consider the case in which $g : (\mathbf{x}, Y, a, t) \mapsto g(\mathbf{x}, Y, a, t)$ is independent of Y.*

Proof. We can assume without loss of generality that the measures P_a are Gaussian. Indeed if, for instance, P_a is a mixture of ℓ gaussians, $P_a = w_1 P_{a,1} + w_2 P_{a,2} + \cdots + w_\ell P_{a,\ell}$ then we can replace effectively the partition element C_a^N by a finer partition $C_{a,1}^N, \ldots, C_{a,\ell}^N$ whereby $C_{a,1}^N \cup \cdots \cup C_{a,\ell}^N = C_a^N$ and $|C_{a,1}^N|, \ldots, |C_{a,\ell}^N|$ are multinomial with parameters $(w_1, \ldots, w_\ell)$. Notice that this finer partition is random, but $|C_{a,i}^N|/N \to c_a w_i$ almost surely, and therefore the theorem applies.

Assume therefore that the P_a are gaussian. By replacing $g(\mathbf{x}, Y, a, t)$ by $g'(\mathbf{x}, Y, a, t) = g(\mathbf{x}, Q_a Y + v_a, a, t)$ for suitable matrices Q_a, and vectors v_a, we can always assume $Y_a \sim \mathsf{N}(0, \mathrm{I}_{\tilde{q} \times \tilde{q}})$ for all a. Assume therefore $Y_a \sim \mathsf{N}(0, \mathrm{I}_{\tilde{q} \times \tilde{q}})$. Enlarge the space by letting $k' = k + \tilde{q}$, $N' = (\tilde{q} + 1)N$ and $C_a^{N'} = \{N\ell + 1, \ldots, N(\ell + 1)\}$, for $a = k + \ell > k$, while $C_a^{N'} = C_a^N$ for $a \le k$. We further let $q' = q + \tilde{q}$ and define new functions $g' : \mathbb{R}^{q'} \times \mathbb{R}^{\tilde{q}} \times [k'] \times \mathbb{N} \to \mathbb{R}^{q'}$ independent of the second argument (Y) as follows. For $\mathbf{x} \in \mathbb{R}^q$, $\tilde{\mathbf{x}} \in \mathbb{R}^{\tilde{q}}$, we let

$$
\begin{aligned}
g'_r\Big((\mathbf{x}, \tilde{\mathbf{x}}), Y, a, t\Big) &= g_r(\mathbf{x}, \tilde{\mathbf{x}}, a, t) && \text{for } r \in \{1, \ldots, q\}, a \in \{1, \ldots, k\}, \\
g'_r\Big((\mathbf{x}, \tilde{\mathbf{x}}), Y, a, t\Big) &= 0 && \text{for } r \in \{q+1, \ldots, q+\tilde{q}\}, a \in \{1, \ldots, k\}, \\
g'_r\Big((\mathbf{x}, \tilde{\mathbf{x}}), Y, a, t\Big) &= 0 && \text{for } r \in \{1, \ldots, q\}, a \in \{k+1, \ldots, k+\tilde{q}\}, \\
g'_{q+\ell}\Big((\mathbf{x}, \tilde{\mathbf{x}}), Y, k+\ell', t\Big) &= \mathbf{1}(\ell = \ell') && \text{for } \ell, \ell' \in \{1, \ldots, \tilde{q}\}.
\end{aligned}
$$

We further use matrix A' constructed as follows: $A'_{ij} = A_{ij}$ for $i, j \le N$, and $A_{ij} \sim \mathsf{N}(0, 1/N)$ if $i > N$ or $j > N$. (Notice that $\mathbb{E}\{(A'_{ij})^2\} = 2/N'$ but this amounts just to an overall rescaling and is of course immaterial.) Clearly the functions g' do not depend on Y as claimed. Further, $\tilde{\mathbf{x}} \sim \mathsf{N}(0, \mathrm{I}_{\tilde{q} \times \tilde{q}})$ at all iterations. Hence the new iteration is identical to the original one when restricted on $\{x_i(r) : i \le N, r \le q\}$. $\square$

3 Proofs of Theorems 3 and 4

In this section we consider the AMP iteration (1.4), and prove Theorem 3 and Theorem 4, and indeed generalize the latter.

We extend the state evolution (1.10) by defining for each $t \geq s \geq 0$ and for all $a \in [k]$, a positive semidefinite matrix $\Sigma_a^{t,s} \in \mathbb{R}^{(2q) \times (2q)}$ as follows. For boundary conditions, we set

$$\widehat{\Sigma}_a^{0,0} = \begin{pmatrix} \widehat{\Sigma}_a^0 & \widehat{\Sigma}_a^0 \\ \widehat{\Sigma}_a^0 & \widehat{\Sigma}_a^0 \end{pmatrix}, \quad \widehat{\Sigma}_a^{t,0} = \begin{pmatrix} \widehat{\Sigma}_a^t & 0 \\ 0 & \widehat{\Sigma}_a^0 \end{pmatrix}, \quad \widehat{\Sigma}_a^{0,t} = \begin{pmatrix} \widehat{\Sigma}_a^0 & 0 \\ 0 & \widehat{\Sigma}_a^t \end{pmatrix}, \tag{3.1}$$

with $\widehat{\Sigma}_a^t$ defined per Eq. (1.10). For any $s, t \geq 1$, we set recursively

$$\Sigma_a^{t,s} = \sum_{b=1}^{k} c_b W_{ab} \widehat{\Sigma}_b^{t-1,s-1}, \tag{3.2}$$

$$\widehat{\Sigma}_a^{t,s} = \mathbb{E}\left\{ X_a X_a^\mathsf{T} \right\}, \quad X_a \equiv [g(Z_a^t, Y_a, a, t), g(Z_a^s, Y_a, a, s)], \quad (Z_a^t, Z_a^s) \sim \mathsf{N}(0, \Sigma_a^{t,s}). \tag{3.3}$$

Recall that $[g(Z_a^t, Y_a, a, t), g(Z_a^s, Y_a, a, s)] \in \mathbb{R}^{2q}$ is the vector obtained by concatenating $g(Z_a^t, Y_a, a, t)$ and $g(Z_a^s, Y_a, a, s)$. Note that taking $s = t$ in (3.2), we recover the recursion for Σ_a^t given by Eq. (1.10). Namely, for all t we have

$$\Sigma_a^{t,t} = \begin{pmatrix} \Sigma_a^t & \Sigma_a^t \\ \Sigma_a^t & \Sigma_a^t \end{pmatrix}. \tag{3.4}$$

Theorem 5. *Let* $\{(A(N), \mathcal{F}_N, x^{0,N})\}_{N \geq 1}$ *be a polynomial and converging sequence of instances and denote by* $\{x^t\}_{t \geq 0}$ *the corresponding AMP orbit.*

Fix $s, t \geq 1$. *If* $s \neq t$, *further assume that the initial condition* $x^{0,N}$ *is obtained by letting* $\mathbf{x}_i^{0,N} \sim Q_a$ *independent and identically distributed, with* Q_a *a finite mixture of Gaussians for each* a. *Then, for each* $a \in [k]$, *and each locally Lipschitz function* $\psi : \mathbb{R}^q \times \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \to \mathbb{R}$ *such that* $|\psi(\mathbf{x}, \mathbf{x}', y)| \leq K(1 + \|y\|_2^2 + \|\mathbf{x}\|_2^2 + \|\mathbf{x}'\|_2^2)^K$, *we have, in probability,*

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{j \in C_a^N} \psi(\mathbf{x}_j^t, \mathbf{x}_j^s, Y(j)) = \mathbb{E}\left[\psi(Z_a^t, Z_a^s, Y_a)\right],$$

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

Throughout this section, we will assume that $\{(A(N), \mathcal{F}_N, x^{0,N})\}$, $\{(\tilde{A}(N), \mathcal{F}_N, x^{0,N})\}$, etc. are (C, d)-*regular polynomial* sequences of AMP instances. We will often omit explicit mention of this hypothesis. Notice that Theorem 3 holds *per realization* of the functions $\mathcal{F}_N$. Because of this, and of Remark 1, we will consider hereafter $\mathcal{F}_N$ to be non-random.

The rest of this section is organized as follows. In subsection 3.1 we introduce two new iterations that are useful intermediary steps for our analysis. We show that the corresponding variables admit representations as sums over trees in Sec. 3.2 and use them to prove basic properties of these recursions in Secs. 3.3, 3.4, and 3.5. Theorems 3 and 5 are then proved in Secs. 3.6, 3.7. Because of Eq. (3.4), Theorem 4 follows as a special case of Theorem 5. Indeed, we will show that both statements are equivalent through a reduction argument. Depending on the application, Theorem 5 might be a more convenient formulation of the state evolution and will be used in Section 4.

3.1 Message passing iteration

We define two new message passing sequences corresponding to the instance $(A, \mathcal{F}, x^{0,N})$. For each $i \in [N]$ we use the short notation $[N] \setminus i$ to denote the set $[N] \setminus \{i\}$. We now define the sequence of vectors $(\mathbf{z}_{i \to j}^t)_{t \in \mathbb{N}}$, where for each $i \neq j \in [N]$, $\mathbf{z}_{i \to j}^t$ is a vector in $\mathbb{R}^q$ or equivalently for each $t \in \mathbb{N}$, we can see $(\mathbf{z}_{i \to j}^t)$ as an $N \times N$ matrix with entries in $\mathbb{R}^q$ (diagonal elements are never used). The initial condition is denoted by $\mathbf{z}_{i \to j}^0 \in \mathbb{R}^q$ for any $i, j \in [N]$ and is independent of j, such that $\mathbf{z}_{i \to j}^0 = \mathbf{x}_i^{0,N}$ for all $j \neq i$. The r-th coordinate of the vector $\mathbf{z}_{i \to j}^{t+1}$ is defined by the following recursion for $t \geq 0$,

$$z_{i \to j}^{t+1}(r) \quad = \quad \sum_{\ell \in [N] \setminus j} A_{\ell i} \, f_r^\ell(\mathbf{z}_{\ell \to i}^t, t), \tag{3.5}$$

where $f_r^\ell(\cdot, t) : \mathbb{R}^q \to \mathbb{R}$ is the rth coordinate of $f^\ell(\cdot, t)$.

We also define for each $i \in [N]$ and $t \geq 0$, the vector $\mathbf{z}_i^{t+1} \in \mathbb{R}^q$ by

$$z_i^{t+1}(r) \quad = \quad \sum_{\ell \in [N]} A_{\ell i} \, f_r^\ell(\mathbf{z}_{\ell \to i}^t, t). \tag{3.6}$$

Our first result establishes universality of the moments of $\mathbf{z}_{i \to j}^t$ for polynomial sequences of instances.

Proposition 6. *Let $(A(N), \mathcal{F}_N, x^{0,N})_{N \geq 1}$ and $(\tilde{A}(N), \mathcal{F}_N, x^{0,N})_{N \geq 1}$ be any two (C, d)-regular polynomial sequences of AMP instances, that differ only in the distribution of the random matrices $A(N)$ and $\tilde{A}(N)$. Assume that for all N and all $i < j$, $\mathbb{E}\{A_{ij}^2\} = \mathbb{E}\{\tilde{A}_{ij}^2\}$. Denote by $\mathbf{z}_i^t$ the orbit (respectively $\tilde{\mathbf{z}}_i^t$) defined by (3.6) while iterating (3.5) with matrix A (respectively $\tilde{A}$). Then for any $t \geq 1$ and any $\mathbf{m} = (m(1), \ldots, m(q)) \in \mathbb{N}^q$, there exists K independent of N such that, for any $i \in [N]$:*

$$\left| \mathbb{E}\left[(\mathbf{z}_i^t)^{\mathbf{m}} \right] - \mathbb{E}\left[(\tilde{\mathbf{z}}_i^t)^{\mathbf{m}} \right] \right| \leq K N^{-1/2}. \tag{3.7}$$

The proof of this proposition is provided in Section 3.3.

Note 2. *In this statement and and in the rest of this section, K is always understood as a function of d, t, q, m, C which may vary from line to line but which is independent of N.*

Our second message passing sequence is defined as follows: for a (C, d)-regular sequence of instances $(A(N), \mathcal{F}_N, x^{0,N})_{N \geq 1}$, we define for each N, an i.i.d. sequence of $N \times N$ random matrices $\{A^t\}_{t \in \mathbb{N}}$ such that $A^0 = A(N)$. Then we define $(\mathbf{y}_{i \to j}^t)$ by $\mathbf{y}_{i \to j}^0 = \mathbf{x}_i^{0,N}$ and for $t \geq 0$

$$y_{i \to j}^{t+1}(r) \quad = \quad \sum_{\ell \in [N] \setminus j} A_{\ell i}^t \, f_r^\ell(\mathbf{y}_{\ell \to i}^t, t), \tag{3.8}$$

and

$$y_i^{t+1}(r) \quad = \quad \sum_{\ell \in [N]} A_{\ell i}^t \, f_r^\ell(\mathbf{y}_{\ell \to i}^t, t). \tag{3.9}$$

The asymptotic analysis of y^t is particularly simple because an independent random matrix A^t is used at each iteration. In particular, it is easy to establish state evolution for y^t. Our next result shows that y^t provides a good approximation for z^t.

Proposition 7. *Let* $(A(N), \mathcal{F}_N, x^{0,N})_{N \geq 1}$ *be a* (C, d)*-regular polynomial sequence of instances. Let* $\mathbf{z}_i^t$ *and* $\mathbf{y}_i^t$ *be the sequences of vectors obtained by iterating (3.5)-(3.6) and (3.8)-(3.9) respectively. Then for any* $t \geq 1$ *and any* $\mathbf{m} = (m(1), \ldots, m(q)) \in \mathbb{N}^q$*, there exists* K *independent of* N *such that, for any* $i \in [N]$*:*

$$\left| \mathbb{E}\left[(\mathbf{z}_i^t)^{\mathbf{m}} \right] - \mathbb{E}\left[(\mathbf{y}_i^t)^{\mathbf{m}} \right] \right| \leq K N^{-1/2}.$$

The proof of this proposition is provided in Section 3.4.

Finally, recall that we defined the sequences $(\mathbf{x}_i^t)_{t \in \mathbb{N}}$ with $\mathbf{x}_i^t \in \mathbb{R}^q$, by $\mathbf{x}_i^0$ and for $t \geq 0$,

$$x_i^{t+1}(r) = \sum_\ell A_{\ell i} f_r^\ell(\mathbf{x}_\ell^t, t) - \sum_\ell A_{\ell i}^2 \sum_s f_s^i(\mathbf{x}_i^{t-1}, t-1) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{x}_\ell^t, t)$$

Proposition 8. *Let* $(A(N), \mathcal{F}_N, x^{0,N})_{N \geq 1}$ *be a* (C, d)*-regular sequence of instances. Denote by* $\{x^t\}_{t \geq 0}$ *the corresponding AMP sequence and by* $\{z^t\}_{t \geq 0}$ *the sequence defined by (3.6) while iterating (3.5). Then for any* $t \geq 1$ *and* $m(1), \ldots, m(q) \geq 0$*, there exists* K *independent of* N *such that, for any* $i \in [N]$*,:*

$$\left| \mathbb{E}\left[(\mathbf{x}_i^t)^{\mathbf{m}} \right] - \mathbb{E}\left[(\mathbf{z}_i^t)^{\mathbf{m}} \right] \right| \leq K N^{-1/2}.$$

The proof of this proposition is provided in Section 3.5.

3.2   Tree representation

By assumption of Proposition 6, we have for each $\ell \in [N]$ and $r \in [q]$,

$$f_r^\ell(\mathbf{z}, t) = \sum_{i_1 + \cdots + i_q \leq d} c_{i_1, \ldots, i_q}^\ell(r, t) \prod_{s=1}^q z(s)^{i_s}, \tag{3.10}$$

where each coefficient $c_{i_1, \ldots, i_q}^\ell(r, t)$ belongs to $\mathbb{R}$ and has absolute value bounded by C (uniformly in $\ell \in [N]$, $i_1, \ldots, i_q$, and $t \in \mathbb{N}$).

We now introduce families of finite rooted labeled trees that will allow us to get a simple expression for the $z_{i \to j}^t(r)$'s and $z_i^t(r)$, see Lemma 1 below. For a vertex v in a rooted tree T different from the root, we denote by $\pi(v)$ the parent of v in T. We denote the root of T by $\circ$. We consider that the edges of T are directed towards the root and write $(u \to v) \in E(T)$ if $\pi(u) = v$. The unlabeled trees that we consider are such that the root and the leaves have degree one; each other vertex has degree at most $d + 1$, i.e. has at most d children. We now describe the possible labels on such trees. The label of the root is in $[N]$, the label of a leaf is in $[N] \times [q] \times \mathbb{N}^q$ and all other vertices have a label in $[N] \times [q]$. For a vertex v different from the root or a leaf, we denote its label by $(\ell(v), r(v))$ and call $\ell(v)$ its type and $r(v)$ its mark. The label (or type) of the root is also denoted by $\ell(\circ)$; the label of a leaf v is denoted by $(\ell(v), r(v), v[1], \ldots v[q])$. For a vertex $u \in T$, we denote $|u|$ its generation in the tree, i.e. its graph-distance from the root. Also for a vertex $u \in T$ (which is not a leaf), we denote by $u[r]$ the number of children of u with mark $r \in [q]$ (with the convention $u[0] = 0$). The children of such a node are ordered with respect to their mark: the labels of the children of u are then $(\ell^1, 1), \ldots, (\ell^{u[1]}, 1), (\ell^{u[1]+1}, 2), \ldots, (\ell^{u[1]+\cdots+u[q]}, q)$, where each $(\ell^{u[0]+\cdots+u[i]}, \ldots, \ell^{u[0]+\cdots+u[i+1]-1})$ is a $u[i+1]$-tuple with coordinates in $[N]$. We denote by $L(T)$ the set of leaves of a tree T, i.e. the set

of vertices of T with no children. For $v \in L(T)$, its label $(\ell(v), r(v), v[1], \ldots v[q])$ is such that for all $i \in [q]$, $v[i] \in \mathbb{N}$ and $v[1] + \cdots + v[q] \leq d$. We will distinguish between two types of leaves: those with maximal depth $t = \max\{|v|, \ v \in L(T)\}$ and the remaining ones. If $v \in L(T)$ and $|v| \leq t - 1$, then we impose $v[1] = \cdots = v[q] = 0$. This case corresponds to 'natural' leaves and since they have no children, the notation is consistent with the notation introduced for other nodes of the tree. For all other leaves, we do not make this assumption so that $v[1] + \cdots + v[q]$ can take any value in $[d]$. These leaves are 'artificial' and can be thought of as leaves resulting from cutting a larger tree after generation t so that the vector of the $v[r]$'s keeps the information on the number of children with mark r in the original tree.

Definition 9. *We denote by $\mathcal{T}^t$ the set of labeled trees T with t generations as above that satisfy the following conditions:*

1. *If $v_1 = \circ, v_2, \ldots, v_k$ is a path starting from the root (i.e. with $\pi(v_{i+1}) = v_i$ for $i \geq 1$), then the corresponding sequence of types $\ell(v_i)$ is non-backtracking. i.e., for any $1 \leq i \leq k - 2$, the three labels $\ell(v_i), \ell(v_{i+1})$ and $\ell(v_{i+2})$ are distinct.*

2. *If $u \in L(T)$ and $|u| \leq t - 1$ (i.e. u is a 'natural' leaf), then we have $v[1] + \cdots + v[q] = 0$.*

3. *If $u \in L(T)$ and $|u| = t$ (i.e. u is an 'artificial' leaf) then we have $v[1] + \cdots + v[q] \leq d$.*

We also denote by $\overline{\mathcal{T}}^t$ the set of trees that satisfy conditions 2 and 3, but not necessarily the non-backtracking condition 1. Hence $\mathcal{T}^t \subseteq \overline{\mathcal{T}}^t$.

We also let $\mathcal{U}^t$ be the same set of trees in which marks have been removed (i.e. we identify any two trees that differ in the marks but not on type). Analogously, $\overline{\mathcal{U}}^t$ is the set of trees in which marks have been removed, but do not necessarily the non-backtracking condition 1.

For a labeled tree $T \in \mathcal{T}^t$ and a set of coefficients $\mathbf{c} = (c^\ell_{i_1, \ldots, i_q}(r, t))$, we define three weights:

$$
\begin{aligned}
A(T) &= \prod_{(u \to v) \in E(T)} A_{\ell(u)\ell(v)}, \\
\Gamma(T, \mathbf{c}, t) &= \prod_{(u \to v) \in E(T)} c^{\ell(u)}_{u[1], \ldots, u[q]}(r(u), t - |u|), \\
x(T) &= \prod_{v \in L(T)} \prod_{s=1}^{q} \left(x^{0,N}_{\ell(v)}(s) \right)^{v[s]}.
\end{aligned}
$$

We define

(a) $\mathcal{T}^t_{i \to j}(r) \subset \mathcal{T}^t$ the family of trees such that: (i) The root has type i; (ii) The root has only one child, call it v; (iii) The type of v is $\ell(v) \notin \{i, j\}$ and its mark is $r(v) = r$.

(b) $\mathcal{T}^t_i(r) \subset \mathcal{T}^t$ the family of trees such that: (i) The root has type i; (ii) The root has only one child, call it v; (iii) The type of v is $\ell(v) \neq i$ and its mark is $r(v) = r$.

The sets of trees $\mathcal{U}^t_i(r)$ and $\mathcal{U}^t_{i \to j}(r)$ are obtained from $\mathcal{T}^t_i(r)$ and $\mathcal{T}^t_{i \to j}(r)$ by removing marks.

Lemma 1. *Let $(A(N), \mathcal{F}_N, x^{0,N})_{N \geq 1}$ be a polynomial sequence of AMP instances. Denote by $\mathbf{z}_i^t$ the orbit defined by (3.6) while iterating (3.5) with matrix A. Then,*

$$z_{i \to j}^t(r) = \sum_{T \in \mathcal{T}_{i \to j}^t(r)} A(T) \Gamma(T, \mathbf{c}, t) x(T), \tag{3.11}$$

$$z_i^t(r) = \sum_{T \in \mathcal{T}_i^t(r)} A(T) \Gamma(T, \mathbf{c}, t) x(T). \tag{3.12}$$

Proof. We first prove (3.11) by induction on t. For $t = 1$ we have, by definition

$$z_{i \to j}^1(r) = \sum_{\ell \in [N] \backslash j} \sum_{i_1 + \cdots + i_q \leq d} A_{\ell i} \, c_{i_1, \ldots, i_q}^\ell(r, 0) \prod_{s=1}^q \left(x_{\ell \to i}^{0,N}(s) \right)^{i_s}$$

This expression corresponds exactly to equation (3.11) since trees in $\mathcal{T}_{i \to j}^1(r)$ have a root with label i and with one child with label $(\ell, r, i_1, \ldots, i_q)$ for some $\ell \notin \{i, j\}$ and $i_1 + \cdots + i_q \leq d$.

To prove the induction, we start with Eq. (3.5), which yields

$$z_{i \to j}^{t+1}(r) = \sum_{\ell \in [N] \backslash j} A_{\ell i} \sum_{i_1 + \cdots + i_q \leq d} c_{i_1, \ldots, i_q}^\ell(r, t) \prod_{s=1}^q \left(z_{\ell \to i}^t(s) \right)^{i_s}$$

Using the induction hypothesis, we get

$$\prod_{s=1}^q \left(z_{\ell \to i}^t(s) \right)^{i_s} = \prod_{s=1}^q \left(\sum_{T \in \mathcal{T}_{\ell \to i}^t(s)} A(T) \Gamma(T, \mathbf{c}, t) x(T) \right)^{i_s}$$

$$= \sum_{[\mathcal{T}_{\ell \to i}^t(s)]^{i_1 + \cdots + i_q}} \prod_{s=1}^q \prod_{k=1}^{i_s} A(T_k^s) \Gamma(T_k^s, \mathbf{c}, t) x(T_k^s),$$

where the last expression is a sum over all $(i_1 + \cdots + i_q)$-tuples of trees with the first i_1 trees in $\mathcal{T}_{\ell \to i}^t(1)$, the following i_2 in $\mathcal{T}_{\ell \to i}^t(2)$, and so on.

Hence, we get

$$z_{i \to j}^{t+1}(r) = \sum_{\ell \in [N] \backslash j} \sum_{i_1 \ldots i_q} \sum_{[\mathcal{T}_{\ell \to i}^t(s)]^{i_1 + \cdots + i_q}} A_{\ell i} \, c_{i_1, \ldots, i_q}^\ell(r, t) \prod_{s=1}^q \prod_{k=1}^{i_s} A(T_k^s) \Gamma(T_k^s, \mathbf{c}, t) x(T_k^s). \tag{3.13}$$

The claim now follows by observing that the set of trees in $\mathcal{T}_{i \to j}^{t+1}(r)$ is in bijection with the set of pairs constituted by a label (ℓ, r) with $\ell \notin \{i, j\}$ and a $(i_1 + \cdots + i_q)$-tuple of trees with exactly i_s trees belonging to $\mathcal{T}_{\ell \to i}^k(s)$ for $s \in [q]$. Indeed, take a root with label i and one child say v, with label (ℓ, r) for some $\ell \notin \{i, j\}$ and with a $(i_1 + \cdots + i_q)$-tuple of trees with exactly i_s trees belonging to $\mathcal{T}_{\ell \to i}^t(s)$ for $s \in [q]$. Now take v as the root of these $(i_1 + \cdots + i_q)$ trees, the order in the tuple giving the order of the subtrees of v. Note that the root of each subtree in $\mathcal{T}_{\ell \to i}^t(s)$ has type ℓ and in the resulting tree will get mark r. The proof of (3.12) follows by the same argument, the only change is that in the sum in (3.13), we need now to include $\ell = j$. $\square$

3.3 Proof of Proposition 6

We are now in position to prove Proposition 6.

Proof. For notational simplicity, we consider the case $m(r) = m$, and $m(s) = 0$ for all $s \in [q] \setminus r$. Thanks to Lemma 1, we have

$$\mathbb{E}\left[\left(z_i^t(r)\right)^m\right] = \sum_{T_1,\ldots,T_m \in \mathcal{T}_i^t(r)} \left[\prod_{\ell=1}^m \Gamma(T_\ell, \mathbf{c}, t)\right] \mathbb{E}\left[\prod_{\ell=1}^m x(T_\ell)\right] \mathbb{E}\left[\prod_{\ell=1}^m A(T_\ell)\right]. \tag{3.14}$$

Since $\mathbf{c}$ is fixed in this section, we omit to write it in $\Gamma(T, t)$. Notice that the general case $\mathbf{m} = (m(1),\ldots,m(q)) \in \mathbb{N}^q$ admits a very similar representation whereby the sum over $T_1,\ldots,T_m \in \mathcal{T}_i^t(r)$ is replaced by sums over $T_1,\ldots,T_{m(1)} \in \mathcal{T}_i^t(1)$, $T_1,\ldots,T_{m(2)} \in \mathcal{T}_i^t(2)$, $\ldots$, $T_1,\ldots,T_{m(q)} \in \mathcal{T}_i^t(q)/$ The argument goes through essentially unchanged.

We have $\Gamma(T_\ell, t) \leq C^{d^{t+1}}$. We first concentrate on the term $\mathbb{E}[\prod_{\ell=1}^m A(T_\ell)]$. Recall that, from subgaussian property of entries of A: $\mathbb{E}\left(e^{\lambda A_{ij}}\right) \leq e^{\frac{C\lambda^2}{2N}}$. Now using Lemma 12 from Appendix D we get for all $i < j \in [N]$

$$\mathbb{E}\left[|A_{ij}|^s\right] \leq 2\left(\frac{s}{e}\right)^s \lambda^{-s} e^{\frac{C\lambda^2}{2N}} \leq 2C^{\frac{s}{2}} \left(\frac{s}{e}\right)^{\frac{s}{2}} N^{-\frac{s}{2}}, \tag{3.15}$$

obtained by taking $\lambda = \sqrt{Ns/C}$.

For a labeled tree T, we define $\phi(T) = \{\phi(T)_{ij} \in \mathbb{N}, i \leq j \in [N]\}$ where $\phi(T)_{ij}$ is the number of occurrences in T of an edge $(u \to v)$ with endpoints having types $\ell(u), \ell(v) \in \{i, j\}$. Hence we have

$$A(T) = \prod_{i<j\in[N]} A_{ij}^{\phi(T)_{ij}} \quad \text{and} \quad \mathbb{E}\left[\prod_{\ell=1}^m A(T_\ell)\right] = \prod_{i<j\in[N]} \mathbb{E}\left[A_{ij}^{\sum_{\ell=1}^m \phi(T_\ell)_{ij}}\right]. \tag{3.16}$$

Since the mean of each entry of the matrix A is zero, in Equation (3.14), we can restrict the sum to $T_1,\ldots,T_m$ such that for all $i < j \in [N]$, $\sum_{\ell=1}^m \phi(T_\ell)_{ij} < 2$ implies $\sum_{\ell=1}^m \phi(T_\ell)_{ij} = 0$.

We now concentrate on the sum restricted to $T_1,\ldots,T_m$ such that moreover there exists $i < j \in [N]$ such that $\sum_{\ell=1}^m \phi(T_\ell)_{ij} \geq 3$. For such a m-tuple $T_1,\ldots,T_m$, we denote $\mu = \mu(T_1,\ldots,T_m) = \sum_{i<j} \sum_{\ell=1}^m \phi(T_\ell)_{ij}$. Let $\mathbf{G}$ be the graph obtained by taking the union of the T_ℓ's and identifying () vertices v with the same type $\ell(v)$. We define $e(T_1,\ldots,T_m) = \sum_{i<j} \mathbf{1}(\sum_{\ell=1}^m \phi(T_\ell)_{ij} \geq 1)$ which is the number of edges counted without multiplicity in $\mathbf{G}$. Since there exists $i < j$ with $\sum_{\ell=1}^m \phi(T_\ell)_{ij} \geq 3$, we have $3 + 2(e(T_1,\ldots,T_m) - 1) \leq \mu$, i.e. $e(T_1,\ldots,T_m) \leq \frac{\mu-1}{2}$. Using Eq. (3.15), we get

$$\left|\mathbb{E}\left[\prod_{\ell=1}^m A(T_\ell)\right]\right| \leq \prod_{i<j\in[N]} \mathbb{E}\left[|A_{ij}|^{\sum_{\ell=1}^m \phi(T_\ell)_{ij}}\right]$$

$$\leq \left(2C^{\frac{\mu}{2}}\left(\frac{\mu}{e}\right)^{\frac{\mu}{2}}\right)^{(\mu-1)/2} N^{-\frac{\mu}{2}}, \tag{3.17}$$

since in the product on the right-hand side of (3.16), there are $e(T_1,\ldots,T_m)$ terms different form one, i.e. at most $(\mu-1)/2$ contributing terms.

14

We now compute an upper bound on

$$\sum_{T_1,\ldots,T_m}^{(\mu)} \mathbb{E}\left[\left|\prod_{\ell=1}^m x(T_\ell)\right|\right],$$

where the sum $\sum^{(\mu)}$ ranges on m-tuple of trees in $\mathcal{T}_i^t(r)$ such that $\sum_{i<j}\sum_{\ell=1}^m \phi(T_\ell)_{ij} = \mu$. First note that for any $\mathbf{x} \in \mathbb{R}^q$, we have for any $p \geq 2$:

$$\|\mathbf{x}\|_p^p \leq \|\mathbf{x}\|_2^p \leq \max\left(\exp(\|\mathbf{x}\|_2^2), p^p\right).$$

Hence the condition $\frac{1}{N}\sum_{i=1}^N \exp(\|\mathbf{x}_i^{0,N}\|_2^2/C) \leq C$ ensures that for any $p \geq 2$,

$$\frac{1}{N}\sum_{i=1}^N \|\mathbf{x}_i^{0,N}\|_p^p \leq C_p.$$

Therefore,

$$
\begin{aligned}
\sum_{T_1,\ldots,T_m}^{(\mu)} \prod_{\ell=1}^m |x(T_\ell)| &\leq \left(q^m \sum_{j=1}^N \sum_{s=1}^q \left(1 + |x_j^{0,N}(s)| + \cdots + |x_j^{0,N}(s)|^{md}\right)\right)^{\frac{\mu-1}{2}} \quad (3.18)\\
&= (q^m N)^{\frac{\mu-1}{2}}\left(q + \sum_{k=1}^{md}\frac{1}{N}\sum_{j=1}^N \|\mathbf{x}_j^{0,N}\|_k^k\right)^{\frac{\mu-1}{2}}\\
&\leq \left(q^m(q + \sum_{k=1}^{md} C_k)\right)^{\frac{\mu-1}{2}} N^{\frac{\mu-1}{2}},
\end{aligned}
$$

where the last inequality is valid for $N \geq C$. To see why (3.18) is true, note that the graph $\mathbf{G}$ is connected since all trees $T_1,\ldots,T_m$ have the same type i at the root. Therefore, the number of vertices in $\mathbf{G}$ is at most $e(T_1,\ldots,T_m) + 1 \leq \frac{\mu-1}{2} + 1$. Since all T_ℓ's have the same root which has type i, $\mathbf{G}$ has at most $\frac{\mu-1}{2}$ distinct vertices which are distinct from the one associated to the root. In particular, all trees $T_1,\ldots,T_m$ together have at most $\frac{\mu-1}{2}$ distinct types among their leaves. The factor q^m comes from the fact that for each type j there are at most q^m choices for its m marks r corresponding to the m trees. Now each leaf with type j will contribute a factor $\prod_{s=1}^q \left(x_j^{0,N}(s)\right)^{n_s}$ with $\sum_s n_s \leq md$.

It is now easy to conclude, since we can decompose the sum in (3.14) in two terms, the first term say $S_1(A)$ consists of the contribution of the m-tuples $T_1,\ldots,T_m$ such that for all i,j, $\sum_{\ell=1}^m \phi(T_\ell)_{ij} \in \{0,2\}$ while the second term denoted by $S_2(A)$ consists of the remaining contribution. We have $S_1(A) = S_1(\tilde{A})$ and, using (3.17) and (3.18), we get:

$$|S_2(A)| \leq \sum_{\mu \leq md^{t+1}} C^{d^{t+1}+\frac{\mu-1}{2}} C' N^{\frac{\mu-1}{2}} N^{-\frac{\mu}{2}} = O\left(N^{-\frac{1}{2}}\right), \quad (3.19)$$

which concludes the proof Proposition 6. Here we used the fact that all values μ, q, and $\{C_k\}_{k=0}^{md}$ are independent of N. $\qquad\square$

We end this section by showing that the term $S_1(A)$ can be further reduced. This result will be useful in the sequel and we state it as the following lemma.

Lemma 2. *Recall that we denoted by $S_1(A)$ the term in the sum (3.14), consisting of the contribution of the m-tuples $T_1, \ldots, T_m$ such that for all i, j, $\sum_{\ell=1}^{m} \phi(T_\ell)_{ij} \in \{0, 2\}$. We further decompose $S_1(A) = T(A) + R(A)$ in two terms where the first term $T(A)$ corresponds to the sum over trees $T_1, \ldots T_m$ such that the resulting graph $\mathbf{G}$ obtained by taking the union of the T_ℓ's and identifying vertices v with the same type $\ell(v)$, is a tree (each edge having multiplicity two). Then there exists K (independent of N) such that:*

$$\left| \mathbb{E}\left[z_i^t(r)^m \right] - T(A) \right| = KN^{-1/2},$$
$$\left| \mathbb{E}\left[z_i^t(r)^m \right] \right| \leq K, \qquad \left| \mathbb{E}\left[z_{i \to j}^t(r)^m \right] \right| \leq K.$$

Proof. We have by definition $\mathbb{E}\left[(z_i^t(r)^m) \right] = T(A) + R(A) + S_2(A)$, so that thanks to (3.19), we need only to show that $R(A) = O\left(N^{-1/2} \right)$.

For any m-tuple $T_1, \ldots, T_m$ such that for all i, j, $\sum_{\ell=1}^{m} \phi(T_\ell)_{ij} \in \{0, 2\}$, we have with the same notation as above: $e(T_1, \ldots, T_m) = \frac{\mu}{2}$. The number of vertices in $\mathbf{G}$ is at most $1 + e(T_1, \ldots, T_m)$ with equality if and only if $\mathbf{G}$ is a tree (remember that $\mathbf{G}$ is always connected as all trees T_ℓ's share the same root). Hence for the cases that $\mathbf{G}$ is not a tree it has at most $\frac{\mu}{2} - 1$ vertices that serve as leaves of a tree among $T_1, \ldots, T_m$. By the same argument as above we get

$$|T(A)| \quad \leq \quad \sum_{\mu \leq md^{t+1}} KN^{\frac{\mu}{2}} N^{-\frac{\mu}{2}} = O(1) \tag{3.20}$$

$$|R(A)| \quad \leq \quad \sum_{\mu \leq md^{t+1}} KN^{\frac{\mu}{2}-1} N^{-\frac{\mu}{2}} = O(N^{-1}), \tag{3.21}$$

and the claim follows. $\qquad\qquad\square$

3.4   Proof of Proposition 7

The proof follows the same approach as for Proposition 6. For notational simplicity, we consider the case $m(r) = m$, and $m(s) = 0$ for all $s \in [q] \setminus r$. The general case follows by the same argument. For $\mathbf{y}$, we are using a different matrix at each iteration and we need to define a new weight associated to trees $T \in \mathcal{T}^t$ as follows:

$$\overline{A}(T, t) \quad = \quad \prod_{(u \to v) \in E(T)} A_{\ell(u)\ell(v)}^{t-|u|}. \tag{3.22}$$

In the particular case where the sequence $\{A^t\}_{t \in \mathbb{N}}$ is constant (i.e., equals to A), this expression reduces to $A(T)$ defined previously. Similar to Lemma 1 for $\mathbf{x}$, we have now

$$y_{i \to j}^t(r) \quad = \quad \sum_{T \in \mathcal{T}_{i \to j}^t(r)} \overline{A}(T, t) \Gamma(T, \mathbf{c}, t) x(T),$$

$$y_i^t(r) \quad = \quad \sum_{T \in \mathcal{T}_i^t(r)} \overline{A}(T, t) \Gamma(T, \mathbf{c}, t) x(T),$$

16

so that we get

$$\mathbb{E}\left[(y_i^t(r))^m\right] = \sum_{T_1,\ldots,T_m \in \mathcal{T}_i^t(r)} \left[\prod_{\ell=1}^m \Gamma(T_\ell, \mathbf{c}, t)\right] \mathbb{E}\left[\prod_{\ell=1}^m x(T_\ell)\right] \mathbb{E}\left[\prod_{\ell=1}^m \overline{A}(T_\ell, t)\right]. \qquad (3.23)$$

For a labeled tree T, we define $\varphi(T) = \{\varphi(T)_{ij}^g \geq 0,\ i \leq j \in [N],\ d \geq 1\}$ where $\varphi(T)_{ij}^g$ is the number of occurrences in T of an edge $(u \to v)$ with endpoints having labels $\ell(u), \ell(v) \in \{i, j\}$ and with generation $|u| = g$. In particular, we have $\sum_g \varphi(T)_{ij}^g = \phi(T)_{ij}$ which was defined in the proof of Proposition 6. Hence we have with $\mu = \sum_{i<j} \sum_{\ell=1}^m \phi(T_\ell)_{ij}$,

$$\left|\mathbb{E}\left[\prod_{\ell=1}^m \overline{A}(T_\ell, t)\right]\right| \overset{(a)}{=} \prod_{i<j\in[N]} \prod_g \left|\mathbb{E}\left[A_{ij}^{\sum_{\ell=1}^m \varphi(T_\ell)_{ij}^g}\right]\right|$$

$$\leq \prod_{i<j\in[N]} \prod_g \mathbb{E}\left[|A_{ij}|^{\sum_{\ell=1}^m \varphi(T_\ell)_{ij}^g}\right]$$

$$\overset{(b)}{\leq} \left(2C^{\frac{\mu}{2}}\left(\frac{\mu}{e}\right)^{\frac{\mu}{2}}\right)^{(\mu-1)/2} N^{-\frac{\mu}{2}} \qquad (3.24)$$

where (a) holds since $\{A^t\}_{t\in\mathbb{N}}$ is an iid sequence with the same distribution as $A(N)$, and (b) follows by the same argument as in (3.17). The inequality (3.24) implies that the bounds (3.19) and (3.21) are still valid with the weight of a tree given by (3.22) (the term $\mathbb{E}\left[\prod_{\ell=1}^m x(T_\ell)\right]$ can be treated as in previous section).

As in the proof of Proposition 6, we define the graph $\mathbf{G}$ obtained by taking the union of the T_ℓ's and identifying vertices v with the same type $\ell(v)$. By Lemma 2, we need only to concentrate on the term $T(\overline{A})$ corresponding to m-tuples $T_1, \ldots, T_m$ such that each edge in $\mathbf{G}$ has multiplicity 2 and such that $\mathbf{G}$ is a tree. Indeed, the proposition will follow, once we prove

$$T(A) = T(\overline{A}), \qquad (3.25)$$

where $T(A)$ was defined in Lemma 2 and $T(\overline{A})$ is the corresponding term with the weight of a tree given by (3.22). First note that for any $T_1, \ldots, T_m$ such that $\mathbb{E}\left[\prod_{\ell=1}^m \overline{A}(T_\ell, t)\right] \neq 0$, we have

$$\mathbb{E}\left[\prod_{\ell=1}^m \overline{A}(T_\ell, t)\right] = \mathbb{E}\left[\prod_{\ell=1}^m A(T_\ell)\right].$$

Now suppose that we have $\mathbb{E}\left[\prod_{\ell=1}^m A(T_\ell)\right] \neq 0 = \mathbb{E}\left[\prod_{\ell=1}^m \overline{A}(T_\ell, t)\right]$. This can only happen, if an edge in $\mathbf{G}$ connecting types say i and j has multiplicity 2 but appears at different generations in the original trees T_ℓ's. Suppose this edge appears twice in say T_1 at on the same branch and at different generations, i.e. there exists $(a \to b)$ and $(c \to d) \in E(T_1)$ with $\{\ell(a), \ell(b)\} = \{\ell(c), \ell(d)\} = \{i, j\}$, $|a| < |c|$ and the edge $(a \to b)$ is on the path that connects c, d to the root. Thanks to the non-backtracking property, these two edges cannot be adjacent, i.e. $a \neq d$. But then these edges create a cycle in $\mathbf{G}$, contradiction. Suppose now that these edge appears in T_1 and T_2 in different generations, i.e. there exists $(a \to b) \in E(T_1)$ and $(c \to d) \in E(T_2)$ with $\{\ell(a), \ell(b), \ell(c), \ell(d)\} = \{i, j\}$ and $|a| < |c|$. Then the same reasoning shows that they will create a cycle in $\mathbf{G}$ since b and d are connected to the roots of T_1 and T_2 respectively which are both identify to a single vertex in $\mathbf{G}$. The latter argument can be used for the case where both edges belong to the same tree T_1 but they lie in different branches. Hence we obtain again a contradiction.

3.5 Proof of Proposition 8

Proof. As in the proof of Proposition 6, we will rely on a representation of $x_i^t(r)$ based on labeled trees defined as in Section 3.2. In the present case it is however more convenient to work with trees from which marks have been removed, i.e. we identify any two trees in which the vertex marks are different but the types are the same. Notice that Eqs. (3.11), (3.12) imply

$$z_{i \to j}^t(r) = \sum_{T \in \mathcal{U}_{i \to j}^t(r)} A(T)\Gamma'(T, \mathbf{c}, t)x(T), \tag{3.26}$$

$$z_i^t(r) = \sum_{T \in \mathcal{U}_i^t(r)} A(T)\Gamma'(T, \mathbf{c}, t)x(T), \tag{3.27}$$

where $\Gamma'(T, \mathbf{c}, t)$ is obtained by summing $\Gamma(T, \mathbf{c}, t)$ over all trees T that coincide up to marks. In the following, with a slight abuse of notation, we will write $\Gamma(T, \mathbf{c}, t)$ instead of $\Gamma'(T, \mathbf{c}, t)$.

In a directed labeled graph, we define a backtracking path of length 3 as a path $a \to b \to c \to d$ such that $\ell(a) = \ell(c)$ and $\ell(b) = \ell(d)$. We define a backtracking star as a set of vertices $a \to b \to c$ and $a'(\neq a) \to b$ such that $\ell(a) = \ell(a') = \ell(c)$. We define $\mathcal{B}^t$ as the set of rooted labeled trees T in $\overline{\mathcal{U}}^t$, that satisfy the following conditions:

- If $u \to v \in E(T)$, then $\ell(u) \neq \ell(v)$ and there exists in T at least one backtracking path of length 3 or one backtracking star.

Then, we define $\mathcal{B}_i^t$ as the subset of trees in $\mathcal{B}^t$ with root having type i and only one child with type ℓ with $\ell \neq i$.

Lemma 3. *Under the same assumptions as in Proposition 8, we have*

$$x_i^t(r) = z_i^t(r) + \sum_{T \in \mathcal{B}_i^t} A(T)\widetilde{\Gamma}(T, t, r)x(T),$$

for some $\widetilde{\Gamma}(T, t, r)$ which is bounded uniformly as $|\widetilde{\Gamma}(T, t, r)| \leq K(d, C, t)$.

Proof. Following the same argument as in Lemma 1, it is easy to prove by induction on t that we can find $\widetilde{\Gamma}(T, t, r)$ such that

$$x_i^t(r) = \sum_{T \in \overline{\mathcal{U}}_i^t} A(T)\widetilde{\Gamma}(T, t, r)x(T), \tag{3.28}$$

with $|\widetilde{\Gamma}(T, t, r)| \leq K(d, C, t)$. The terms $A_{i\ell}f_r^\ell(\mathbf{x}_\ell^t, t)$ can be handled exactly as in Lemma 1. Concerning the terms $A_{i\ell}^2 f_s^i(\mathbf{x}_i^{t-1}, t-1)\frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{x}_\ell^t, t)$, it can be interpreted as a sum on the following trees in $\overline{\mathcal{U}}$: the type of the root is i and the root has one child with type ℓ. This child has at most $d-1$ subtrees in $\overline{\mathcal{U}}^t$ coming from the term $\frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{x}_\ell^t, t)$ (which is a polynomial with degree at most $d-1$) and one child say u with type i. This child u is the root of at most d subtrees in $\overline{\mathcal{U}}^{t-1}$ coming from the term $f_s^i(\mathbf{x}_i^{t-1}, t-1)$. We see that the resulting tree is in $\overline{\mathcal{U}}^{t+1}$. Now to see that $|\widetilde{\Gamma}(T, t, r)| \leq K(d, C, t)$, note that each polynomial $f_r^\ell(\cdot, t)$ (resp. $\frac{\partial f_r^\ell}{\partial x(s)}(\cdot, t)$) has coefficients bounded by C (resp. dC) so that taking into account the contribution of each term in decomposition (3.28), we easily get

$$|\widetilde{\Gamma}(T, t+1, r)| \leq dC^2 \left[K(d, C, t)^d + K(d, C, t)^{d-1} K(d, C, t-1)^d \right].$$

18

It remains to prove that $\widetilde{\Gamma}(T,t,r)$ agrees with the expression in Lemma 1, cf. Eq. (3.26), (3.27), for $T \in \mathcal{U}_i^t(r)$ and is zero for trees in $\overline{\mathcal{U}}^t \setminus \mathcal{B}_i^t$. The proof of this fact will proceed by induction on t. The cases $t = 0, 1$ are clear since $\mathcal{B}_i^t = \emptyset$. For $t \geq 1$, we define

$$z_{\ell,i}^t(r) = A_{i\ell} f_r^i(\mathbf{z}_{i \to \ell}^{t-1}, t-1), \quad e_\ell^t(r) = \sum_{T \in \mathcal{B}_\ell^t} A(T)\widetilde{\Gamma}(T,t,r)x(T), \quad d_{\ell,i}^t(r) = z_{\ell,i}^t(r) + e_\ell^t(r)$$

so that we have by the induction hypothesis, $\mathbf{x}_\ell^t = \mathbf{z}_{\ell \to i}^t + \mathbf{z}_{\ell,i}^t + \mathbf{e}_\ell^t = \mathbf{z}_{\ell \to i}^t + \mathbf{d}_{\ell,i}^t$.

Since $f_r^\ell(\,\cdot\,, t)$ is a polynomial, we have

$$f_r^\ell(\mathbf{x}_\ell^t, t) = f_r^\ell(\mathbf{z}_{\ell \to i}^t, t) + \sum_s \left(z_{\ell,i}^t(s) + e_\ell^t(s)\right) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{z}_{\ell \to i}^t, t)$$

$$+ \sum_{n_1 + \cdots + n_q \geq 2} \prod_{s=1}^q \frac{\left(d_{\ell,i}^t(s)\right)^{n_s}}{n_s!} \frac{\partial^{n_1 + \cdots + n_q} f_r^\ell}{\partial x(1)^{n_1} \ldots \partial z(q)^{n_q}}(\mathbf{z}_{\ell \to i}^t, t),$$

where the last sum contains a finite number of non-zero terms.

Multiplying by $A_{i\ell}$ and summing over $\ell \in [N]$, the first term on the right hand side gives exactly $z_i^{t+1}(r)$. The second term gives:

$$\sum_\ell A_{\ell i}^2 \sum_s f_s^i(\mathbf{z}_{i \to \ell}^{t-1}, t-1) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{z}_{\ell \to i}^t, t) + \sum_\ell A_{\ell i} \sum_s e_\ell^t(s) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{z}_{\ell \to i}^t, t)\,.$$

From now on and to lighten the notation, we omit the second argument of the functions f_r^ℓ. Hence we have

$$\begin{aligned} x_i^{t+1}(r) &= z_i^{t+1}(r) - \sum_\ell A_{\ell i}^2 \sum_s \left(f_s^i(\mathbf{x}_i^{t-1}) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{x}_\ell^t) - f_s^i(\mathbf{z}_{i \to \ell}^{t-1}) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{z}_{\ell \to i}^t) \right) \\ &\quad + \sum_\ell A_{\ell i} \sum_s e_\ell^t(s) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{z}_{\ell \to i}^t) \\ &\quad + \sum_\ell A_{\ell i} \sum_{n_1 + \cdots + n_q \geq 2} \prod_{s=1}^q \frac{\left(d_{\ell,i}^t(s)\right)^{n_s}}{n_s!} \frac{\partial^{n_1 + \cdots + n_q} f_r^\ell}{\partial x(1)^{n_1} \ldots \partial x(q)^{n_q}}(\mathbf{z}_{\ell \to i}^t). \end{aligned}$$

(3.29)

We now show that each contribution on the right hand side (except $z_i^{t+1}(r)$) can be written as a sum of terms $A(T)\widetilde{\Gamma}(T, t+1, r, x^0)$ over trees $T \in \mathcal{B}_i^{t+1}$ that we construct explicitly.

First consider the terms of the form: $A_{\ell i} e_\ell^t(s) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{z}_{\ell \to i}^t)$. By definition $e_\ell^t(s)$ can be written as a sum over trees in $\mathcal{B}_\ell^t$ and by Lemma 1, the r-th component of $\mathbf{z}_{\ell \to i}^t$ can be written as a sum over trees in $\mathcal{U}_{\ell \to i}^t(r)$. Hence by the same argument as in the proof of Lemma 1, we see that $A_{\ell i} e_\ell^t(s) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{x}_{\ell \to i}^t)$ can be written as a sum over trees with root having type i, one child say v with type ℓ. This vertex v is the root of a tree in $\mathcal{B}_\ell^t$ (corresponding to the factor $e_\ell^t(s)$) and a set of trees in $\mathcal{U}_{\ell \to i}^t(1), \ldots, \mathcal{U}_{\ell \to i}^t(q)$ (corresponding to the factor $\frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{z}_{\ell \to i}^t)$). This tree clearly belongs to $\mathcal{B}_i^{t+1}$.

We now treat the terms in the first line. Again, we have

$$f_s^i(\mathbf{x}_i^{t-1}) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{x}_\ell^t) = f_s^i(\mathbf{z}_{i \to \ell}^{t-1}) \frac{\partial f_r^\ell}{\partial x(s)}(\mathbf{z}_{\ell \to i}^t) + g(\mathbf{d}_{i,\ell}^{t-1}, \mathbf{d}_{\ell,i}^t, \mathbf{z}_{i \to \ell}^{t-1}, \mathbf{z}_{\ell \to i}^t),$$

19

where g is a polynomial with either a positive power of a component of $\mathbf{d}_{i,\ell}^{t-1}$ or of $\mathbf{d}_{\ell,i}^t$. Hence, we only need to construct trees in $\mathcal{B}_i^{t+1}(r)$ corresponding to terms of the following form: for $\sum_s (a_s + b_s) \geq 1$,

$$A_{\ell i}^2 \prod_s \left(d_{i,\ell}^{t-1}(s)\right)^{a_s} \left(d_{\ell,i}^t(s)\right)^{b_s} \left(z_{i\to\ell}^{t-1}(s)\right)^{c_s} \left(z_{\ell\to i}^t(s)\right)^{d_s}.$$

Let first consider the term: $A_{\ell i}^2 \prod_s \left(z_{i\to\ell}^{t-1}(s)\right)^{c_s} \left(z_{\ell\to i}^t(s)\right)^{d_s}$. It can be interpreted as a sum on the following family of trees: the type of the root is i and the root has one child with type ℓ. This child has d_s subtrees in $\mathcal{U}_{\ell\to i}^t(s)$ and one child denoted u with type i. This child u has c_s subtrees in $\mathcal{U}_{i\to\ell}^{t-1}(s)$. Note that the only backtracking path in such a tree is the path from u to the root with types i, ℓ, i. In particular such a tree does not belong to $\mathcal{B}_i^t(r)$.

We assume now that there exists s with $a_s \geq 1$. We need to interpret the multiplication by $d_{i,\ell}^{t-1}(s) = z_{i,\ell}^{t-1}(s) + e_i^{t-1}(s)$. First consider the case of $e_i^{t-1}(s)$, this corresponds to add a subtree in $\mathcal{B}_i^{t-1}$ to the vertex u. As in previous analysis, we clearly obtain a tree in $\mathcal{B}_i^{t+1}$. The term $z_{i,\ell}^{t-1}(s)$ corresponds to adding a child of type ℓ to the vertex u which is the root of a subtree in $\mathcal{U}_{\ell\to i}^{t-2}(s)$, in particular we introduce a backtracking path of length 3 so that again the resulting tree is in $\mathcal{B}_i^{t+1}$. Similarly if $b_s \geq 1$, the multiplication by $d_{\ell,i}^t(s)$ will correspond to add a subtree to the child of the root, resulting in either adding a backtracking path of length 3 or adding a backtracking star.

The last term of the form

$$A_{\ell i} \prod_{s=1}^q \frac{\left(d_{\ell,i}^t(s)\right)^{n_s}}{n_s!} \frac{\partial^{n_1+\cdots+n_q} f_r^\ell}{\partial x(1)^{n_1} \ldots \partial x(q)^{n_q}}(\mathbf{z}_{\ell\to i}^t),$$

with $n_1 + \cdots + n_q \geq 2$ can be analyzed by the same kind of argument by noticing that the factor $A_{i\ell} z_{\ell,i}^t(s) z_{\ell,i}^t(s')$ corresponds to a backtracking star. □

The proof of Proposition 8 follows from the same arguments as in the proof of Proposition 6. Once more, for simplicity, we only consider the case $m(r) = m$ and $m(s) = 0$ for $s \neq r$, the general case of $\mathbf{m} = (m(1), m(2), \ldots, m(q)) \in \mathbb{N}^q$ being completely analogous. We represent both moments $\mathbb{E}[x_i^t(r)^m]$ and $\mathbb{E}[z_i^t(r)^m]$ using Lemma 1 (in the form given in Eqs. (3.26), (3.27)) and Lemma 3. The expectation $\mathbb{E}[x_i^t(r)^m]$ is represented as a sum over trees $T_1, \ldots, T_m \in \mathcal{U}_i^t(r) \cup \mathcal{B}_i^t(r)$, while $\mathbb{E}[z_i^t(r)^m]$ is given by a sum over trees $T_1, \ldots, T_m \in \mathcal{U}_i^t(r)$. In order to complete the proof we need to show that the contribution of terms that have at least one tree in $\mathcal{B}_i^t(r)$ vanishes as $N \to \infty$.

The factor $\prod_{\ell=1}^m \widetilde{\Gamma}(T_\ell, t, r)$ is bounded by $K(d, C, t)^m$. which is independent of N. Hence, we only need to prove that

$$\sum_{T_1 \in \mathcal{B}_i^t(r)} \sum_{T_j \in \mathcal{T}_i^t(r_j) \cup \mathcal{B}_i^t(r_j), j \in [2,m]} \mathbb{E}\left[\prod_{\ell=1}^m A(T_\ell) x(T_\ell)\right] = O\left(N^{-\frac{1}{2}}\right). \tag{3.30}$$

This statement directly follows from previous analysis, since in the graph $\mathbf{G}$ obtained by taking the union of the T_ℓ's and identifying vertices v with the same type $\ell(v)$, there is at least one edge with multiplicity 3, due to the backtracking path of length 3 or the backtracking star in T_1. So that previous analysis shows that the term in (3.30) is of order $O\left(N^{-\frac{1}{2}}\right)$. □

3.6   Proof of Theorem 3

Let $\{p_{N,i}\}_{N\geq 0, 1\leq i\leq N}$ be a collection of multivariate polynomials $p_{N,i} : \mathbb{R}^q \to \mathbb{R}$ with degrees bounded by D, and coefficients bounded in magnitude by B:

$$p_{N,i}(\mathbf{x}) = \sum_{m(1)+\cdots+m(q)\leq D} c^{N,i}_{m(1),\ldots,m(q)} x(1)^{m(1)} \cdots x(q)^{m(q)} . \tag{3.31}$$

By Propositions 6 and 8, we have,

$$|\mathbb{E}p_{N,i}(\mathbf{x}^t_i) - \mathbb{E}p_{N,i}(\tilde{\mathbf{x}}^t_i)| \leq \sum_{m(1)+\cdots+m(q)\leq D} |c^{N,i}_{m(1),\ldots,m(q)}| \, |\mathbb{E}[(\mathbf{x}^t_i)^{\mathbf{m}}] - \mathbb{E}[(\tilde{\mathbf{x}}^t_i)^{\mathbf{m}}]| \leq KD^q BN^{1/2} \tag{3.32}$$

whence the thesis follows.

3.7   Proof of Theorem 5

An important simplification is provided by the following.

Remark 2. *It is sufficient to prove Theorem 5 for $t = s$.*
   (Hence, Theorem 4 implies Theorem 5.)

Proof. Indeed consider a converging sequence $\{(A(N), \mathcal{F}_N, x^{0,N})\}_{N\geq 1}$ and fix $h = t - s > 0$. For the sake of simplicity, and in view of Remark 1 we can assume $\mathcal{F}_N$ to be given by the polynomial function $g : \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \times [k] \times \mathbb{N} \to \mathbb{R}^q$, $(\mathbf{x}, Y, a, t) \mapsto g(\mathbf{x}, Y, a, t)$ that does not depend on the random variable Y. With an abuse of notation we will write $g(\mathbf{x}, a, t)$ in place of $g(\mathbf{x}, Y, a, t)$.

   We will construct a new converging sequence of instances $\{(A(N), \widetilde{\mathcal{F}}_N, \tilde{x}^{0,N})\}_{N\geq 1}$ with variables $\tilde{\mathbf{x}}^t_i \in \mathbb{R}^{2q}$ and such that, letting $\tilde{\mathbf{x}}^t_i = (\mathbf{u}^t_i, \mathbf{v}^t_i)$, $\mathbf{u}^t_i, \mathbf{v}^t_i \in \mathbb{R}^q$, the pair $(\mathbf{u}^t_i, \mathbf{v}^t_i)$ is distributed as $(\mathbf{x}^t_i, \mathbf{x}^{t-h}_i)$ asymptotically as $N \to \infty$.

   The new sequence of initial conditions is constructed as follows

  1. The initial condition is given by $\tilde{\mathbf{x}}^0_i = (0, 0)$.

  2. The independent randomness is given by $Y(i) = \mathbf{x}^0_i$. Notice that, for $i \in C^N_a$, we have $Y(i) \sim_{i.i.d.} Q_a$ and hence we let $P_a = Q_a$.

  3. The partitions C^N_a, $a \in [k]$ and matrices $A(N)$ are kept unchanged.

  4. The collection of functions in $\widetilde{\mathcal{F}}_N$ is determined by the polynomial function $\widetilde{g} : \mathbb{R}^{2q} \times \mathbb{R}^{\tilde{q}} \times [k] \times \mathbb{N} \to \mathbb{R}^{2q}$, $(\tilde{\mathbf{x}}, Y, a, t) \mapsto \widetilde{g}(\tilde{\mathbf{x}}, Y, a, t)$. Writing $\widetilde{g}(\,\cdot\,) = [\widetilde{g}^{(1)}(\,\cdot\,), \widetilde{g}^{(2)}(\,\cdot\,)]$, with $\widetilde{g}^{(1)}(\,\cdot\,), \widetilde{g}^{(2)}(\,\cdot\,) \in \mathbb{R}^q$, we let, for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^q$.

$$g^{(1)}\big((\mathbf{u}, \mathbf{v}), Y, a, t\big) = \begin{cases} g(Y, a, t) & \text{if } t = 0, \\ g(\mathbf{u}, a, t) & \text{if } t > 0, \end{cases} \tag{3.33}$$

$$g^{(2)}\big((\mathbf{u}, \mathbf{v}), Y, a, t\big) = \begin{cases} g(Y, a, t) & \text{if } t \leq h, \\ g(\mathbf{v}, a, t) & \text{if } t > h. \end{cases} \tag{3.34}$$

As a consequence of this construction, $\mathbf{u}^t_i = \mathbf{x}^t_i$ for all $i \in [N]$, $t \geq 1$, and $\mathbf{v}^t_i = \mathbf{x}^{t-h}_i$ for all $t \geq h + 1$. This completes the reduction. $\qquad \square$

As a consequence of this remark, it is sufficient to prove Theorem 4, and by Remark 1 we can limit ourselves to the case in which $g : (\mathbf{x}, Y, a, t) \mapsto g(\mathbf{x}, Y, a, t)$ does not depend on Y and hence this argument will be dropped. We begin by considering the expectation of moments of $\mathbf{x}_i^t$.

Proposition 10. *Let $(A(N), \mathcal{F}_N, x^0)_{N \geq 0}$ be a polynomial and converging sequence of AMP instances, and denote by $\{x^t\}_{t \geq 0}$ the corresponding AMP orbit. Then we have for any $i = i(N) \in C_a^N$, $t \geq 1$, $\mathbf{m} = (m(1), \ldots, m(q)) \in \mathbb{N}^q$,*

$$\lim_{N \to \infty} \mathbb{E}\left[(\mathbf{x}_i^t)^{\mathbf{m}}\right] = \mathbb{E}\left[(Z_a^t)^{\mathbf{m}}\right],$$

where $Z_a^t \sim \mathsf{N}\left(0, \Sigma_a^t\right)$.

Proof. By Propositions 7 and 8, we need only to prove the statement for the AMP orbit y^t. We will indeed prove by induction on t that for any $i \in C_a^N$ and any $j \neq i$,

$$\lim_{N \to \infty} \mathbb{E}\left[(\mathbf{y}_{i \to j}^t)^{\mathbf{m}}\right] = \mathbb{E}\left[(Z_a^t)^{\mathbf{m}}\right], \tag{3.35}$$

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} (\mathbf{y}_{i \to j}^t)^{\mathbf{m}} = \mathbb{E}\left[(Z_a^t)^{\mathbf{m}}\right] \quad \text{in probability}. \tag{3.36}$$

For $t \geq 1$, let $\mathfrak{F}_t$ be the σ-algebra generated by $A^0, \ldots, A^{t-1}$. We will show, using the central limit theorem, that the random vector $(y_{i \to j}^{t+1}(1), \ldots, y_{i \to j}^{t+1}(q))$ given $\mathfrak{F}_t$ converge in distribution to a centered Gaussian random vector. More precisely, by (3.8) and the induction hypothesis, the following limit holds in probability,

$$
\begin{aligned}
\lim_{N \to \infty} \mathbb{E}\left[y_{i \to j}^{t+1}(r) y_{i \to j}^{t+1}(s) \big| \mathfrak{F}_t\right] &= \lim_{N \to \infty} \sum_{\substack{\ell \in [N] \setminus j \\ \ell \in C_b^N}} \mathbb{E}\left[(A_{\ell i}^t)^2\right] g_r(\mathbf{y}_{\ell \to i}^t, b, t) g_s(\mathbf{y}_{\ell \to i}^t, b, t) \\
&= \sum_{b=1}^{k} c_b W_{ab} \mathbb{E}\left[g_r(Z_b^t, b, t) g_s(Z_b^t, b, t)\right] = \Sigma_a^{t+1}(r, s).
\end{aligned}
$$

Since for all $r \in [q]$ from (3.8) we have $\mathbb{E}[y_{i \to j}^{t+1}(r)] = 0$, from the central limit theorem, it follows that $\mathbf{y}_{i \to j}^{t+1}$ converges to a centered Gaussian vector with covariance Σ_a^{t+1}. Since all the moments of $\mathbf{y}_{i \to j}^{t+1}$ are bounded uniformly in N by Proposition 7 and Lemma 2, the induction claim, Eq. (3.35) follows, for iteration $t + 1$.

In the base case $t = 0$ the same conclusion holds because

$$
\begin{aligned}
\lim_{N \to \infty} \mathbb{E}\left[y_{i \to j}^1(r) y_{i \to j}^1(s)\right] &= \lim_{N \to \infty} \sum_{\substack{\ell \in [N] \setminus j \\ \ell \in C_b^N}} \mathbb{E}\left[(A_{\ell i}^0)^2\right] g_r(\mathbf{y}_{\ell \to i}^0, b, 0) g_s(\mathbf{y}_{\ell \to i}^0, b, 0) \\
&= \sum_{b=1}^{k} c_b W_{ab} \widehat{\Sigma}_b^0(r, s),
\end{aligned}
$$

where the second identity holds by assumption.

Next consider the induction claim Eq. (3.36). Recall the representation introduced in Section 3.4:

$$y_{i \to j}^t(r) = \sum_{T \in \mathcal{T}_{i \to j}^t(r)} \overline{A}(T, t) \Gamma(T, \mathbf{c}, t) x(T),$$

$$\overline{A}(T, t) = \prod_{(u \to v) \in E(T)} A_{\ell(u)\ell(v)}^{t - |u|}.$$

Using this representation of $\mathbf{y}_{i \to j}^t$, $\mathbf{y}_{k \to j}^t$ it is easy to show that, for $i \neq k$, $i, k \in C_a^N$

$$\left| \mathbb{E}\left[(\mathbf{y}_{i \to j}^t)^{\mathbf{m}} (\mathbf{y}_{k \to j}^t)^{\mathbf{m}} \right] - \mathbb{E}\left[(\mathbf{y}_{i \to j}^t)^{\mathbf{m}} \right] \mathbb{E}\left[(\mathbf{y}_{k \to j}^t)^{\mathbf{m}} \right] \right| \leq \varepsilon(N), \qquad (3.37)$$

for some function $\varepsilon(N) \to 0$ as $N \to \infty$ ar $\mathbf{m}, C, d, t$ fixed. Indeed, the above expectations can be represented as sums over $m = m(1) + m(2) + \cdots + m(q)$ trees $T_1, \ldots, T_m \in \mathcal{T}_{i \to j}^t$ and m trees $T_1', \ldots, T_m' \in \mathcal{T}_{k \to j}^t$. Let $\mathbf{G}$ be the simple graph obtained by identifying vertices of the same type in $T_1, \ldots, T_m, T_1', \ldots, T_m'$.

By Lemma 2 and the argument in the proof of Proposition 6, all the terms in which $\mathbf{G}$ has cycles, or an edge of $\mathbf{G}$ correspond to more than 2 edges in the union of $T_1, \ldots, T_m, T_1', \ldots, T_m'$ add up to a vanishing contribution in the $N \to \infty$ limit. Further, all the terms in which $\mathbf{G}$ is the union of two disconnected components (one containing i, and the other containing k) are identical in $\mathbb{E}\left[(\mathbf{y}_{i \to j}^t)^{\mathbf{m}} (\mathbf{y}_{k \to j}^t)^{\mathbf{m}} \right]$ and $\mathbb{E}\left[(\mathbf{y}_{i \to j}^t)^{\mathbf{m}} \right] \mathbb{E}\left[(\mathbf{y}_{k \to j}^t)^{\mathbf{m}} \right]$ and hence cancel out. We are therefore left with the sum over trees $T_1, \ldots, T_m, T_1', \ldots, T_m'$ such that $\mathbf{G}$ is itself a connected tree, with edges covered exactly twice. Assume, to be definite, that $\mathbf{G}$ has μ vertices and hence $\mu - 1$ edges. The weight of such a term is bounded by

$$K \mathbb{E}\left\{ \prod_{i=1}^m \overline{A}(T_i, t) \prod_{i=1}^m \overline{A}(T_i', t) \right\} \leq K N^{-\mu + 1}$$

On the other hand, the number of such terms is bounded by $K N^{\mu - 2}$ (because the type has to be assigned to μ vertices, but 2 of these are fixed to i and k), and hence the overall contribution of these terms vanishes as well.

From Eq. (3.37) and using the fact that $\mathbb{E}[(\mathbf{y}_{i \to j}^t)^{2\mathbf{m}}] \leq K$ (because of Lemma 2 and Proposition 7), we have

$$\lim_{N \to \infty} \mathrm{Var}\left\{ \frac{1}{|C_a^N|} \sum_{i \in C_a^N} (\mathbf{y}_{i \to j}^t)^{\mathbf{m}} \right\}$$

$$\leq \lim_{N \to \infty} \frac{1}{|C_a^N|^2} \sum_{i,k \in C_a^N} \left| \mathbb{E}\left[(\mathbf{y}_{i \to j}^t)^{\mathbf{m}} (\mathbf{y}_{k \to j}^t)^{\mathbf{m}} \right] - \mathbb{E}\left[(\mathbf{y}_{i \to j}^t)^{\mathbf{m}} \right] \mathbb{E}\left[(\mathbf{y}_{k \to j}^t)^{\mathbf{m}} \right] \right| = 0.$$

Equation (3.36) follows for iteration $t + 1$ by applying Chebyshev inequality to the sequence

$$\left\{ \frac{1}{|C_a^N|} \sum_{i \in C_a^N} (\mathbf{y}_{i \to j}^t)^{\mathbf{m}} \right\}_{N \geq 0},$$

and using (3.35). $\qquad \square$

We are now ready to prove Theorem 5 in the case in which $\psi : \mathbb{R}^q \to \mathbb{R}$ is a polynomial.

Proposition 11. *Let $(A(N), \mathcal{F}_N, x^0)_{N \geq 0}$ be a polynomial and converging sequence of AMP instances, and denote by $\{x^t\}_{t \geq 0}$ the corresponding AMP orbit. Then we have for any $t \geq 1$, $\mathbf{m} = (m(1), \ldots, m(q)) \in \mathbb{N}^q$,*

$$\lim_{N \to \infty} \mathrm{Var}\Big\{ \frac{1}{|C_a^N|} \sum_{i \in C_a^N} (\mathbf{x}_i^t)^{\mathbf{m}} \Big\} = 0 \,. \tag{3.38}$$

Proof. In order to prove (3.38), we fix $t \geq 1$ and $a \in [k]$, and construct a modified sequence of AMP instances as follows. The new sequence has $N' = 2N$ and $k' = k + 1$. The new partition of the variable indices $\{1, \ldots, N\}$ is the same as in the original instances, with the addition of $C_{k+1}^N = \{N + 1, \ldots, 2N = N'\}$. Further we set, for $\varphi : \mathbb{R}^q \to \mathbb{R}$ a polynomial,

1. For $i, j \leq N$: $A'_{ij} = A_{ij}$ and when $i > N$ or $j > N$ define $A'_{ij} \sim \mathsf{N}(0, 1/N)$ independently.

2. $g'(\mathbf{x}, b, t') = g(\mathbf{x}, b, t')$ for $b \in [k]$, $t' \leq t - 1$; $g'(\mathbf{x}, b, t) = 0$ for $b \in [k] \setminus a$; $g'_1(\mathbf{x}, a, t) = \varphi(\mathbf{x})$, $g'_r(\mathbf{x}, a, t) = 0$, for $r \geq 2$; $g'(\mathbf{x}, k + 1, t') = 0$ for all t'.

   The definition of $g'(\mathbf{x}, a, t')$ for $t' > t$ is irrelevant for our purposes.

Since $g'(\mathbf{x}, k + 1, t') = 0$ for all t', the orbit $(\mathbf{x}_i^{t'} : i \leq N, t' \leq t)$ is not affected by the new variables. Further, by the general AMP equation (1.6), we have, for $i \in C_{k+1}^N$

$$x_i^{t+1}(1) = \sum_{j \in C_a^N} A_{ij} \varphi(\mathbf{x}_j^t) \,. \tag{3.39}$$

Notice that the $\{A_{ij}\}_{j \in C_a^N}$ in this equation are independent of $\mathbf{x}_j^t$. Hence

$$\mathbb{E}\{x_i^{t+1}(1)^4\} = \sum_{j_1, \ldots, j_4 \in C_a^N} \mathbb{E}\{A_{ij_1} A_{ij_2} A_{ij_3} A_{ij_4}\} \, \mathbb{E}\{\varphi(\mathbf{x}_{j_1}^t) \varphi(\mathbf{x}_{j_2}^t) \varphi(\mathbf{x}_{j_3}^t) \varphi(\mathbf{x}_{j_4}^t)\} \tag{3.40}$$

$$= \frac{3}{N^2} \sum_{j_1, j_2 \in C_a^N} \mathbb{E}\{\varphi(\mathbf{x}_{j_1}^t)^2 \varphi(\mathbf{x}_{j_2}^t)^2\} \,. \tag{3.41}$$

On the other hand, using Proposition 10 (once for iteration $t + 1$ and $i \in C_{k+1}^N$, and another time for iteration t and $i \in C_a^N$) we get

$$\lim_{N \to \infty} \mathbb{E}\{x_i^{t+1}(1)^4\} = \mathbb{E}\{(Z_{k+1}^{t+1}(1))^4\} = 3(\Sigma_{k+1}^{t+1}(1, 1))^2 = 3c_a^2 \mathbb{E}\{\varphi(Z_a^t)^2\}^2 \,, \qquad i \in C_{k+1}^N, \tag{3.42}$$

$$\lim_{N \to \infty} \mathbb{E}\{\varphi(\mathbf{x}_i^t)^2\} = \mathbb{E}\{\varphi(Z_a^t)^2\} \,, \qquad\qquad\qquad i \in C_a^N, \tag{3.43}$$

where $Z_a^t \sim \mathsf{N}(0, \Sigma_a^t)$. Comparing these equations with Eq.(3.41) we conclude that

$$\lim_{N \to \infty} \frac{1}{N^2} \sum_{j_1, j_2 \in C_a^N} \mathbb{E}\{\varphi(\mathbf{x}_{j_1}^t)^2 \varphi(\mathbf{x}_{j_2}^t)^2\} = \Big\{ \lim_{N \to \infty} \frac{1}{N} \sum_{j \in C_a^N} \mathbb{E}[\varphi(\mathbf{x}_j^t)^2] \Big\}^2 \,. \tag{3.44}$$

Equivalently

$$\lim_{N \to \infty} \text{Var}\left\{\frac{1}{|C_a^N|} \sum_{i \in C_a^N} \varphi(\mathbf{x}_i^t)^2\right\} = 0. \tag{3.45}$$

Taking $\varphi(\mathbf{x}) = \mathbf{x^k}$, we obtain Eq.(3.38) for $\mathbf{m}$ even. In order to establish Eq.(3.38) for general $\mathbf{m}$ we take, for instance, $\varphi(\mathbf{x}) = 1 + \varepsilon \mathbf{x^m}$ and use the fact that the limit must vanish for all ε. $\qquad\square$

At this point we can prove Theorem 5.

Proof of Theorem 5. By Remark 1 and Remark 2, we reduced ourselves to the case $t = s$, and $Y(i) = 0$ (equivalently, $Y(i)$, is absent).

Consider the empirical measure on $\mathbb{R}^q$ given by

$$\mu_a^N = \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \delta_{\mathbf{x}_i^t}.$$

Proposition 10 shows the convergence of expected the moments of μ_a^N to moments that determine the Gaussian distribution. Proposition 11 combined with Chebyshev inequality implies

$$\lim_{N \to \infty} \mu_a^N\big((\mathbf{x}_i^t)^{\mathbf{m}}\big) = \mathbb{E}\left[(Z_a^t)^{\mathbf{m}}\right],$$

in probability. The proof follows using the relation between convergence in probability and convergence almost sure along subsequences, together with the moment method. $\qquad\square$

4 Non-symmetric matrices

In this section we consider a slightly different setting that turns out to be a special case of the one introduced in Section 1.3.

Definition 12. *A converging sequence of (polynomial) bipartite AMP instances $\{(A(n), f, h, x^{0,n})\}_{n \geq 1}$ is defined by giving for each n:*

1. *A matrix $A(n) \in \mathbb{R}^{m \times n}$ with $m = m(n)$ such that $\lim_{n \to \infty} m(n)/n = \delta > 0$. Further, $A(n) = (A_{ij})_{i \leq m, j \leq n}$ is a matrix with the entries A_{ij} independent subgaussian random variables with common scale factor C/n and first two moments $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/m$.*

2. *Two functions $f : \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \times \mathbb{N} \to \mathbb{R}^q$, and $h : \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \times \mathbb{N} \to \mathbb{R}^q$ such that, for each $t \geq 0$, $f(\cdot, \cdot, t)$ and $h(\cdot, \cdot, t)$ are polynomials.*

3. *An initial condition $x^{0,n} = (\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0) \in \mathcal{V}_{q,n} \simeq (\mathbb{R}^q)^n$, with $\mathbf{x}_i^0 \in \mathbb{R}^q$, such that, in probability,*

$$\sum_{i=1}^n \exp\{\|\mathbf{x}_i^{0,n}\|_2^2/C\} \leq nC, \tag{4.1}$$

$$\lim_{n \to \infty} \frac{1}{m(n)} \sum_{i=1}^n f(\mathbf{x}_i^0, Y(i), 0) f(\mathbf{x}_i^0, Y(i), 0)^{\mathsf{T}} = \Xi^0. \tag{4.2}$$

25

4. *Two collections of i.i.d. random variables $(Y(i), i \in [n])$ and $(W(j), j \in [m])$ with $Y(i) \sim_{i.i.d.} Q$ and $W(j) \sim_{i.i.d.} P$. Here Q and P are finite mixture of Gaussians on $\mathbb{R}^{\tilde{q}}$.*

Throughout this section, we will refer to non-bipartite AMP instances as per Definition 5, as to *symmetric instances*. With these ingredients, we define the AMP orbit as follows.

Definition 13. *The* approximate message passing orbit *corresponding to the bipartite instance (A, f, h, x^0) is the sequence of vectors $\{x^t, z^t\}_{t \geq 0}$, $x^t \in \mathcal{V}_{q,n}$ $z^t \in \mathcal{V}_{q,m}$ defined as follows, for $t \geq 0$,*

$$z^t = A f(x^t, Y; t) - \mathsf{B}_t h(z^{t-1}, W; t-1), \tag{4.3}$$
$$x^{t+1} = A^{\mathsf{T}} h(z^t, W; t) - \mathsf{D}_t f(x^t, Y; t), \tag{4.4}$$

where $f(\cdots)$, $h(\cdots)$ are applied componentwise (see below for an explicit formulation). Here $\mathsf{B}_t : \mathcal{V}_{q,m} \to \mathcal{V}_{q,m}$ is the linear operator defined by letting, for $v' = \mathsf{B}_t v$, and any $j \in [m]$,

$$\mathbf{v}'_j = \left(\sum_{k \in [n]} A^2_{jk} \frac{\partial f}{\partial \mathbf{x}} (\mathbf{x}^t_k, Y(k); t) \right) \mathbf{v}_j. \tag{4.5}$$

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

$$\mathbf{v}'_i = \left(\sum_{l \in [m]} A^2_{li} \frac{\partial h}{\partial \mathbf{z}} (\mathbf{z}^t_l, W(l); t) \right) \mathbf{v}_i. \tag{4.6}$$

For the sake of clarity, it is useful to rewrite the iteration (4.3), (4.4) explicitly, by components:

$$\mathbf{z}^t_i = \sum_{j \in [n]} A_{ij} f(\mathbf{x}^t_j, Y(j); t) - \sum_{k \in [n]} A^2_{jk} \frac{\partial f}{\partial \mathbf{x}} (\mathbf{x}^t_k, Y(k); t) h(\mathbf{z}^{t-1}_i, W(i); t-1) \quad \text{for all } i \in [m],$$

$$\mathbf{x}^{t+1}_j = \sum_{i \in [m]} A_{ij} h(\mathbf{y}^t_i, W(i); t) - \sum_{l \in [m]} A^2_{lj} \frac{\partial h}{\partial \mathbf{z}} (\mathbf{z}^t_l, W(l); t) f(x^t_j, Y(j); t) \quad \text{for all } j \in [n].$$

We will state and prove a state evolution result that is analogous to Theorem 5 for the present case. Since the proof is by reduction to the symmetric case, the same argument also implies a universality statement of the type of Theorem 3. However, we will not state explicitly any universality statement in this case. We begin by introducing the appropriate state evolution recursion. In analogy with Eq. (1.10), we introduce two sequences of positive semidefinite matrices $\{\Sigma^t\}_{t \geq 0}$, $\{\Xi^t\}_{t \geq 0}$ by letting Ξ^0 be given as per Eq. (4.2) and defining, for all $t \geq 1$,

$$\Sigma^t = \mathbb{E}\left\{ h(Z^{t-1}, W, t-1) h(Z^{t-1}, W, t-1)^{\mathsf{T}} \right\}, \qquad Z^{t-1} \sim \mathsf{N}(0, \Xi^{t-1}), \ W \sim P, \tag{4.7}$$

$$\Xi^t = \frac{1}{\delta} \mathbb{E}\left\{ f(X^t, Y, t) f(X^t, Y, t)^{\mathsf{T}} \right\}, \qquad X^t \sim \mathsf{N}(0, \Sigma^t), \ Y \sim Q. \tag{4.8}$$

We also define a two-times recursion analogous to Eqs. (3.2), (3.3). Namely, we introduce the boundary condition

$$\Xi^{0,0} = \begin{pmatrix} \Xi^0 & \Xi^0 \\ \Xi^0 & \Xi^0 \end{pmatrix}, \qquad \Xi^{t,0} = \begin{pmatrix} \Xi^t & 0 \\ 0 & \Xi^0 \end{pmatrix}, \qquad \Xi^{0,t} = \begin{pmatrix} \Xi^0 & 0 \\ 0 & \Xi^t \end{pmatrix}, \tag{4.9}$$

26

with Ξ^t defined per Eq. (4.7), (4.8). For any $s, t \geq 1$, we set recursively

$$\Sigma^{t,s} = \mathbb{E}\left\{ \mathcal{Z}_{t-1,s-1}\, \mathcal{Z}_{t-1,s-1}^{\mathsf{T}} \right\}, \tag{4.10}$$

$$\mathcal{Z}_{t-1,s-1} \equiv [h(Z^{t-1}, W, t-1), h(Z^{s-1}, W, s-1)], \tag{4.11}$$

$$\Xi^{t,s} = \mathbb{E}\left\{ \mathcal{X}_{t,s}\, \mathcal{X}_{t,s}^{\mathsf{T}} \right\}, \tag{4.12}$$

$$\mathcal{X}_{t,s} \equiv [f(X^t, Y, t), f(X^s, Y, s)]. \tag{4.13}$$

(Recall that $[u, v]$ denotes the column vector obtained by concatenating u and v.)

Theorem 6. *Let $\{(A(n), f, h, x^{0,n})\}_{n \geq 1}$ be a polynomial and converging sequence of* bipartite AMP *instances, and denote by $\{x^t, z^t\}_{t \geq 0}$ the corresponding AMP orbit.*

Fix $s, t \geq 1$. If $s \neq t$, further assume that the initial condition $x^{0,n}$ is obtained by letting $\mathbf{x}_i^{0,n} \sim R$ independent and identically distributed, with R a finite mixture of Gaussians. Then, for each locally Lipschitz function $\psi : \mathbb{R}^q \times \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \to \mathbb{R}$ such that $|\psi(\mathbf{x}, \mathbf{x}', y)| \leq K(1 + \|y\|_2^2 + \|\mathbf{x}\|_2^2 + \|\mathbf{x}'\|_2^2)^K$, we have, in probability,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j \in [n]} \psi(\mathbf{x}_j^t, \mathbf{x}_j^s, Y(j)) = \mathbb{E}\left[\psi(X^t, X^s, Y) \right], \tag{4.14}$$

$$\lim_{N \to \infty} \frac{1}{m(n)} \sum_{j \in [m]} \psi(\mathbf{z}_j^t, \mathbf{z}_j^s, W(j)) = \mathbb{E}\left[\psi(Z^t, Z^s, W) \right], \tag{4.15}$$

where $(X^t, X^s) \sim \mathsf{N}(0, \Sigma^{t,s})$ is independent of $Y \sim Q$, and $(Z^t, Z^s) \sim \mathsf{N}(0, \Xi^{t,s})$ is independent of $W \sim P$.

Proof. The proof follows by constructing a suitable polynomial and converging sequence of symmetric instances, recognizing that a suitable subset of the resulting orbit corresponds to the orbit $\{x^t, z^t\}$ of interest, and applying Theorem 5.

Specifically, given a converging sequence of bipartite instances $(A(n), f, h, x^{0,n})$, we construct a symmetric instance $(A_s(N), g, x_s^{0,N})$ with (below we use the subscript s to refer to the symmetric instance):

1. The symmetric instance has dimensions $N = n + m$ and $q_s = q$, $\tilde{q}_s = \tilde{q}$.

2. We partition the index set in $k = 2$ subsets: $[N] = C_1^N \cup C_2^N$, with $C_1^N = \{1, \ldots, m\}$ and $C_2^N = \{m + 1, \ldots, m + n\}$. In particular $c_1 = \delta/(1 + \delta)$ and $c_2 = 1/(1 + \delta)$.

3. The symmetric random matrix A' is given by

$$A_s = \begin{pmatrix} 0 & A \\ A^{\mathsf{T}} & 0 \end{pmatrix}.$$

   In particular $W_{11} = W_{22} = 0$ and $W_{12} = W_{21} = (1 + \delta)/\delta$.

4. The vertex labels are $Y_s(i) = W(i)$ for $i \leq m$ and $Y_s(i) = Y(i - m)$ for $i > m$. In particular, these are independent random variables with distribution $Y_s(i) \sim P_1 = Q$ if $i \in C_1^N$ and $Y_s(i) \sim P_2 = P$ if $i \in C_2^N$.

27

5. The initial condition is given by $\mathbf{x}_{s,i}^{0,N} = 0$ for $i \in C_1^N$ and $\mathbf{x}_{s,i}^{0,N} = \mathbf{x}_{i-m}^{0,n}$ for $i \in C_2^N$.

6. Finally, for any $\mathbf{x} \in \mathbb{R}^q$, $Y \in \mathbb{R}^{\tilde{q}}$, $t \geq 0$, we let

$$
\begin{aligned}
g(\mathbf{x}, Y, a = 1, 2t) &= f(\mathbf{x}, Y, t), & (4.16) \\
g(\mathbf{x}, Y, a = 2, 2t+1) &= h(\mathbf{x}, Y, t), . & (4.17)
\end{aligned}
$$

The definition of $g(\mathbf{x}, Y, a = 1, 2t+1)$ and $g(\mathbf{x}, Y, a = 2, 2t)$ is irrelevant for our purposes.

The proof is concluded by recognizing that, for all $t \geq 0$,

$$
\begin{aligned}
\mathbf{x}_{s,i}^{2t+1} &= \mathbf{z}_i^t, & \text{for } i \in C_1^N, \\
\mathbf{x}_{s,i}^{2t} &= \mathbf{x}_{i-m}^t, & \text{for } i \in C_2^N,
\end{aligned}
$$

$\square$

We finish this section with a lemma that establishes continuity of the AMP trajectories with respect to Gaussian perturbations of the matrix A. This fact will be used in the next section. (Notice that an analogous Lemma holds by the same argument for converging, non-bipartite, instances.)

Lemma 4. *Let $\{(A(n), f, h, x^{0,n})\}_{n \geq 1}$ be a polynomial converging sequence of bipartite AMP instances and denote by $\{x^t, z^t\}_{t \geq 0}$ the corresponding AMP orbit. For each n, let $G(n) \in \mathbb{R}^{m(n) \times n}$ be a random matrix with i.i.d. entries $G(n)_{ij} \sim \mathsf{N}(0, 1/m(n))$, independent of $A(n)$. Consider the perturbed sequence $\{(\tilde{A}(n) = A(n) + \nu\, G(n), f, h, x^{0,n})\}_{n \geq 1}$, with $\nu \in \mathbb{R}^+$ and denote by $\{\tilde{x}^t, \tilde{z}^t\}_{t \geq 0}$ the corresponding AMP orbit. Then for any t there exists a constant K independent of n such that*

$$
\mathbb{E}\{\|\mathbf{x}_i^t - \tilde{\mathbf{x}}_i^t\|_2^2\} \leq K\left(\nu^2 + n^{-1/2}\right), \quad \mathbb{E}\{\|\mathbf{z}_i^t - \tilde{\mathbf{z}}_i^t\|_2^2\} \leq K\left(\nu^2 + n^{-1/2}\right).
$$

Proof. Consider the difference $[\mathbf{x}_i^t(r) - \tilde{\mathbf{x}}_i^t(r)]$. By the tree representation in Section 3.2 and Lemma 3, this difference can be written as a polynomial in A and G whereby each monomial has the form

$$
\Gamma(T, t)x(T)\left\{ \prod_{(u \to v) \in E(T)} \tilde{A}_{\ell(u)\ell(v)} - \prod_{(u \to v) \in E(T)} A_{\ell(u)\ell(v)} \right\}. \tag{4.18}
$$

Enumerating the edges in T as $(u_1, v_1), \ldots, (u_k, v_k)$ the quantity in parenthesis reads

$$
\sum_{i=1}^{k} \prod_{j=1}^{i-1} A_{\ell(u_j),\ell(v_j)} \cdot \nu\, G_{\ell(u_i),\ell(v_i)} \cdot \prod_{j=i+1}^{k} \tilde{A}_{\ell(u_j),\ell(v_j)}. \tag{4.19}
$$

In other words, the sum over trees T is replaced by a sum over trees with one distinguished edge, and the edge carries weight $\nu\, G_{\ell(u_i),\ell(v_i)}$. The expectation $\mathbb{E}\{\|\mathbf{x}_i^t - \tilde{\mathbf{x}}_i^t\|_2^2\}$ is given by a sum over pairs of such marked trees. Using the fact that the entries of the matrix $\tilde{A}(n)$ are still independent subgaussian with scale factor $C/(n + \nu^2 C m(n)) \leq C'/n$, it is easy to see that the argument in Lemma 2 and (3.30) are still valid. Hence, up to errors bounded by $K\, n^{-1/2}$ the only terms that contribute to this sum are those over pair of trees such that the graph $\mathbf{G}$ obtained by identifying vertices of the same type has only double edges. In particular for the distinguished edge, we can use the following upper bound instead of (3.15): $\mathbb{E}\left[|\nu G_{ij}|^2\right] = \frac{\nu^2}{m(n)} \leq K\frac{\nu^2}{n}$ and this yields a factor ν^2 (by the same argument as in the proof of Lemma 2 to get (3.20)). $\square$

5 Proof of universality of polytope neighborliness

In this section we prove Theorem 2, deferring several technical steps to the Appendix.

Hypothesis 1 Throughout this section $\{A(n)\}_{n\geq 0}$ is a sequence of random matrices whereby $A(n) \in \mathbb{R}^{m\times n}$ has independent entries that satisfy $\mathbb{E}\{A(n)_{ij}\} = 0$, $\mathbb{E}\{A(n)_{ij}^2\} = 1/m$ and are subgaussian with scale factor s/m, with s independent of m, n.

Notice that these matrices differ by a factor $1/\sqrt{m}$ from the matrices in the statement of Theorem 2. Since neighborliness is invariant under scale transformations, this change is immaterial.

The approach we will follow is based on the equivalence between weak neighborliness and compressed sensing reconstruction developed in [Don05b, Don05a, DT05b, DT05a]. Within compressed sensing, one considers the problem of reconstructing a vector $x_0 \in \mathbb{R}^n$ from a vector of linear 'observations' $y = Ax_0$ with $y \in \mathbb{R}^m$ and $m \leq n$. The measurement matrix $A \in \mathbb{R}^{m\times n}$ is assumed to be known. An interesting approach towards reconstructing x_0 from the linear observations y consists in solving a convex program:

$$\widehat{x}(y) = \arg\min \left\{ \|x\|_1 \text{ such that } x \in \mathbb{R}^n, \ y = Ax, \right\}. \tag{5.1}$$

Hence one says that ℓ_1 minimization *succeeds* if the above arg min is uniquely defined and $\widehat{x}(y) = x_0$. Remarkably, this event only depends on the support of x_0, $\text{supp}(x_0) = \{i \in [n] : x_{0,i} \neq 0\}$ [Don05b]. This motivates the following abuse of terminology. We say that, for a given matrix A, ℓ_1 minimization *succeeds* for a fraction f of vectors x_0 with[3] $\|x_0\|_0 \leq k$ if it does succeed for at least $f\binom{n}{k}$ choices of $\text{supp}(x_0)$ out of the $\binom{n}{k}$ possible ones. Analogously, that ℓ_1 minimization *fails* for a fraction f of vectors x_0 if it does succeed at most for $(1-f)\binom{n}{k}$ choices of $\text{supp}(x_0)$.

Success of ℓ_1 minimization turns out to be intimately related to the neighborliness properties of the polytope AC^n.

Theorem 7 (Donoho, 2005). *Fix $\delta \in (0,1)$. For each $n \in \mathbb{N}$, let $m(n) = \lfloor n\delta \rfloor$ and $A(n) \in \mathbb{R}^{m(n)\times n}$ be a random matrix. Then the sequence $\{A(n)C^n\}_{n\geq 0}$ has weak neighborliness ρ in probability if and only if the following happens:*

1. *For any $\rho_- < \rho$, there exists $\varepsilon_n \downarrow 0$ such that, for a fraction larger than $(1 - \varepsilon_n)$ of vectors x_0 with $\|x_0\|_0 = m(n)\,\rho_-$ the ℓ_1 minimization* succeeds *with high probability (with respect to the choice of the random matrix $A(n)$).*

2. *Viceversa, for any $\rho_+ > \rho$, there exists $\varepsilon_n \downarrow 0$ such that, for a fraction larger than $(1 - \varepsilon_n)$ of vectors x_0 with $\|x_0\|_0 = m(n)\,\rho_+$ the ℓ_1 minimization* fails *with high probability (with respect to the choice of the random matrix $A(n)$).*

This is indeed a rephrasing of Theorem 2 in [Don05b].

In view of this result, Theorem 2 follows from the following result on compressed sensing with random sensing matrices.

---

[3]As customary in this domain, we denote by $\|v\|_0$ the number of non-zero entries in $v \in \mathbb{R}^q$ (which of course is not a norm).

Theorem 8. *Fix $\delta \in (0,1)$. For each $n \in \mathbb{N}$, let $m(n) = \lfloor n\delta \rfloor$ and define $A(n) \in \mathbb{R}^{m(n) \times n}$ to be a random matrix with independent subgaussian entries, with mean 0, variance $1/m$, and common scale factor s/m. Further assume $A_{ij}(n) = \tilde{A}_{ij}(n) + \nu_0 \, G_{ij}(n)$ where $\nu_0 > 0$ is independent of n and $\{G_{ij}(n)\}_{i\in[m],j\in[n]}$ is a collection of i.i.d. $\mathsf{N}(0,1/m)$ random variables independent of $\tilde{A}(n)$.*

Consider either of the following two cases:

1. *The matrix $A(n)$ has i.i.d. entries and $\{x_0(n)\}_{n\geq 1}$ is any fixed sequence of vectors with $\lim_{n\to\infty} \|x_0(n)\|_0/m(n) = \rho$.*

2. *The matrix $A(n)$ has independent but not identically distributed entries. The vectors $x_0(n)$ have i.i.d. entries independent of $A(n)$, with $\mathbb{P}\{x_{0,i}(n) \neq 0\} = \rho\delta$.*

Then the following holds. If $\rho < \rho_*(\delta)$ then ℓ_1 minimization succeeds with high probability. Viceversa, if $\rho > \rho_*(\delta)$, then ℓ_1 minimization fails with high probability. (Here probability is with respect to the realization of the random matrix $A(n)$ and, eventually, $x_0(n)$.)

The rest of this section is devoted to the proof of Theorem 8. Indeed, as shown below, this immediately implies Theorem 2.

Proof of Theorem 2. Take $x_0(n)$ to be a sequence of independent vectors with independent entries such that $\mathbb{P}_\rho\{x_0(n)_i = 1\} = \rho\delta$ and $\mathbb{P}_\rho\{x_0(n)_i = 0\} = 1 - \rho\delta$. Then, by the law of large numbers we have $\lim_{n\to\infty} \|x_0(n)\|_0/m(n) = \rho$ almost surely. Let $A(n) \in \mathbb{R}^{m(n)\times n}$ be a matrix with i.i.d. entries as per Hypothesis 1 above, with $m(n) = \lfloor n\delta \rfloor$ and $y(n) = A(n)x_0(n)$. Applying Theorem 8, we have, for any $\rho_- < \rho_*(\delta)$ and $\rho_+ > \rho_*(\delta)$

$$\lim_{n\to\infty} \mathbb{P}_{\rho_-}\{\widehat{x}(y(n)) = x_0(n)\} = 1\,, \tag{5.2}$$

$$\lim_{n\to\infty} \mathbb{P}_{\rho_+}\{\widehat{x}(y(n)) = x_0(n)\} = 0\,, \tag{5.3}$$

where $\mathbb{P}_{\rho_\pm}\{\,\cdot\,\}$ denotes probability with respect to the law just described when $\rho = \rho_\pm$. Let $V(\rho; m, n)$ be the fraction of vectors x_0 with $\|x_0\| = \lfloor m\rho \rfloor$ on which ℓ_1 reconstruction succeeds. Since in Eqs. (5.2), (5.3), support of $x_0(n)$ is uniformly random given its size, and the probability of success is monotone decreasing in the support size [Don05b], the above equations imply

$$\lim_{n\to\infty} \mathbb{E}\{V(\rho_-; m, n)\} = 1\,, \tag{5.4}$$

$$\lim_{n\to\infty} \mathbb{E}\{V(\rho_+; m, n)\} = 0\,, \tag{5.5}$$

Using Markov inequality, Eqs. (5.4), (5.5) coincide (respectively) with assumptions 1 and 2 in Theorem 7. The claim follows by applying this theorem. $\square$

Let us now turn to the proof of Theorem 8. The following Lemma provides a useful sufficient condition for successful reconstruction. Here and below, for a convex function $F : \mathbb{R}^q \to \mathbb{R}$, $\partial F(x)$ denotes the subgradient of F at $x \in \mathbb{R}^q$. In particular $\partial \|x\|_1$ denotes the subgradient of the ℓ_1 norm at x. Further, for $R \subseteq [n]$, A_R denotes the submatrix of A formed by columns with index in R. The singular values of a matrix $M \in \mathbb{R}^{d_1 \times d_2}$ are denoted by $\sigma_{\max}(M) \equiv \sigma_1(M) \geq \sigma_2(M) \geq \cdots \geq \sigma_{\min(d_1,d_2)}(M) \equiv \sigma_{\min}(M)$.

Lemma 5. *For any $c_1, c_2, c_3 > 0$, there exists $\varepsilon_0(c_1, c_2, c_3) > 0$ such that the following happens. If $x_0 \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $y = Ax_0 \in \mathbb{R}^m$, are such that*

1. *There exists $v \in \partial\|x_0\|_1$ and $z \in \mathbb{R}^m$ with $v = A^{\mathsf{T}}z + w$ and $\|w\|_2 \leq \sqrt{n}\,\varepsilon$, with $\varepsilon \leq \varepsilon_0(c_1, c_2, c_3)$.*

2. *For $c \in (0,1)$, let $S(c) \equiv \{i \in [n] : |v_i| \geq 1 - c\}$. Then, for any $S' \subseteq [n]$, $|S'| \leq c_1 n$, the minimum singular value of $A_{S(c_1) \cup S'}$ satisfies $\sigma_{\min}(A_{S(c_1) \cup S'}) \geq c_2$.*

3. *The maximum singular value of A satisfies $c_3^{-1} \leq \sigma_{\max}(A)^2 \leq c_3$.*

Then x_0 is the unique minimizer of $\|x\|_1$ over $x \in \mathbb{R}^n$ such that $y = Ax$.

The proof of this lemma is deferred to Appendix B.

The proof of Theorem 8 consists in two parts. For $\rho > \rho_*(\delta)$, we shall exhibit a vector x with $\|x\|_1 < \|x_0\|_1$ and $y = Ax$. For $\rho < \rho_*(\delta)$ we will show that assumptions of Lemma 5 hold. In particular, we will construct a subgradient v as per assumption 1. In both tasks, we will use an iterative message passing algorithm analogous to the one in Section 4. The algorithm is defined by the following recursion initialized with $x^0 = 0$:

$$x^{t+1} = \eta(x^t + A^{\mathsf{T}}z^t; \alpha\sigma_t), \tag{5.6}$$

$$z^t = y - Ax^t + \mathsf{b}_t\, z^{t-1}, \tag{5.7}$$

where $\eta(u; \theta) = \text{sign}(u)\,(u - \theta)_+$, α is a non-negative constant, and b_t is a diagonal matrix whose precise definition is immaterial here and will be given in the proof of Proposition 14 below. Notice two important differences with respect to the treatment in Section 4:

- The iteration in Eqs. (5.6), (5.7) does not take immediately the form in Eqs. (4.3), (4.4). For instance the nonlinear mapping $\eta(\,\cdot\,; \alpha\sigma_t)$ is applied *after* multiplication by A^{T}. This mismatch can be resolved by a simple change of variables.

- The nonlinear mapping $\eta(\,\cdot\,; \alpha\sigma_t)$ is not a polynomial. This point will be addressed by constructing suitable *polynomial approximations* of η.

We refer to Appendix A for further details.

For $t \geq 0$, σ_t is defined by the one-dimensional recursion

$$\sigma_{t+1}^2 = \frac{1}{\delta}\mathbb{E}\{[\eta(X + \sigma_t\,Z; \alpha\sigma_t) - X]^2\}, \tag{5.8}$$

where expectation is with respect to the independent random variables $Z \sim \mathsf{N}(0,1)$, $X \sim p_X$, and $\sigma_0^2 = \mathbb{E}\{X^2\}/\delta$.

Proposition 14. *Let $\{(x_0(n), A(n), y(n))\}_{n \geq 0}$ be a sequence of triples with $A(n)$ random as per Hypothesis 1, $\{x_{0,i}(n) : i \in [n]\}$ independent and identically distributed with $x_{0,i}(n) \sim p_X$ a finite mixture of Gaussians on $\mathbb{R}$, and $y(n) = A(n)x_0(n)$.*

Then, for each n there exist a sequence of vectors $\{x^t(n), z^t(n)\}_{t \geq 0}$, with $x^t(n) = x^t \in \mathbb{R}^n$, $z^t(n) = z^t \in \mathbb{R}^m$, such that the following happens for every t.

1. *There exists a diagonal matrix $\mathsf{b}_t = \mathsf{b}_t(n)$ such that*

$$z^t = y - Ax^t + \mathsf{b}_t\, z^{t-1}, \tag{5.9}$$

$$\lim_{n \to \infty} \max_{i \in [m]} (\mathsf{b}_t)_{ii} = \lim_{n \to \infty} \min_{i \in [m]} (\mathsf{b}_t)_{ii} = \frac{1}{\delta}\mathbb{P}\{|X + \sigma_{t-1}Z| \geq \alpha\sigma_{t-1}\}. \tag{5.10}$$

where the limit holds in probability.

2. *In probability*

$$\lim_{n\to\infty} \frac{1}{n} \|x^{t+1} - \eta(x^t + A^\mathsf{T} z^t; \alpha\sigma_t)\|_2^2 = 0\,. \tag{5.11}$$

3. *For any locally Lipschitz function $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $|\psi(x,y)| \le C(1 + x^2 + y^2)$, in probability*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \psi(x_{0,i}, x_i^t + (A^\mathsf{T} z^t)_i) = \mathbb{E}\,\psi(X, X + \sigma_t Z)\,. \tag{5.12}$$

4. *There exist a two functions $o(a;c)$ and $o(a,b;c)$, with $o(a;c) \to 0$, $o(a,b;c) \to 0$ as $c \to 0$ at a,b fixed, such that the following holds. Assume $A_{ij}(n) = \tilde{A}_{ij}(n) + \nu\, G_{ij}(n)$ where $\nu > 0$ is independent of n and $\{G_{ij}(n)\}_{i\in[m],j\in[n]}$ is a collection of i.i.d. $\mathsf{N}(0, 1/m)$ random variables independent of $\tilde{A}(n)$. Then there exists a sequence of vectors $\{\tilde{x}^t, \tilde{z}^t\}_{t\ge0}$ that is independent of G such that, for any $t \ge 0$,*

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\big\{\big((x^t + A^\mathsf{T} z^t)_i - (\tilde{x}^t + \tilde{A}^\mathsf{T}\tilde{z}^t)_i\big)^2\big\} \le o(t;\nu) + o(t,\nu;n^{-1})\,, \tag{5.13}$$

$$\frac{1}{m} \sum_{i=1}^{m} \mathbb{E}\big\{\big(z_i^t - \tilde{z}_i^t\big)^2\big\} \le o(t;\nu) + o(t,\nu;n^{-1})\,. \tag{5.14}$$

The proof is deferred to Appendix A.

We also need a generalization of the last proposition for functions of the estimates x^t, x^s at two distinct iteration numbers $t \ne s$. To this objective, we introduce the generalization of the state evolution equation (5.8). Namely, we define $\{R_{s,t}\}_{s,t\ge0}$ recursively for all $s, t \ge 0$ by letting

$$R_{s+1,t+1} = \frac{1}{\delta} \mathbb{E}\big\{[\eta(X + Z_s; \alpha\sigma_s) - X][\eta(X + Z_t; \alpha\sigma_t) - X]\big\}\,. \tag{5.15}$$

Here the expectation is with respect to $X \sim p_X$ and the independent Gaussian vector $[Z_s, Z_t]$ with zero mean and covariance given by $\mathbb{E}\{Z_s^2\} = R_{s,s}$, $\mathbb{E}\{Z_t^2\} = R_{t,t}$ and $\mathbb{E}\{Z_t Z_s\} = R_{t,s}$. The boundary condition is fixed by letting $R_{0,0} = \mathbb{E}\{X^2\}/\delta$ and defining, for each $t \ge 0$,

$$R_{0,t+1} = \frac{1}{\delta} \mathbb{E}\big\{[\eta(X + Z_t; \alpha\sigma_t) - X][-X]\big\}\,, \tag{5.16}$$

with $Z_t \sim \mathsf{N}(0, R_{t,t})$. This uniquely determine the doubly infinite array $\{R_{t,s}\}_{t,s\ge0}$. Notice in particular that $R_{t,t} = \sigma_t^2$ for all $t \ge 0$. (This is easily checked by induction over t).

Proposition 15. *Under the assumptions of Proposition 14 the sequence $\{x^t(n), z^t(n)\}_{t\ge0}$ constructed there further satisfies the following. For any fixed $t, s \ge 0$, and any Lipschitz continuous functions $\psi : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, in probability*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \psi\Big(x_{0,i}, x_i^s + (A^\mathsf{T} z^s)_i, x_i^t + (A^\mathsf{T} z^t)_i\Big) = \mathbb{E}\,\psi(X, X + Z_s, X + Z_t)\,, \tag{5.17}$$

$$\lim_{n\to\infty} \frac{1}{m} \sum_{i=1}^{n} \phi(z_i^s, z_i^t) = \mathbb{E}\,\phi(Z_s, Z_t)\,, \tag{5.18}$$

where expectation is with respect to $X \sim p_X$ and the independent Gaussian vector (Z_s, Z_t) with zero mean and covariance given by $\mathbb{E}\{Z_s^2\} = R_{s,s}$, $\mathbb{E}\{Z_t^2\} = R_{t,t}$ and $\mathbb{E}\{Z_t Z_s\} = R_{t,s}$.

The proof of this proposition is in Appendix A.

Finally, we need some analytical estimates on the recursions (5.8) and (5.15). Part of these estimates were already proved in [DMM09, DMM11, BM12], but we reproduce them here for the reader's convenience. Proofs of the others are provided in Appendix C.

Lemma 6. *Let p_X be a probability measure on the real line such that $p_X(\{0\}) = 1 - \varepsilon$ and $\mathbb{E}_{p_X}\{X^2\} < \infty$, fix $\delta \in (0,1)$ and set $\rho = \delta\varepsilon$. For this choice of parameters, consider the sequences $\{\sigma_t^2\}_{t\geq 0}$, $\{R_{s,t}\}_{s,t\geq 0}$ defined as per Eqs. (5.8), (5.15).*

If $\rho < \rho_*(\delta)$ then

(a1) *There exists $\alpha_1(\varepsilon,\delta)$, $\alpha_2(\varepsilon,\delta)$, $\alpha_*(\varepsilon)$ with $0 < \alpha_1(\varepsilon,\delta) < \alpha_*(\varepsilon) < \alpha_2(\varepsilon,\delta) < \infty$, and $\omega_*(\varepsilon,\delta) \in (0,1)$ such that the following happens. For each $\alpha \in (\alpha_1, \alpha_2)$, $\sigma_t^2 = B\,\omega^t(1 + o_t(1))$ as $t \to \infty$, with $\omega \in (0,1)$.*

Further, for each $\omega \in [\omega_*(\varepsilon,\delta), 1)$ there exists $\alpha_- \in (\alpha_1, \alpha_*]$ and $\alpha_+ \in [\alpha_*, \alpha_2)$ (distinct as long as $\omega > \omega_*$) such that, letting $\alpha \in \{\alpha_-, \alpha_+\}$, $\sigma_t^2 = B\,\omega^t(1 + o_t(1))$.

Finally, for all $\alpha \in [\alpha_*, \alpha_2)$, we have $\varepsilon + 2(1-\varepsilon)\Phi(-\alpha) < \delta$.

(a2) *For any $\alpha \in [\alpha_*(\varepsilon), \alpha_2(\varepsilon,\delta))$, we have $\lim_{t\to\infty} R_{t,t-1}/(\sigma_t\sigma_{t-1}) = 1$.*

(a3) *Assume p_X to be such that $\max(p_X((0,a)), p_X((-a,0))) \leq Ba^b$ for some $B, b > 0$ (in particular this is the case if p_X has an atom at 0 and is absolutely continuous in a neighborhood of 0). Fixing again $\alpha \in [\alpha_*(\varepsilon), \alpha_2(\varepsilon,\delta))$, and $c \in \mathbb{R}_+$,*

$$\lim_{t_0 \to \infty} \sup_{t,s \geq t_0} \mathbb{P}\{|X + Z_s| \geq c\,\sigma_s\,;\ |X + Z_t| < c\,\sigma_t\} = 0\,, \tag{5.19}$$

where (Z_s, Z_t) is a gaussian vector with $\mathbb{E}\{Z_s^2\} = \sigma_s^2$, $\mathbb{E}\{Z_s^2\} = \sigma_s^2$, $\mathbb{E}\{Z_s Z_t\} = R_{s,t}$.

Viceversa, if $\rho > \rho_*(\delta)$, then there exists $\alpha_0(\delta, p_X) > \alpha_{\min}(\delta) > 0$ such that

(b1) *For any $\alpha > \alpha_{\min}(\delta)$, we have $\lim_{t\to\infty} \sigma_t^2 = \sigma_*^2 > 0$ and, for $\alpha \geq \alpha_0$, $\lim_{t\to\infty}[R_{t,t} - 2R_{t,t-1} + R_{t-1,t-1}] = 0$.*

(b2) *Letting $\alpha = \alpha_0(\delta, p_X)$, we have $\mathbb{P}\{|X + \sigma_* Z| \geq \alpha\sigma_*\} = \delta$.*

(b3) *Consider the probability distribution $p_X = (1-\varepsilon)\delta_0 + \varepsilon\gamma$ with $\gamma(\mathrm{d}x) = \exp(-x^2/2)/\sqrt{2\pi}\,\mathrm{d}x$ the standard Gaussian measure. Then, setting $\alpha = \alpha_0(\delta, p_X)$, we have $\lim_{t\to\infty} \mathbb{E}\{|\eta(X + \sigma_t Z; \alpha\sigma_t)|\} < \mathbb{E}\{|X|\}$, where $Z \sim \mathsf{N}(0,1)$ independent of X.*

We are now in position to prove Theorem 8. For greater convenience of the reader, we distinguish the cases $\rho < \rho_*(\delta)$ and $\rho > \rho_*(\delta)$. Before considering these cases, we will establish some common simplifications.

5.1 Proof of Theorem 8, common simplifications

Consider first case 1. By exchangeability of the columns of $A(n)$, it is sufficient to prove the claim for the sequence of random vectors obtained by permuting the entries of $x_0(n)$ uniformly at random. Hence $x_0(n)$ is a vector with a uniformly random support $\mathrm{supp}(x_0(n)) = S_n$, with deterministic size $|S_n|$ such that $|S_n|/n \to \varepsilon$. Further, the success of ℓ_1 minimization is an event that is monotone

decreasing in the support $\mathrm{supp}(x_0(n))$ [Don05b]. Therefore we can replace the deterministic support size, with a random size $|S_n| \sim \mathrm{Binom}(n, \varepsilon)$ (which concentrates tightly around $n\varepsilon$).

Finally, since success of ℓ_1 minimization only depends on the support of $x_0(n)$ [Don05b], we can replace the non-zero entries by arbitrary values. We will take advantage of this fact and assume that all the non-zero entries of $x_0(n)$ are i.i.d. $\mathsf{N}(0, 1)$. We conclude that it is sufficient to prove that ℓ_1 minimization succeeds/fails with high probability if the vectors $x_0(n)$ have i.i.d. entries with distribution $p_X = (1 - \varepsilon)\delta_0 + \varepsilon\gamma$, where $\gamma(\mathrm{d}x) = \exp(-x^2/2)/\sqrt{2\pi}\,\mathrm{d}x$.

Consider next case 2, in which the entries of $x_0(n)$ are i.i.d. with $\mathbb{P}\{x_{0,i}(n) \neq 0\} = \rho\delta = \varepsilon$. Again, exploiting the fact that the success of ℓ_1 minimization depends only on the support of $x_0(n)$, we can assume that its entries have common distribution $p_X = (1 - \varepsilon)\delta_0 + \varepsilon\gamma$.

Summarizing this discussion, in order to prove the Theorem both in case 1 and case 2, it will be sufficient to do so for the following setting

Remark 3. *In the proof of Theorem 8, we can assume the vectors $x_0(n)$ to be random with i.i.d. entries with common distribution $p_X = (1 - \varepsilon)\delta_0 + \varepsilon\gamma$, and the matrices $A(n)$.*

5.2 Proof of Theorem 8, $\rho < \rho_*(\delta)$

Fix $\rho < \rho_*(\delta)$. We will prove that the hypotheses 1, 2, 3 of Lemma 5 hold with high probability for fixed $c_1, c_2, c_3 > 0$, and ε arbitrarily small. This implies the claim (i.e. that ℓ_1 minimization succeeds) by applying the Lemma. Notice that hypothesis 3 holds with high probability for some $c_3 = c_3(\delta)$ by classical estimates on the extreme eigenvalues of sample covariance matrices [BS98, BS05].

We next consider hypothesis 1 of Lemma 5. In order to construct the subgradient v used there, we consider the sequence of vectors $\{x^t, z^t\}_{t \geq 0}$ defined by as per Proposition 14. We fix $\alpha \in (\alpha_1(\varepsilon), \alpha_2(\varepsilon))$ as per Lemma 6.(a) so that $\sigma_t^2 = A\omega^t(1 + o(1))$ with $\omega \in (0, 1)$ to be chosen close enough to 1. Also, we introduce the notation $\theta_t \equiv \alpha\sigma_t$. We let $v^t \in \mathbb{R}^n$ be defined by

$$v_i^t = \begin{cases} \mathrm{sign}(x_{0,i}) & \text{if } i \in S, \\ \frac{1}{\theta_{t-1}}\big(x^{t-1} + A^\mathsf{T}z^{t-1} - \widehat{x}^t\big)_i & \text{otherwise,} \end{cases} \tag{5.20}$$

$$\widehat{x}^t \equiv \eta(x^{t-1} + A^\mathsf{T}z^{t-1}; \theta_{t-1}). \tag{5.21}$$

Notice that, by definition of the function $\eta(\,\cdot\,;\,\cdot\,)$ we have $|x_i^{t-1} - (A^\mathsf{T}z^{t-1})_i - \widehat{x}_i^t| \leq \theta_{t-1}$, and hence $v^t \in \partial\|x_0\|_1$. We can write

$$v^t = \frac{1}{\theta_{t-1}}A^\mathsf{T}z^t + \xi^t + \beta^t + \zeta^t, \tag{5.22}$$

$$\xi^t \equiv \frac{1}{\theta_{t-1}}\big(x^{t-1} + A^\mathsf{T}z^{t-1} - x^t - A^\mathsf{T}z^t\big), \tag{5.23}$$

$$\beta^t \equiv \frac{1}{\theta_{t-1}}\big(x^t - \widehat{x}^t\big), \tag{5.24}$$

$$\zeta^t \equiv \begin{cases} \mathrm{sign}(x_{0,i}) - \frac{1}{\theta_{t-1}}\big(x^{t-1} + A^\mathsf{T}z^{t-1} - \widehat{x}^t\big)_i & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases} \tag{5.25}$$

This part of the proof is completed by showing that there exists $h(t)$ with $\lim_{t \to \infty} h(t) = 0$ such that, for each t, with high probability we have $\|\xi^t\|_2^2/n \leq (1 - \sqrt{\omega})^2/\alpha^2 + h(t)$, $\|\beta^t\|_2^2/n \leq h(t)$, and

34

$\|\zeta^t\|_2^2/n \leq h(t)$. Indeed, if this is true, we can then choose t sufficiently large and $\alpha \in (\alpha_*(\varepsilon), \alpha_2(\varepsilon, \delta))$ so that $\|\xi^t + \beta^t + \zeta^t\|_2^2$ is small enough as to satisfy the condition 1 of Lemma 5.

First consider ξ^t. Applying Proposition 15 to $\psi(x, y_1, y_2) = (y_1 - y_2)^2$, we have, in probability

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\|\xi^t\|_2^2 &= \lim_{n\to\infty} \frac{1}{n\alpha^2\sigma_{t-1}^2} \|x^t + A^\mathsf{T}z^t - x^{t-1} - A^\mathsf{T}z^{t-1}\|_2^2 \\
&= \frac{1}{\alpha^2\sigma_{t-1}^2} \left[R_{t,t} - 2R_{t,t-1} + R_{t-1,t-1} \right] \\
&= \frac{1}{\alpha^2\sigma_{t-1}^2} \left[\sigma_t^2 - 2\sigma_t\sigma_{t-1} + \sigma_{t-1}^2 \right] + 2\frac{\sigma_t}{\sigma_{t-1}}\left[1 - \frac{R_{t,t-1}}{\sigma_t\sigma_{t-1}} \right] \\
&= \frac{1}{\alpha^2}(1 - \sqrt{\omega})^2 + h(t) \, .
\end{aligned}
$$

Here the last equality follows from the fact that $\sigma_t^2/\sigma_{t-1}^2 \to \omega$ by Lemma 6.(a1) and $R_{t,t-1}/(\sigma_t\sigma_{t-1}) \to 1$ by Lemma 6.(a2). This implies the claim for ξ^t.

Next, consider β^t. By Proposition 14.2

$$
\lim_{n\to\infty} \frac{1}{n} \|x^t - \widehat{x}^t\|_2^2 = \lim_{n\to\infty} \frac{1}{n} \|x^t - \eta(x^{t-1} + A^\mathsf{T}z^{t-1}; \alpha\sigma_{t-1})\|_2^2 = 0 \, , \tag{5.26}
$$

and hence $\|\beta^t\|_2^2/n \leq h(t)$ with high probability for any $h(t) > 0$

Finally consider ζ^t, and define $R(y; \theta) = y - \eta(y; \theta)$. We have

$$
R(y; \theta) = \begin{cases} +1 & \text{for } y \geq \theta, \\ y/\theta & \text{for } -\theta < y < \theta, \\ -1 & \text{for } y \leq -\theta. \end{cases}
$$

Using Proposition 14.3, we can show that

$$
\lim_{n\to\infty} \frac{1}{n}\|\zeta^t\|_2^2 = \mathbb{E}\{[\text{sign}(X) - R(X + \sigma_{t-1}Z; \alpha\sigma_{t-1})]^2 \mathbf{1}_{X\neq 0}\} \, . \tag{5.27}
$$

Notice that this apparently requires applying Proposition 14 to the function $\psi(x, y) = [\text{sign}(x) - R(y; \theta)]^2 \mathbf{1}_{x\neq 0}$ which is non-Lipschitz in x. However we can define a Lipschitz approximation, with parameter $r > 0$:

$$
\psi_r(x, y) = \begin{cases} [x/r - R(y; \theta)]^2 |x|/r & \text{for } |x| \leq r, \\ [1 - R(y; \theta)] & \text{for } |x| > r. \end{cases} \tag{5.28}
$$

Notice that ψ_r is bounded and Lipschitz continuous. We further have $|\psi_r(x, y) - \psi(x, y)| \leq 4\,\mathbf{1}(x \neq 0; |x| \leq r)$, whence

$$
\limsup_{n\to\infty} \left| \frac{1}{n}\|\zeta^t\|_2^2 - \frac{1}{n}\sum_{i=1}^n \psi_r(x_{0,i}, x_i^{t-1} + A^\mathsf{T}z^{t-1}) \right| \leq \limsup_{n\to\infty} \frac{4}{n}\sum_{i=1}^n \mathbf{1}(x_{0,i} \neq 0; |x_{0,i}| \leq r) \leq 8\,r \tag{5.29}
$$

The last inequality holds almost surely by the law of large numbers using $\gamma([-r, r]) < 2r$. Analogously

$$
\left| \mathbb{E}\psi(X, X + \sigma_{t-1}Z) - \mathbb{E}\psi_r(X, X + \sigma_{t-1}Z) \right| \leq 4\,\mathbb{P}(X \neq 0; |X| \leq r) \leq 8\,r \, . \tag{5.30}
$$

35

Hence the claim (5.27) follows by applying Proposition 14.3 to $\psi_r(x, y)$, using Eqs. (5.29), (5.30), and letting $r \to 0$.

We conclude by noting that the right-hand side of Eq. (5.27) converges to 0 as $t \to \infty$ by dominated convergence, since $\sigma_t \to 0$. Therefore

$$\lim_{n \to \infty} \frac{1}{n} \|\zeta^t\|_2^2 \leq \frac{h(t)}{2}.$$

this completes our proof of assumption 1 of Lemma 5.

We finally consider hypothesis 2. Let $S_t(c)$ be defined as there, for the subgradient v^t, namely

$$\begin{aligned}
S_t(c) &\equiv \left\{ i \in [n] \; : \; |v_i^t| \geq 1 - c \right\} \\
&= S \cup \left\{ i \in [n] \setminus S \; : \; |x^{t-1} + A^\mathsf{T} z^{t-1}| \geq (1 - c)\theta_{t-1} \right\}.
\end{aligned}$$

Recall that by assumption $A_{ij} = \tilde{A}_{ij} + \nu\, G_{ij}$ whereby $G_{ij} \sim \mathsf{N}(0, 1/m)$ and (eventually redefining $\tilde{A}_{ij}$) we can freely choose $\nu \in [0, \nu_0]$. Let $\{\tilde{x}^t, \tilde{z}^t\}_{t \geq 0}$ be a sequence of vectors defined as per Proposition 14.4, and define $\tilde{v}^t$ as v^t, but replacing x^t, z^t, A by $\tilde{x}^t, \tilde{z}^t, \tilde{A}$

$$\tilde{v}_i^t = \begin{cases} \operatorname{sign}(x_{0,i}) & \text{if } i \in S, \\ \frac{1}{\theta_{t-1}} \big(\tilde{x}^{t-1} + \tilde{A}^\mathsf{T} \tilde{z}^{t-1} - \hat{x}^t \big)_i & \text{otherwise,} \end{cases} \tag{5.31}$$

$$\hat{x}^t \equiv \eta(\tilde{x}^{t-1} + \tilde{A}^\mathsf{T} \tilde{z}^{t-1}; \theta_{t-1}). \tag{5.32}$$

We further define

$$\begin{aligned}
\widetilde{S}_t(c) &\equiv \left\{ i \in [n] \; : \; |\tilde{v}_i^t| \geq 1 - c \right\} \\
&= S \cup \left\{ i \in [n] \setminus S \; : \; |\tilde{x}^{t-1} + \tilde{A}^\mathsf{T} \tilde{z}^{t-1}| \geq (1 - c)\theta_{t-1} \right\}.
\end{aligned}$$

We claim that the following two claims hold for some $t_* \geq 0$ independent of n:

Claim 1. There exists $c_1, \hat{c}_2 > 0$ (independent of ν) such that for all $S' \subseteq [n]$, $|S'| \leq 2c_1 n$, the minimum singular value of $A_{\widetilde{S}_{t_*}(2c_1) \cup S'}$, satisfies $\sigma_{\min}(A_{\widetilde{S}_{t_*}(2c_1) \cup S'}) \geq \hat{c}_2 \nu$ with probability converging to 1 as $n \to \infty$.

Claim 2. For all $t \geq t_*$,

$$\mathbb{P}\left\{ |S_t(c_1) \setminus \widetilde{S}_{t_*}(2c_1)| \geq n\, c_1 \right\} = o_1(t_*; \nu) + o_2(t_*, \nu; n^{-1}),$$

where $o_1(t_*, \nu)$ vanishes as $\nu \to 0$ at t_*, c_1, c_2 fixed, and $o_2(t_*, \nu; n^{-1})$ vanishes as $n^{-1} \to 0$ at t_*, ν, c_1, c_2 fixed.

These claims immediately imply that hypothesis 2 of Lemma 5 holds with probability converging to one as $n \to \infty$. Indeed, if $|S'| \leq nc_1$, then (by Claim 2) $S_t(c_1) \cup S' \subseteq \widetilde{S}_{t_*}(2c_1) \cup S''$ where $|S''| \leq 2nc_1$ with probability larger than $1 - o_1(t_*; \nu) - o_2(t_*, \nu; n^{-1})$. By Claim 1, we hence have $\sigma_{\min}(A_{S_t(c_1) \cup S'}) \geq c_2 \equiv \hat{c}_2 \nu$. The thesis follows since ν can be chosen as small as we want. (Notice that once t_* is fixed to satisfy these claims, we can still choose $t \geq t_*$ arbitrarily to satisfy hypothesis 1 of Lemma 5, as per the argument above.)

In order to prove Claim 1, above first notice that, for any $b \geq 0$

$$\mathbb{P}\Big\{ \min_{\substack{S' \subseteq [n] \\ |S'| \leq 2c_1 n}} \sigma_{\min}(A_{\widetilde{S}_{t_*}(2c_1) \cup S'}) < \hat{c}_2 \nu \Big\}$$

$$\leq \mathbb{P}\Big\{ \min_{\substack{S' \subseteq [n] \\ |S'| \leq 2c_1 n}} \sigma_{\min}(A_{\widetilde{S}_{t_*}(2c_1) \cup S'}) < \hat{c}_2 \nu;\ |\widetilde{S}_{t_*}(2c_1)| \leq bn \Big\} + \mathbb{P}\big\{ |\widetilde{S}_{t_*}(2c_1)| > bn \big\}$$

$$\leq e^{nH(2c_1)} \max_{\substack{S' \subseteq [n] \\ |S'| \leq 2c_1 n}} \mathbb{P}\Big\{ \sigma_{\min}(A_{\widetilde{S}_{t_*}(2c_1) \cup S'}) < \hat{c}_2 \nu;\ |\widetilde{S}_{t_*}(2c_1)| \leq bn \Big\} + \mathbb{P}\big\{ |\widetilde{S}_{t_*}(2c_1)| > bn \big\},$$

$$(5.33)$$

where in the last line $H(c)$ denotes the binary entropy of b and we used $\binom{n}{nc} \leq \exp\{nH(c)\}$. We want to show that t_*, b, c_1, c_2, ν can be chosen so that both contributions vanish as $n \to \infty$.

Consider any $b \in (0, \delta)$ and restrict $c_1 \in (0, (\delta - b)/2)$. Then the matrix $A_{\widetilde{S}_{t_*}(2c_1) \cup S'}$ has $n\delta$ rows and $n\delta - \Theta(n)$ columns. Further $A = \tilde{A} + \nu G$ with $\widetilde{S}_{t_*}(2c_1)$ (and hence $\widetilde{S}_{t_*}(2c_1) \cup S'$) independent of G. We can therefore use an upper bound on the condition number of randomly perturbed deterministic matrices proved by Buergisser and Cucker [BC10] (see also Appendix D) to show that

$$\mathbb{P}\Big\{ \sigma_{\min}(A_{\widetilde{S}_{t_*}(2c_1) \cup S'}) < \hat{c}_2 \nu;\ |\widetilde{S}_{t_*}(2c_1)| \leq bn \Big\} \leq (a_1 \hat{c}_2)^{n(\delta - b - 2c_1) + 1} \qquad (5.34)$$

with $a_1 = a_1((b + 2c_1)/\delta)$ bounded as long as $(b + 2c_1)/\delta < 1$ We can therefore select $\hat{c}_2 = 1/(2a_1)$ and select c_1 small enough so that $H(2c_1) \leq (1/2)(\delta - b - 2c_1) \log 2$. This ensures that the first term in Eq. (5.33) vanishes as $n \to \infty$.

We are left with the task of selecting $b \in (0, \delta)$, $t_* \geq 0$, so that the second term vanishes as well, since then we can take $c_1 \in (0, (\delta - b)/2)$. To this hand notice that by Proposition 14 (and using the fact that $X + \sigma_{t-1}Z$ has a density) we have, in probability,

$$\lim_{n \to \infty} \frac{1}{n} |S_{t_*}(c)| = \mathbb{P}\big\{ |X + \sigma_{t_* - 1}Z| \geq (1 - c)\theta_{t_* - 1} \big\},$$

and further, since $\sigma_t \to 0$ as $t \to \infty$ (cf. Lemma 6.(a1)) and $\theta_t = \alpha \sigma_t$, we have

$$\lim_{t_* \to \infty} \mathbb{P}\big\{ |X + \sigma_{t_* - 1}Z| \geq (1 - c)\theta_{t_* - 1} \big\} = \varepsilon + 2(1 - \varepsilon)\Phi(-(1 - c)\alpha).$$

On the other hand, by Lemma 6.(a1), and since $\alpha \in [\alpha_*, \alpha_2)$, we have $\varepsilon + 2(1 - \varepsilon)\Phi(-\alpha) < \delta$. Hence there exist $b_0 \in (0, \delta)$ and $c_1 > 0$ so that for all t_* large enough $|S_{t_*}(3c_1)| \leq nb_0$ with high probability. Taking $b \in (b_0, \delta)$ and using Markov inequality (with $t'_* = t_* - 1$)

$$\mathbb{P}\big\{ |\widetilde{S}_{t_*}(2c_1)| > bn \big\} \leq \frac{1}{(b - b_0)n}\mathbb{E}\big\{ |\widetilde{S}_{t_*}(2c_1) \setminus S_{t_*}(3c_1)| \big\} + \mathbb{P}\big\{ |S_{t_*}(3c_1)| > b_0 n \big\}$$

$$\leq \frac{1}{(b - b_0)c_1^2 \theta_{t_* - 1}^2 n} \sum_{i=1}^{n} \mathbb{E}\big\{ ((x^{t'_*} + A^\mathsf{T} z^{t'_*})_i - (\tilde{x}^{t'_*} + \tilde{A}^\mathsf{T} \tilde{z}^{t'_*})_i)^2 \geq c_1^2 \theta_{t'_*}^2 \big\} + \mathbb{P}\big\{ |S_{t_*}(3c_1)| > b_0 n \big\}$$

$$\leq o_1(t_*; \nu) + o_2(t_*, \nu; n^{-1}) + \mathbb{P}\big\{ |S_{t_*}(3c_1)| > b_0 n \big\},$$

where the last inequality follows from Proposition 14.4. LL terms can be made arbitrarily small by choosing ν small and n large enough.

In order to conclude the proof, we need to show that Claim 2 holds for eventually larger t_*. First notice that, applying again Proposition 14.4, we get

$$\mathbb{P}\big\{\big|S_{t_*}(c_1) \setminus \widetilde{S}_{t_*}(2c_1)\big| \geq nc_1/2\big\} \leq \frac{2}{nc_1}\mathbb{E}\big\{\big|S_{t_*}(c_1) \setminus \widetilde{S}_{t_*}(2c_1)\big|\big\}$$

$$\leq \frac{2}{nc_1}\sum_{i=1}^{n}\mathbb{E}\big\{\big((x^{t'_*} + A^{\mathsf{T}}z^{t'_*})_i - (\tilde{x}^{t'_*} + \tilde{A}^{\mathsf{T}}\tilde{z}^{t'_*})_i\big)^2 \geq c_1^2\theta_{t'_*}^2\big\} \leq o_1(t_*;\nu) + o_2(t_*,\nu;n^{-1}). \quad (5.35)$$

By Proposition 15, and using the fact that the vector $(X + Z_{t_*}, X + Z_t)$ has a density, we have, in probability,

$$\lim_{n\to\infty}\frac{1}{n}|S_t(c_1) \setminus S_{t_*}(c_1)| = \mathbb{P}\big\{|X + Z_{t_*-1}| \geq (1-c_1)\sigma_{t_*-1}; \ |X + Z_{t-1}| < (1-c_1)\sigma_{t-1}\big\} \leq h(t_*),$$

where, by Lemma 6.(a3), $h(t_*)$ vanishes as $t_* \to \infty$. Given any $c_1 > 0$, we can therefore choose t_* so that, with high probability $|S_t(c_1) \setminus S_{t_*}(c_1)| \leq nc_1/2$. Combining with Eq. (5.35), we obtain the desired Claim.

5.3   Proof of Theorem 8, $\rho > \rho_*(\delta)$

Fix a small number $h > 0$. By Lemma 6.(b), there exists $\Delta = \Delta(\delta, \varepsilon) > 0$ independent of h, such that, for $\alpha = \alpha_0(\delta, p_X)$ and t large enough

$$\left|\frac{1}{\delta}\mathbb{P}\big\{|X + \sigma_t Z| > \alpha\sigma_t\big\} - 1\right| \quad \leq \quad h, \quad (5.36)$$

$$\big|R_{t,t} - 2R_{t,t-1} + R_{t-1,t-2}\big| \quad \leq \quad h^2, \quad (5.37)$$

$$\mathbb{E}\{|\eta(X + \sigma_t Z; \alpha\sigma_t)|\} \quad < \quad \mathbb{E}\{|X|\} - 2\Delta, \quad (5.38)$$

as well as $\sigma_{t-1}^2 \leq 2\sigma_*^2$. By Propositions 14, 15 (and noting that $X + \sigma_t Z$ has a distribution that is absolutely continuous with respect to Lebesgue measure), we have, with high probability,

$$\max_{i\in[m]}\big|(\mathsf{b}_t - 1)_{ii}\big| \quad \leq \quad 2h, \quad (5.39)$$

$$\|z^t - z^{t-1}\|_2 \quad \leq \quad 2h\sqrt{n}, \quad (5.40)$$

$$\|x^t\|_1 \quad \leq \quad \|x_0\|_1 - n\Delta, \quad (5.41)$$

$$\|z^t\|_2 \quad \leq \quad 2\sigma_*\sqrt{n}. \quad (5.42)$$

Namely Eq. (5.36) implies (5.39), Eq. (5.37) implies (5.40), Eq. (5.38) implies (5.41), and the assumption $\sigma_{t-1}^2 \leq 2\sigma_*^2$ implies (5.42).

Using Eq. (5.9) together with the above, we get

$$\|y - Ax^t\|_2 \leq \|z^t - z^{t-1}\|_2 + \max_{i\in[m]}|(\mathsf{b}_t)_{ii} - 1|\,\|z^{t-1}\|_2 \leq 2h\sqrt{n}\,(1 + 2\sigma_*). \quad (5.43)$$

Define $\tilde{x} = x^t + A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}(y - Ax^t)$ (notice that the sample covariance matrix AA^{T} has full rank with high probability [BS98, BS05]). Notice that, by construction $A\tilde{x} = y$. Then, with high probability

$$\|\tilde{x} - x^t\|_2 \leq \sigma_{\max}(A)\sigma_{\min}(A)^{-2}\|y - Ax^t\|_2 \leq C(\delta)(1 + 2\sigma_*)h\sqrt{n}, \quad (5.44)$$

38

where $\sigma_{\max}(A)$, $\sigma_{\min}(A)$ are the maximum and minimum non-zero singular values of A. The second inequality holds with high probability for $\delta \in (0,1)$ by standard estimates on the singular values of random matrices [BS98, BS05]. Using Eq. (5.41) together with triangular inequality and $\|\tilde{x} - x^t\|_1 \leq \sqrt{n}\|\tilde{x} - x^t\|_2$ we finally get

$$\|\tilde{x}\|_1 \leq \|x_0\|_1 - n\Delta + C(\delta)(1 + 2\sigma_*)hn < \|x_0\|_1 \tag{5.45}$$

where the second inequality follows from the fact that $h > 0$ can be taken arbitrarily small (by letting t large) while Δ, C and σ_* are fixed. We conclude that x_0 cannot be the solution of the ℓ_1 minimization problem (5.1).

Acknowledgments

A Proof of Proposition 14 and 15

In this appendix we prove Proposition 14 and 15 by a suitable application of Theorem 6. Before passing to these proofs, we establish a corollary of Theorem 6 that allows to control iterations of the form (5.6), (5.7), with $\eta(\,\cdot\,;\,\cdot\,)$ replaced by a general polynomial.

A.1 A general corollary

For $x_0 = x_0(n) \in \mathbb{R}^n$ and $A = A(n) \in \mathbb{R}^{m \times n}$ as per Hypothesis 1 in Section 5, we define $y = y(n) \in \mathbb{R}^m$ by

$$y = Ax_0. \tag{A.1}$$

Let $D \in \mathbb{R}^{n \times n}$ be the diagonal matrix with diagonal entries equal to the square column norms of A, that is $D_{ii} = \sum_{j \in [m]} A_{ji}^2$, and $D_{ij} = 0$ for $i \neq j$. Further define $u_0 = u_0(n) \in \mathbb{R}^n$ as follows

$$u_{0,i} = (D_{ii} - 1)x_{0,i} = \Big(\sum_{j \in [m]} A_{ji}^2 - 1 \Big) x_{0,i}. \tag{A.2}$$

Let $x^0 = (I - D^{-1})x_0$ (notice that D is invertible with high probability) and define iteratively

$$z^t = y - Ax^t + \mathsf{b}_t\, z^{t-1}, \qquad (\mathsf{b}_t)_{ii} = \sum_{j \in [n]} A_{ij}^2 \eta_{t-1}' \Big(D_{jj}x_j^{t-1} + (A^\mathsf{T}z^{t-1})_j - u_{0,j} \Big), \tag{A.3}$$

$$x^{t+1} = \eta_t(Dx^t + A^\mathsf{T}z^t - u_0), \tag{A.4}$$

where, for each t, $\eta_t : \mathbb{R} \to \mathbb{R}$ is a polynomial and, for $v \in \mathbb{R}^n$, $\eta_t(v) = (\eta_t(v_1), \ldots, \eta_t(v_n))$. Further $\mathsf{b}_t \in \mathbb{R}^{m \times m}$ is a diagonal matrix with entries given as in Eq. (A.3).

We next introduce the corresponding state evolution recursion. Namely, we define $\{\widetilde{R}_{s,t}\}_{s,t\geq 0}$ recursively for all $s, t \geq 0$ by letting

$$\widetilde{R}_{s+1,t+1} = \frac{1}{\delta} \mathbb{E}\big\{[\eta_s(X + Z_s) - X][\eta_t(X + Z_t) - X]\big\}. \tag{A.5}$$

Here expectation is with respect to $X \sim p_X$ and the independent Gaussian vector $[Z_s, Z_t]$ with zero mean and covariance given by $\mathbb{E}\{Z_s^2\} = \widetilde{R}_{s,s}$, $\mathbb{E}\{Z_t^2\} = \widetilde{R}_{t,t}$ and $\mathbb{E}\{Z_t Z_s\} = \widetilde{R}_{t,s}$. The boundary condition is fixed by letting $\widetilde{R}_{0,0} = \mathbb{E}\{X^2\}/\delta$ and defining, for each $t \geq 0$,

$$\widetilde{R}_{0,t+1} = \frac{1}{\delta} \mathbb{E}\big\{[\eta_t(X + Z_t) - X][-X]\big\}, \tag{A.6}$$

with $Z_t \sim \mathsf{N}(0, \widetilde{R}_{t,t})$. This uniquely determines the doubly infinite array $\{\widetilde{R}_{t,s}\}_{t,s\geq 0}$.

Corollary 16. *Let $\{(x_0(n), A(n), y(n))\}_{n\geq 0}$ be a sequence of triples with $A(n)$ having independent subgaussian entries with $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/m$, $\{x_{0,i}(n) : i \in [n]\}$ independent and identically distributed with $x_{0,i}(n) \sim p_X$, and p_X a finite mixture of Gaussians. Define $\{x^t, z^t\}_{t\geq 0}$ as per Eqs. (A.3), (A.4).*

Then, for any fixed $t, s \geq 0$, and any Lipschitz continuous functions $\psi : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, in probability

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \psi\left(x_{0,i}, x_i^s + (A^\mathsf{T} z^s)_i, x_i^t + (A^\mathsf{T} z^t)_i\right) = \mathbb{E}\,\psi(X, X + Z_s, X + Z_t), \tag{A.7}$$

$$\lim_{n\to\infty} \frac{1}{m} \sum_{i=1}^{n} \phi(z_i^s, z_i^t) = \mathbb{E}\,\phi(Z_s, Z_t), \tag{A.8}$$

where expectation is with respect to $X \sim p_X$ and the independent Gaussian vector $[Z_s, Z_t]$ with zero mean and covariance given by $\mathbb{E}\{Z_s^2\} = \widetilde{R}_{s,s}$, $\mathbb{E}\{Z_t^2\} = \widetilde{R}_{t,t}$ and $\mathbb{E}\{Z_t Z_s\} = \widetilde{R}_{t,s}$.

Proof. Define $\tilde{x}^{t+1} = A^\mathsf{T} z^t + D x^t - D x_0$. Then Eqs. (A.3), (A.4) read

$$z^t = A\,f(\tilde{x}^t, x_0; t) + \mathsf{b}_t h(z^{t-1}; t-1), \tag{A.9}$$

$$\tilde{x}^{t+1} = A^\mathsf{T}\,h(x^t; t) + \mathsf{d}_t f(\tilde{x}^t, x_0; t), \tag{A.10}$$

where, for $i \in [m]$, $j \in [n]$,

$$f(x, y; t) = y - \eta_{t-1}(y + x), \qquad h(z; t) = z, \tag{A.11}$$

$$(\mathsf{b}_t)_{ii} = -\sum_{j\in[n]} A_{ij}^2 f'(\tilde{x}_j^t, x_{0,i}; t), \tag{A.12}$$

$$(\mathsf{d}_t)_{jj} = -\sum_{j\in[n]} A_{ij}^2 h'(z; t). \tag{A.13}$$

(Here $f'(x, y; t)$, $h'(x; t)$ denote derivatives with respect to the first argument.) The iteration takes the same form as in Eqs. (4.3), (4.4) with $Y(i) = x_{0,i}$, and $W(i) = 0$, $\mathsf{B}_t = -\mathsf{b}_t$ and $\mathsf{D}_t = -\mathsf{d}_t$. Further, the initial condition x^0 implies $\tilde{x}^0 = -x_0$. Notice that this is dependent on $Y = x_0$, but we

can easily set the initial condition at $\tilde{x}^{-1} = 0$ and define $f(x, y; t = 0) = -y$. We can therefore apply Theorem 6 and conclude that, in probability

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \psi \left(x_{0,i}, D_{ii}(x_i^s - x_{0,i}) + (A^\mathsf{T} z^s)_i, D_{ii}(x_i^t - x_{0,i}) + (A^\mathsf{T} z^t)_i \right) = \mathbb{E} \, \psi(X, Z_s, Z_t) \,, \quad \text{(A.14)}$$

$$\lim_{n \to \infty} \frac{1}{m} \sum_{i=1}^{n} \phi(z_i^s, z_i^t) = \mathbb{E} \, \phi(Z_s, Z_t) \,, \quad \text{(A.15)}$$

where expectations are defined as in the statement of the Corollary. The second of these equations coincides with Eq. (A.8). For the first one, note that $\mathbb{E}\{D_{ii}\} = 1$ and, by a standard Chernoff bound

$$\lim_{n \to \infty} \max \big\{ D_{ii} : \; i \in [n] \big\} \;\; = \;\; 1 \,, \quad \text{(A.16)}$$

$$\lim_{n \to \infty} \min \big\{ D_{ii} : \; i \in [n] \big\} \;\; = \;\; 1 \,. \quad \text{(A.17)}$$

We therefore get

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \psi \left(x_{0,i}, (x^s + A^\mathsf{T} z^s)_i - x_{0,i}, (x_i^t + A^\mathsf{T} z^t)_i - x_{0,i} \right) = \mathbb{E} \, \psi(X, Z_s, Z_t) \,, \quad \text{(A.18)}$$

which coincides with Eq. (A.7) after a redefinition of the function ψ. $\qquad\square$

A.2 Proofs of Propositions 14 and 15

We will start by proving Proposition 14. Since Proposition 15 follows from the same construction, we will only point to the necessary modifications. Before presenting the proof, we recall a basic result in weighted polynomial approximation (here stated for a specific case), see e.g. [Lub07].

Theorem 9. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then for any $\kappa, \xi > 0$ there exists a polynomial $p : \mathbb{R} \to \mathbb{R}$ such that, for all $x \in \mathbb{R}$,*

$$\big| f(x) - p(x) \big| \leq \xi \, e^{\kappa x^2 / 2} \,. \quad \text{(A.19)}$$

Proof of Propositions 14. Since the proposition holds as $n \to \infty$ at t fixed, we shall assume throughout that $t \in \{0, 1, \dots, t_{\max}\}$ for some fixed arbitrarily large $t_{\max}$.

We claim that, for each $\beta, t_{\max} > 0$, we can construct an orbit $\{x^{\beta,t}, z^{\beta,t}\}_{t \geq 0}$ obeying Eqs. (A.3), (A.4) for suitable functions $\eta_t = \eta_t^{(\beta)}$ such that the following holds (with a slight abuse of notation we will drop the parameter β from $x^{\beta,t}, z^{\beta,t}$). For all $0 \leq t \leq t_{\max}$, and all functions ψ as in the statement, we have $z^t = y - Ax^t + \mathsf{b}_t \, z^{t-1}$ by construction. Further, in probability,

$$\lim_{n \to \infty} \max_{i \in [m]} \left| (\mathsf{b}_t)_{ii} - \frac{1}{\delta} \mathbb{P}\big\{ |X + \sigma_{t-1} Z| \geq \alpha \sigma_{t-1} \big\} \right| \leq \beta \,, \quad \text{(A.20)}$$

$$\lim_{n \to \infty} \frac{1}{n} \, \| x^{t+1} - \eta(x^t + A^\mathsf{T} z^t; \alpha \sigma_t) \|_2^2 \leq \beta \,, \quad \text{(A.21)}$$

$$\lim_{n \to \infty} \left| \frac{1}{n} \sum_{i=1}^{n} \psi(x_{0,i}, x_i^t + (A^\mathsf{T} z^t)_i) - \mathbb{E} \, \psi(X, X + \sigma_t Z) \right| \leq \beta \,. \quad \text{(A.22)}$$

Assuming this claim holds, let $\{\beta_\ell\}_{\ell\geq 0}$ be a sequence such that $\lim_{\ell\to\infty}\beta_\ell = 0$. Denote by $\{x^{\ell,t}, z^{\ell,t}\}_{t\geq 0}$ the orbit satisfying Eqs. (A.20), (A.21), (A.22) with $\beta = \beta_\ell$. Let $\eta_t^\ell = \eta_t^{(\beta_\ell)}$ be the corresponding polynomial, and b_t^ℓ be given per Eq. (A.3). Fix an increasing sequence of instance sizes $n_1 < n_2 < n_3 < \ldots$, and let $x^t(n) = x^{\ell,t}(n)$, $z^t(n) = z^{\ell,t}(n)$ for all $n_\ell \leq n < n_{\ell+1}$. Choosing $\{n_\ell\}_{\ell\geq 0}$ that increases rapidly enough we can ensure that, for all $n \geq n_\ell$,

$$\max_{i\in[m]}\left|(\mathsf{b}_t^\ell)_{ii} - \frac{1}{\delta}\mathbb{P}\{|X + \sigma_{t-1}Z| \geq \alpha\sigma_{t-1}\}\right| \leq 2\beta_\ell\,, \tag{A.23}$$

$$\frac{1}{n}\|x^{\ell,t+1} - \eta(x^{\ell,t} + A^\mathsf{T}z^{\ell,t}; \alpha\sigma_t)\|_2^2 \leq 2\beta_\ell\,, \tag{A.24}$$

$$\left|\frac{1}{n}\sum_{i=1}^n \psi(x_{0,i}, x_i^{\ell,t} + (A^\mathsf{T}z^{\ell,t})_i) - \mathbb{E}\,\psi(X, X + \sigma_t Z)\right| \leq 2\beta_\ell\,. \tag{A.25}$$

with probability larger than $1 - \beta_\ell$. Points 1, 2, 3 in the proposition then follow since $\beta_\ell \to 0$.

In order to prove Eqs. (A.20) to (A.22) we proceed as follows. It is easy to check that $\sigma_t > 0$ for all t, cf. Eq. (5.8). We use Theorem 9 to construct polynomials η_t such that

$$\left|\eta(x; \alpha\sigma_t) - \eta_t(x)\right| \leq \xi\,\exp\left\{\frac{x^2}{16\max(\sigma_t^2, s^2)}\right\}\,, \tag{A.26}$$

for all $x \in \mathbb{R}$. Here $\xi > 0$ is a small parameter to be chosen below, and s^2 is the smallest variance of the Gaussians that are combined in p_X. Let $\widetilde{\sigma}_t$ be defined by

$$\widetilde{\sigma}_{t+1}^2 = \frac{1}{\delta}\mathbb{E}\{[\eta_t(X + \widetilde{\sigma}_t Z) - X]^2\}\,, \tag{A.27}$$

with $Z \sim \mathsf{N}(0,1)$ independent from $X \sim p_X$, and $\widetilde{\sigma}_0^2 = \mathbb{E}\{X^2\}/\delta$. Notice that $\widetilde{\sigma}_t^2 = \widetilde{R}_{tt}$. From Eqs. (5.8), (A.26), and (A.27), it is then straightforward to show that $|\sigma_t^2 - \widetilde{\sigma}_t^2| \leq C\xi$ for some $C = C(t)$.

Given polynomials as defined by (A.26), we define $\{x^t, z^t\}_{t\geq 0}$ as per Eqs. (A.3), (A.4), with the initial condition given there. Equation (A.22) follows immediately from Corollary 16 for ξ sufficiently small. Equation (A.21) also follows from the same Corollary, by taking

$$\psi(x_1, x_2, x_3) = \{\eta_t(x_3) - \eta(x_3; \alpha\sigma_t)\}^2\,, \tag{A.28}$$

and then using once again Eq. (A.26) on the resulting expression.

Finally, consider Eq. (A.20). For economy of notation, we write

$$(\mathsf{b}_t)_{ii} = \sum_{j\in[n]} A_{ij}^2 \varphi_j\,, \qquad \varphi_i = \eta_{t-1}'(D_{jj}x_j^{t-1} + (A^\mathsf{T}z^{t-1})_j - u_{0,j})\,, \tag{A.29}$$

and further define

$$\mathsf{b}_t^{\mathrm{av}} = \frac{1}{m}\sum_{j\in[n]} \varphi_j\,. \tag{A.30}$$

Then we have

$$\mathbb{E}\left\{\left((\mathsf{b}_t)_{ii} - \mathsf{b}_t^{\mathrm{av}}\right)^4\right\} = \sum_{j_1,j_2,j_3,j_4 \in [n]} \mathbb{E}\left\{\left(A_{ij_1}^2 - \frac{1}{m}\right)\left(A_{ij_2}^2 - \frac{1}{m}\right)\left(A_{ij_3}^2 - \frac{1}{m}\right)\left(A_{ij_4}^2 - \frac{1}{m}\right)\varphi_{j_1}\varphi_{j_2}\varphi_{j_3}\varphi_{j_3}\right\}$$

$$= \sum_{j_1,j_2,j_3,j_4 \in [n]} E(j_1, j_2, j_3, j_4)$$

Using the tree representation in Section 3.2, it is not hard to prove that the expectation on the right-hand side is bounded as follows

$$E(p, q, r, s) \le \frac{K}{n^6}, \qquad p, q, r, s \text{ distinct,}$$

$$E(q, q, r, s) \le \frac{K}{n^5}, \qquad q, r, s \text{ distinct,}$$

$$E(r, r, s, s) \le \frac{K}{n^4}, \qquad r, s \text{ distinct,}$$

$$E(r, r, r, s) \le \frac{K}{n^4}, \qquad r, s \text{ distinct,}$$

$$E(r, r, r, r) \le \frac{K}{n^3}.$$

Consider for instance the first case, p, q, r, s distinct. Using Lemma 3, each of φ_p, φ_q, φ_r φ_s can be represented as a sum over trees with root type respectively at p, q, r, s. The weight of these trees is as in Lemma 3, times the prefactor $(A_{ip}^2 - m^{-1})\cdots(A_{is}^2 - m^{-1})$. Let μ be the total number of edges in these trees, plus 8 (two for each of the additional factors). Then any non-vanishing contribution is of order $n^{-\mu/2}$. Let $\mathbf{G}$ be the graph obtained by identifying the vertices of the same type in these trees, and $e(\mathbf{G})$ the number of its edges. Since each edge in $\mathbf{G}$ must be covered at least twice by the trees to get a non-zero expectation, and the edges in $(i, p), \ldots, (i, s)$ at least once, we have $2e(\mathbf{G}) + 4 \le \mu$. The number of vertices in $\mathbf{G}$ is at most $e(\mathbf{G}) + 1$ (note that $\mathbf{G}$ is connected because it includes type i connected to p, q, r, s). Of these vertices all but 5 (whose type is i, p, q, r, s) can take an arbitrary type, yielding a combinatorial factor of order $n^{e(\mathbf{G})+1-5} \le n^{\mu/2-6}$. Hence the sum over trees is of order $n^{-\mu/2}n^{\mu/2-6} = n^{-6}$ as claimed.

Summing over $j_1, \ldots, j_4$ de above bounds we obtain $\mathbb{E}\left\{\left((\mathsf{b}_t)_{ii} - \mathsf{b}_t^{\mathrm{av}}\right)^4\right\} \le K/n^2$ and therefore, by Markov inequality

$$\lim_{n\to\infty} \mathbb{P}\left\{\max_{i\in[m]} |(\mathsf{b}_t)_{ii} - \mathsf{b}_t^{\mathrm{av}}| \ge n^{-1/5}\right\} = 0. \tag{A.31}$$

Since by standard concentration bounds $\max_{i\in[n]} D_{ii}$, $\min_{i\in[n]} D_{ii} \to 1$, we obtain, in probability,

$$\lim_{n\to\infty} \max_{i\in[m]}(\mathsf{b}_t)_{ii} = \lim_{n\to\infty} \min_{i\in[m]}(\mathsf{b}_t)_{ii} = \lim_{n\to\infty} \mathsf{b}_t^{\mathrm{av}}$$

$$= \lim_{n\to\infty} \frac{1}{m}\sum_{j\in[n]} \eta'_{t-1}(x_j^{t-1} + (A^\mathsf{T} z^{t-1})_j)$$

$$= \frac{1}{\delta}\mathbb{E}\left\{\eta'_{t-1}(X + \widetilde{\sigma}_{t-1} Z)\right\}$$

43

where, in the last step, we applied Corollary 16 to the polynomials η'_{t-1}, and $X \sim p_X$, $Z \sim \mathsf{N}(0,1)$ are independent. We are left with the task of showing that, by taking ξ small enough in Eq. (A.26), we can ensure that

$$\left| \mathbb{E}\{\eta'_{t-1}(X + \widetilde{\sigma}_{t-1} Z)\} - \mathbb{P}\{|X + \sigma_{t-1} Z| \geq \alpha \sigma_{t-1}\} \right| \leq \beta \delta. \tag{A.32}$$

Indeed integrating by parts with respect to Z the above difference can be written as (for K a finite constant that can depend on t and change from line to line)

$$\left| \frac{1}{\widetilde{\sigma}_{t-1}} \mathbb{E}\{Z \eta_{t-1}(X + \widetilde{\sigma}_{t-1} Z)\} - \frac{1}{\sigma_{t-1}} \mathbb{E}\{Z \eta(X + \widetilde{\sigma}_{t-1} Z; \alpha \sigma_{t-1})\} \right|$$

$$\leq K \mathbb{E}\left| Z \eta_{t-1}(X + \sigma_{t-1} Z) - Z \eta(X + \sigma_{t-1} Z; \alpha \sigma_{t-1}) \right| + K|\sigma_{t-1} - \widetilde{\sigma}_{t-1}|$$

$$\leq K \xi \mathbb{E}\left\{ \exp\left\{ \frac{X^2 + \sigma_{t-1}^2 X^2}{4 \max(\sigma_t^2, s^2)} \right\} \right\} + K|\sigma_{t-1} - \widetilde{\sigma}_{t-1}|$$

$$\leq K\xi + K|\sigma_{t-1} - \widetilde{\sigma}_{t-1}|.$$

The claim follows by noting that, as argued above $|\sigma_{t-1} - \widetilde{\sigma}_{t-1}| \leq K'\xi$.

Consider finally point 4. First recall that we constructed the vectors $\{x^t, z^t\}_{t \geq 0}$, using a sequence of orbits $\{x^{\ell,t}, z^{\ell,t}\}_{t \geq 0}$, indexed by $\ell \in \mathbb{N}$, that obey Eqs. (A.3), (A.4), and letting

$$x^t(n) = x^{\ell,t}(n), \quad z^t(n) = z^{\ell,t}(n), \qquad \text{for all } n, \text{ with } n_\ell \leq n < n_{\ell+1}. \tag{A.33}$$

Claim 17. *There exists a sequence $\{\widetilde{\beta}_\ell\}_{\ell \in \mathbb{N}}$ with $\lim_{\ell \to \infty} \widetilde{\beta}_\ell = 0$ such that, for all $\ell' \geq \ell$,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i \in [n]} \mathbb{E}\{ ((x^{\ell',t} + A^\mathsf{T} z^{\ell',t})_i - (x^{\ell,t} + A^\mathsf{T} z^{\ell,t})_i)^2 \} \leq \widetilde{\beta}_\ell, \tag{A.34}$$

$$\lim_{n \to \infty} \frac{1}{m} \sum_{i \in [m]} \mathbb{E}\{ (z_i^{\ell',t} - z_i^{\ell,t})^2 \} \leq \widetilde{\beta}_\ell. \tag{A.35}$$

The proof of this claim is presented below. It follows from this claim that, by eventually redefining $n_{\ell'}$ to be larger we can ensure

$$\mathbb{E}\{ ((x^{\ell',t} + A^\mathsf{T} z^{\ell',t})_I - (x^{\ell,t} + A^\mathsf{T} z^{\ell,t})_I)^2 \} \leq 2\widetilde{\beta}_\ell,$$
$$\mathbb{E}\{ (z_J^{\ell',t} - z_J^{\ell,t})^2 \} \leq 2\widetilde{\beta}_\ell.$$

for all $n \geq n_{\ell'}$. Here and below expectation is taken also with respect to I uniformly random in $[n]$ and J uniformly random in $[m]$. By Eq. (A.33), for all $n \geq n_\ell$, we also have

$$\mathbb{E}\{ ((x^t + A^\mathsf{T} z^t)_I - (x^{\ell,t} + A^\mathsf{T} z^{\ell,t})_I)^2 \} \leq 2\widetilde{\beta}_\ell,$$
$$\mathbb{E}\{ (z_J^t - z_J^{\ell,t})^2 \} \leq 2\widetilde{\beta}_\ell.$$

Applying Lemma 4, we can then construct $\{\widetilde{x}^t, \widetilde{z}^t\}_{t \geq 0}$ as in the statement at point 4, such that

$$\mathbb{E}\{ ((\widetilde{x}^t + \widetilde{A}^\mathsf{T} \widetilde{z}^t)_I - (x^{\ell,t} + A^\mathsf{T} z^{\ell,t})_I)^2 \} \leq K(\nu^2 + n^{-1/2}),$$
$$\mathbb{E}\{ (\widetilde{z}_J^t - z_J^{\ell,t})^2 \} \leq K(\nu^2 + n^{-1/2}),$$

where K depends on ℓ but not on ν or n. Proof is finished by using triangular inequality and selecting $\ell = \ell(\nu, t)$ diverging slowly enough as $\nu \to 0$. $\square$

We now prove Claim 17.

Proof of Claim 17. To be definite we will focus on Eq. (A.34).

Fis ℓ, $\ell' \in \mathbb{N}$ (not necessarily distinct). By an immediate generalization of Corollary 16, we have, in probability

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i \in [n]} \mathbb{E}\{(x^{\ell,t} + A^\mathsf{T} z^{\ell,t} - x_0)_i (x^{\ell',t} + A^\mathsf{T} z^{\ell',t} - x_0)_i\} = Q_{\ell,\ell'}^t. \tag{A.36}$$

Further, the quantities $Q_{\ell,\ell'}^t$ satisfy the state evolution recursion

$$Q_{\ell,\ell'}^{t+1} = \frac{1}{\delta} \mathbb{E}\left\{ \left[\eta_t^\ell(X + Z_{t,\ell}) - X \right] \left[\eta_t^{\ell'}(X + Z_{t,\ell'}) - X \right] \right\}, \tag{A.37}$$

with initial condition $Q_{\ell,\ell'}^0 = (1/\delta)\mathbb{E}\{X^2\}$. Here expectation is taken with respect to $X \sim p_X$ and the independent centered Gaussian vector $(Z_{t,\ell}, Z_{t,\ell'})$ with covariance given by $\mathbb{E}\{Z_{\ell,t}^2\} = Q_{\ell,\ell}^t$, $\mathbb{E}\{Z_{\ell',t}^2\} = Q_{\ell',\ell'}^t$, $\mathbb{E}\{Z_{\ell,t} Z_{\ell',t}\} = Q_{\ell,\ell'}^t$. In order to prove the claim, it is therefore sufficient to show that

$$\lim_{\ell \to \infty} \sup_{\ell': \ell' \geq \ell} \left| Q_{\ell,\ell'}^t - \sigma_t^2 \right| = 0, \tag{A.38}$$

since this implies $\lim_{\ell \to \infty} \sup_{\ell': \ell' \geq \ell} [Q_{\ell,\ell}^t - 2Q_{\ell,\ell'}^t + Q_{\ell',\ell'}^t] = 0$, which in turn implies the Claim, via Eq. (A.36).

Finally, recall that η_t^ℓ was constructed using Theorem 9, cf. Eq. (A.26), in such a way that, for all $x \in \mathbb{R}$,

$$\left| \eta(x; \alpha\sigma_t) - \eta_t^\ell(x) \right| \leq \xi_\ell \exp\left\{ \frac{x^2}{16 \max(\sigma_t^2, s^2)} \right\}, \tag{A.39}$$

with $\xi_\ell \to 0$ as $\ell \to \infty$. The desired estimate (A.38) then follows by recalling that $\sigma_{t+1}^2 = (1/\delta)\mathbb{E}\{[\eta(X + \sigma_t Z) - X]^2\}$ and using Eq. (A.39) inductively to show that $\left| Q_{\ell,\ell'}^t - \sigma_t^2 \right| \leq K(t)\xi_\ell$. $\square$

We finally sketch the proof of Proposition 15.

Proof of Proposition 15. The sequence $\{x^t, z^t\}_{t \geq 0}$ is constructed as in the previous statement. The proof hence follow by using Corollary 16, and taking ξ small enough in Eq. (A.26), since we can ensure that $|\widetilde{R}_{t,s} - R_{t,s}| \leq \beta'$ for any $\beta' > 0$ and any $t, s \leq t_{\max}$ (as shown above for the case $t = s$). $\square$

B Proof of Lemma 5

Throughout the proof we denote by C_1, C_2, C_3 etc, positive constants that depend uniquely on $c_1, \ldots, c_3$.

Consider the ℓ_1 minimization problem

$$\begin{aligned} \text{minimize} \quad & \|x\|_1, \\ \text{subject to} \quad & y = Ax_0. \end{aligned}$$

and denote by $\widehat{x}$ any minimizer. Further, let v be a subgradient as in the statement, and define, for some $c \in (0, 1)$,

$$S(c) \equiv \left\{ i \in [n] : |v_i| \geq 1 - c \right\}. \tag{B.1}$$

Also, let $\overline{S}(c) = [n] \setminus S(c)$ be the complement of this set. Notice that, by definition of subgradient, we have $v_i = \mathrm{sign}(x_{0,i})$ for all $i \in S$ and $|v_{0,i}| \leq 1$ for all in $\overline{S} \equiv [n] \setminus S$. This implies that $S \subseteq S(c)$.

We have

$$\|\widehat{x}\|_1 = \|x_0\|_1 + \langle v, (\widehat{x} - x_0) \rangle + R_1 + R_2, \tag{B.2}$$

$$R_1 = \|\widehat{x}_{S(c)}\|_1 - \|x_{0,S(c)}\|_1 - \langle v_{S(c)}, (\widehat{x} - x_0)_{S(c)} \rangle, \tag{B.3}$$

$$R_2 = \|\widehat{x}_{\overline{S}(c)}\|_1 - \|x_{0,\overline{S}(c)}\|_1 - \langle v_{\overline{S}(c)}, (\widehat{x} - x_0)_{\overline{S}(c)} \rangle. \tag{B.4}$$

Since $\overline{S}(c) \subseteq \overline{S}$, we have $x_{0,\overline{S}(c)} = 0$ and hence

$$R_2 = \|\widehat{x}_{\overline{S}(c)}\|_1 - \langle v_{\overline{S}(c)}, \widehat{x}_{\overline{S}(c)} \rangle = \sum_{i \in \overline{S}(c)} \left(|\widehat{x}_i| - v_i \widehat{x}_i \right) \geq \sum_{i \in \overline{S}(c)} \left(|\widehat{x}_i| - (1 - c)|\widehat{x}_i| \right) = c\|\widehat{x}_{\overline{S}(c)}\|_1. \quad \text{(B.5)}$$

On the other hand, $v_{S(c)}$ is in the subgradient of $\|x_{S(c)}\|_1$ at $x_{S(c)} = x_{0,S(c)}$. Hence $R_1 \geq 0$. It follows that Eq. (B.2) implies $\|\widehat{x}\|_1 \geq \|x_0\|_1 + \langle v, (\widehat{x} - x_0) \rangle + c\|\widehat{x}_{\overline{S}(c)}\|_1$. Since $\widehat{x}$ is a minimizer, we thus get

$$\|\widehat{x}_{\overline{S}(c)}\|_1 \leq -\frac{1}{c}\langle v, (\widehat{x} - x_0) \rangle = -\frac{1}{c}\langle w, (\widehat{x} - x_0) \rangle \leq \frac{\varepsilon}{c}\sqrt{n}\,\|\widehat{x} - x_0\|_2, \tag{B.6}$$

where in the last step we used Cauchy-Schwarz together with assumption 1. Hereafter we let $r \equiv \widehat{x} - x_0$.

Let $\overline{S}(c) = \cup_{\ell=1}^K S_\ell$ be a partition such that $nc/2 \leq |S_\ell| \leq nc$, and that $|r_i| \leq |r_j|$ for each $i \in S_\ell$, $j \in S_{\ell-1}$. If $|\overline{S}(c)| < nc/2$, such a partition does not exist, but the argument follows by an obvious modification of the one below. Further define $\overline{S}_+ = \cup_{\ell=2}^K S_\ell \subseteq \overline{S}(c)$ and $S_+ = [n] \setminus \overline{S}_+$. We have

$$\|r_{\overline{S}_+}\|_2^2 = \sum_{\ell=2}^K \|r_{S_\ell}\|_2^2 \leq \sum_{\ell=2}^K |S_\ell| \left(\frac{\|r_{S_{\ell-1}}\|_1}{|S_{\ell-1}|} \right)^2 \leq \frac{4}{nc} \sum_{\ell=1}^{K-1} \|r_{S_\ell}\|_1^2 \leq \frac{4}{nc}\|r_{\overline{S}(c)}\|_1^2. \tag{B.7}$$

Fix $c = c_1$. Since $\overline{S}(c) \subseteq \overline{S}$, we have $r_{\overline{S}(c)} = \widehat{x}_{\overline{S}(c)}$ and using Eq. (B.6) we conclude that there exists $C_1 \leq 4/c_1^3$ such that

$$\|r_{\overline{S}_+}\|_2^2 \leq C_1\,\varepsilon^2\|r\|_2^2. \tag{B.8}$$

On the other hand, by definition $Ar = 0$, and hence $A_{S_+} r_{S_+} + A_{\overline{S}_+} r_{\overline{S}_+} = 0$. Since $\overline{S}(c) \subseteq \overline{S}$, we have $S \subseteq S(c) \subseteq S_+$. Further $S_+ \setminus S(c) = S_1$, whence $|S_+ \setminus S(c)| \leq nc = nc_1$. By assumption 2, we have $\sigma_{\min}(A_{S_+}) \geq c_2$ and therefore

$$\|r_{S_+}\|_2 \leq \frac{1}{c_2}\|A_{S_+} r_{S_+}\|_2 = \frac{1}{c_2}\|A_{\overline{S}_+} r_{\overline{S}_+}\|_2 \leq \frac{c_3}{c_2}\|r_{\overline{S}_+}\|_2.$$

Combining this with Eq. (B.8), we deduce that $\|r\|_2 \leq C_2\varepsilon\,\|r\|_2$ for some $C_2 = C_2(c_1, c_2, c_3)$, which in turns implies $r = 0$ provided that $C_2\varepsilon < 1$. The claim hence follows for $\varepsilon_0 = 1/[2C_2(c_1, c_2, c_3)]$.

C Asymptotic analysis of state evolution: Proof of Lemma 6

Before proceeding, we introduce the following piece of notation (following [BM12]). Fix a probability distribution p_X on $\mathbb{R}$, with $p_X(\{0\}) = 1 - \varepsilon$, and $\delta > 0$. For $\theta, \sigma^2 > 0$, we define

$$\mathsf{F}(\sigma^2, \theta) \equiv \frac{1}{\delta} \mathbb{E}\{[\eta(X + \sigma Z; \theta) - X]^2\}, \tag{C.1}$$

where expectation is taken with respect to the independent random variables $X \sim p_X$ and $Z \sim \mathsf{N}(0,1)$. When necessary, we will indicate the dependency on p_X by $\mathsf{F}(\sigma^2, \theta; p_X)$. With this notation the state evolution recursion reads $\sigma_{t+1}^2 = \mathsf{F}(\sigma_t^2, \alpha\sigma_t)$. The following properties of the function F were proved in [DMM09] (but see also [BM12], Appendix A for a more explicit treatment).

Lemma 7 ([DMM09]). *For any $\alpha > 0$, the mapping $\sigma^2 \mapsto \mathsf{F}(\sigma^2, \alpha\sigma)$ is monotone increasing and concave with $\mathsf{F}(0, 0) = 0$ and*

$$\frac{\mathrm{d}}{\mathrm{d}(\sigma^2)} \mathsf{F}(\sigma^2, \alpha\sigma)\Big|_{\sigma=0} = \frac{1}{\delta}\{\varepsilon(1 + \alpha^2) + 2(1 - \varepsilon)\mathbb{E}[(Z - \alpha)_+^2]\}. \tag{C.2}$$

It is also convenient to define

$$\begin{aligned} G_\varepsilon(\alpha) &\equiv \varepsilon(1 + \alpha^2) + 2(1 - \varepsilon)\mathbb{E}\{(Z - \alpha)_+^2\} \\ &= \varepsilon(1 + \alpha^2) + 2(1 - \varepsilon)[(1 + \alpha^2)\Phi(-\alpha) - \alpha\phi(\alpha)]. \end{aligned} \tag{C.3}$$

The first two derivatives of $\alpha \mapsto G_\varepsilon(\alpha)$ will be used in the proof

$$G_\varepsilon'(\alpha) = 2\alpha\varepsilon + 4(1 - \varepsilon)[-\phi(\alpha) + \alpha\Phi(-\alpha)], \tag{C.4}$$

$$G_\varepsilon''(\alpha) = 2\varepsilon + 4(1 - \varepsilon)\Phi(-\alpha). \tag{C.5}$$

In particular, we have the following.

Lemma 8. *For any $\varepsilon \in (0, 1)$, $\alpha \mapsto G_\varepsilon(\alpha)$ is strictly convex in $\alpha \in \mathbb{R}_+$, with a unique minimum on $\alpha_*(\varepsilon) \in (0, \infty)$. Further $G_\varepsilon(0) = 1$ and $\lim_{\alpha \to \infty} G_\varepsilon(\alpha) = \infty$. Finally, the minimum value satisfies*

$$G_\varepsilon(\alpha_*) = \varepsilon + 2(1 - \varepsilon)\Phi(-\alpha_*) = \frac{1}{2}G_\varepsilon''(\alpha_*) \in (0, 1). \tag{C.6}$$

Proof. By inspection of Eq. (C.5), $G_\varepsilon''(\alpha) > 0$ for all $\alpha > 0$, hence $G_\varepsilon(\alpha)$ is strictly convex. Further, from Eq. (C.4), we have $G_\varepsilon'(0) = -4(1 - \varepsilon)\phi(0) < 0$ and $G_\varepsilon'(\alpha) = 2\alpha\varepsilon + O_\alpha(1) > 0$ as $\alpha \to \infty$. Hence $\alpha \mapsto G_\varepsilon(\alpha)$ has a unique minimum $\alpha_*(\varepsilon) \in (0, \infty)$.

Finally, Eq. (C.6) follows immediately by using the condition $G_\varepsilon'(\alpha_*) = 0$ in the expression (C.3). □

In our proof it is more convenient to use the coordinates (δ, ε) instead of (ρ, δ). In terms of the latter, the phase boundary (1.2), (1.3) reads

$$\delta_*(\varepsilon) = \frac{2\phi(\alpha_*(\varepsilon))}{\alpha_*(\varepsilon) + 2[\phi(\alpha_*(\varepsilon)) - \alpha_*(\varepsilon)\Phi(-\alpha_*(\varepsilon))]}, \tag{C.7}$$

$$\alpha_*(\varepsilon) \quad \text{solves} \quad \alpha\varepsilon + 2(1 - \varepsilon)[\alpha\Phi(-\alpha) - \phi(\alpha)] = 0. \tag{C.8}$$

Notice that the use of the symbol $\alpha_*(\varepsilon)$ in the last equations is not an abuse of notation. Indeed comparing Eq. (C.8) with (C.4) we conclude that $\alpha_*(\varepsilon)$ is indeed the unique solution of $G_\varepsilon'(\alpha) = 0$. Further, comparing Eq. (C.7) with Eq. (C.3) we obtain the following.

Lemma 9. *Let $(\delta, \rho_*(\delta))$ be the phase boundary defined by Eqs. (1.2), (1.3). Then, for $\rho, \delta \in [0, 1]$, $\rho > \rho_*(\delta)$ if and only if, for $\varepsilon \in (0, 1)$, $\delta \in (\varepsilon, 1)$*

$$\delta < \delta_*(\varepsilon) \equiv \min_{\alpha > 0} G_\varepsilon(\alpha) \,. \tag{C.9}$$

Viceversa $\rho < \rho_*(\delta)$ if and only if $\delta > \delta_*(\varepsilon)$.

C.1   Proof of Lemma 6.(a): $\rho < \rho_*(\delta)$

Proof of Lemma 6.(a1). We set $\alpha = \alpha_*(\varepsilon) \equiv \arg\min_{\alpha \geq 0} G_\varepsilon(\alpha)$. Hence we have, by Lemma 7, and Lemma 9,

$$\frac{\mathrm{d}}{\mathrm{d}(\sigma^2)} \mathsf{F}(\sigma^2, \alpha_* \sigma)\Big|_{\sigma^2=0} = \frac{1}{\delta} \min_{\alpha > 0} G_\varepsilon(\alpha) = \frac{\delta_*(\varepsilon)}{\delta} \,. \tag{C.10}$$

In particular, by Lemma 9, for $\rho < \rho_*(\delta)$, we have $\frac{\mathrm{d}}{\mathrm{d}(\sigma^2)} \mathsf{F}(\sigma^2, \alpha_* \sigma) \equiv \omega_*(\varepsilon, \delta) \in (0, 1)$. Since, by Lemma 7, $\sigma^2 \mapsto \mathsf{F}(\sigma^2, \alpha_* \sigma)$ is concave, it follows that $\sigma_t^2 = B \omega_*^t[1 + o_t(1)]$.

Let $S \equiv \{\alpha \in \mathbb{R}_+ : G_\varepsilon(\alpha)/\delta < 1\}$. Since $\alpha \mapsto G_\varepsilon(\alpha)$ is strictly convex by Lemma 8, with $G_\varepsilon(0), G_\varepsilon(\infty) > \delta$, we have $S = (\alpha_1, \alpha_2)$ with $0 < \alpha_1 < \alpha_* < \alpha_2 < \infty$. Let $\omega(\alpha) = G_\varepsilon(\alpha)/\delta$. Fixing $\alpha \in (\alpha_1, \alpha_2)$, by concavity of $\sigma^2 \mapsto \mathsf{F}(\sigma^2, \alpha\sigma)$, we have $\sigma_t^2 = B \omega(\alpha)^t[1 + o_t(1)]$. Finally, by continuity of $\alpha \mapsto G_\varepsilon(\alpha)$, we have $\{\omega(\alpha) : \alpha \in (\alpha_1, \alpha_2)\} = [\omega_*, 1)$ and hence any rate $\omega \in [\omega_*, 1)$ can be realized.

Finally by Lemma 8 $G_\varepsilon(\alpha_*) \equiv \varepsilon + 2(1 - \varepsilon)\Phi(-\alpha_*) < \delta$. Since $\alpha \mapsto \varepsilon + 2(1 - \varepsilon)\Phi(-\alpha)$ is decreasing in α, the last claim follows. $\qquad\square$

In the proof of part (a2) we will make use of the following analytical result.

Lemma 10. *For $\varepsilon \in (0, 1)$, $\alpha \geq \alpha_*(\varepsilon)$, consider the function $\mathcal{F}_{\alpha,\varepsilon} : [0, 1] \to \mathbb{R}$ defined by*

$$\mathcal{F}_{\alpha,\varepsilon}(Q) \equiv \frac{1}{G_\varepsilon(\alpha)} \mathbb{E}\Big\{[\eta(X_\infty + Z_1; \alpha) - X_\infty][\eta(X_\infty + Z_2; \alpha) - X_\infty]\Big\}, \tag{C.11}$$

where expectation is taken with respect to X_∞, $\mathbb{P}\{X_\infty = 0\} = 1 - \varepsilon$, $\mathbb{P}\{X_\infty \in \{+\infty, -\infty\}\} = \varepsilon$, and the independent Gaussian vector (Z_1, Z_2) with mean zero and covariance $\mathbb{E}\{Z_1^2\} = \mathbb{E}\{Z_2^2\} = 1$, $\mathbb{E}\{Z_1 Z_2\} = Q$. (The mapping $x \mapsto [\eta(x + a; b) - x]$ is here extended to $x = +\infty, -\infty$ by continuity for any a, b bounded.)

Then $\mathcal{F}_{\alpha,\varepsilon}$ is increasing and convex on $[0, 1]$ with $\mathcal{F}_{\alpha,\varepsilon}(1) = 1$ and $\mathcal{F}'_{\alpha,\varepsilon}(1) < 1$. In particular $\mathcal{F}_{\alpha,\varepsilon}(Q) > Q$ for all $\in [0, 1)$

Proof. It is convenient to change variables and let $Q = e^{-s}$. If we let $\{U_s\}_{s \in \mathbb{R}}$ denote the standard Ornstein-Uhlenbeck process, $\mathrm{d}U_s = -U_s \mathrm{d}s + \sqrt{2}\,\mathrm{d}B_s$ with $\{B_s\}_{s \in \mathbb{R}}$ the standard Brownian motion. Then $\mathcal{F}_{\alpha,\varepsilon}(Q) = \widehat{\mathcal{F}}_{\alpha,\varepsilon}(-\log(Q))$, with

$$\widehat{\mathcal{F}}_{\alpha,\varepsilon}(s) \equiv \frac{1}{G_\varepsilon(\alpha)} \mathbb{E}\Big\{[\eta(X_\infty + U_0; \alpha) - X_\infty][\eta(X_\infty + U_s; \alpha) - X_\infty]\Big\}. \tag{C.12}$$

A simple calculation yields

$$\frac{\mathrm{d}}{\mathrm{d}s} \widehat{\mathcal{F}}_{\alpha,\varepsilon}(s) = -\frac{1}{G_\varepsilon(\alpha)} \mathbb{E}\Big\{\eta'(X_\infty + U_0; \alpha)\eta'(X_\infty + U_s; \alpha)\Big\} e^{-s}, \tag{C.13}$$

48

where $\eta'(\,\cdot\,;\alpha)$ denotes the derivative of η with respect to its first argument. By the spectral decomposition of the Ornstein-Uhlenbeck process, we have, for any function $\psi \in L_2(\mathbb{R})$

$$\mathbb{E}\{\psi(U_0)\psi(U_s)\} = \sum_{k=1}^{\infty} e^{-\lambda_k s} c_k(\psi)^2 \,, \tag{C.14}$$

for some non-negative $\{\lambda_k\}_{k \geq 1}$. In particular $e^s \frac{\mathrm{d}}{\mathrm{d}s} \widehat{\mathcal{F}}_{\alpha,\varepsilon}(s)$ is strictly negative and increasing in s. We therefore obtain

$$\frac{\mathrm{d}}{\mathrm{d}Q} \mathcal{F}_{\alpha,\varepsilon}(Q) = \frac{1}{G_\varepsilon(\alpha)} \mathbb{E}\{\eta'(X_\infty + Z_1;\alpha)\eta'(X_\infty + Z_2;\alpha)\} \,, \tag{C.15}$$

Which is strictly positive and increasing in Q. Hence $Q \mapsto \mathcal{F}_{\alpha,\varepsilon}(Q)$ is increasing and strictly convex. Finally, since $\eta'(y;\alpha) = \mathbf{1}(|y| \geq \alpha)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}Q} \mathcal{F}_{\alpha,\varepsilon}(Q)\bigg|_{Q=1} = \frac{1}{G_\varepsilon(\alpha)} \mathbb{P}\{|X_\infty + Z| > \alpha\} = \frac{1}{G_\varepsilon(\alpha)} \{\varepsilon + 2(1-\varepsilon)\Phi(-\alpha)\} = \frac{G_\varepsilon''(\alpha)}{2G_\varepsilon(\alpha)} \,. \tag{C.16}$$

Since by Lemma 8 $\alpha \mapsto G_\varepsilon(\alpha)$ is strictly increasing over $(\alpha_*(\varepsilon),\infty)$ and by Eq. (C.5) $\alpha \mapsto G_\varepsilon''(\alpha)$ is strictly decreasing over $\mathbb{R}_+$, we have

$$\frac{\mathrm{d}}{\mathrm{d}Q} \mathcal{F}_{\alpha,\varepsilon}(Q)\bigg|_{Q=1} < \frac{G_\varepsilon''(\alpha_*(\varepsilon))}{2G_\varepsilon(\alpha_*(\varepsilon))} = 1 \,, \tag{C.17}$$

where the last equality follows again by Lemma 8. This conclude the proof. $\qquad\square$

We are now in position to prove part $(a2)$ of Lemma 6.

Proof of Lemma 6.$(a2)$. Throughout the proof we fix $\alpha \in (\alpha_*(\varepsilon,\delta),\alpha_2(\varepsilon,\delta))$. Let the sequence $\{\sigma_t^2\}_{t \geq 0}$ be given as per the state evolution equation (5.8). Define $Q_t \equiv R_{t,t-1}/(\sigma_t\sigma_{t-1})$. By Proposition 15, Q_t is the covariance of two gaussian random variables of variance 1. Hence $|Q_t| \leq 1$. Using Eq. (5.15) we further have

$$Q_{t+1} = \mathcal{F}_t(Q_t) \,, \tag{C.18}$$

$$\mathcal{F}_t(Q) = \frac{\sigma_{t-1}}{\delta\sigma_{t+1}} \mathbb{E}\left\{ \left[\eta\Big(\frac{X}{\sigma_t} + Z_1;\alpha\Big) - \frac{X}{\sigma_t}\right]\left[\eta\Big(\frac{X}{\sigma_{t-1}} + Z_2;\alpha\Big) - \frac{X}{\sigma_{t-1}}\right]\right\} \,, \tag{C.19}$$

were expectation is taken with respect to $X \sim p_X$ and the independent Gaussian random vector (Z_1, Z_2) with zero mean and covariance $\mathbb{E}\{Z_1^2\} = 1$, $\mathbb{E}\{Z_2^2\} = 1$, $\mathbb{E}\{Z_1 Z_2\} = Q_t$. By induction it is easy to check that $Q_t \geq 0$ for all t.

For $\alpha \in (\alpha_1, \alpha_2)$, by part $(a1)$ we have $\sigma_t \to 0$. Hence X/σ_t converges in distribution (over the completed real line) to a random variable $X_\infty \sim (1-\varepsilon)\delta_0 + \varepsilon_+\delta_{+\infty} + \varepsilon_-\delta_{-\infty}$ where $\varepsilon_+ \equiv \mathbb{P}\{X > 0\}$, $\varepsilon_- \equiv \mathbb{P}\{X < 0\}$, $\varepsilon = \varepsilon_+ + \varepsilon_-$. Hence the expectation in Eq. (C.19) converges pointwise to

$$\mathbb{E}\left\{ \big[\eta\big(X_\infty + Z_1;\alpha\big) - X_\infty\big]\big[\eta\big(X_\infty + Z_2;\alpha\big) - X_\infty\big]\right\} \,. \tag{C.20}$$

(Notice that this expectation depends on the distribution of X_∞ only through ε, because of the symmetry properties of the function η.)

Further, by the proof of part $(a1)$, as $t \to \infty$ we have $\sigma_t^2 \to 0$ and

$$\sigma_{t+1}^2 = \left.\frac{\mathrm{d}}{\mathrm{d}(\sigma^2)}\mathsf{F}(\sigma^2, \alpha_* \sigma)\right|_{\sigma=0} \sigma_t^2 + o(\sigma_t^2) = \frac{1}{\delta}\, G_\varepsilon(\alpha_*)\, \sigma_t^2 + o(\sigma_t^2)\,. \tag{C.21}$$

Hence

$$\lim_{t\to\infty} \frac{\sigma_{t-1}}{\sigma_{t+1}} = \frac{\delta}{G_\varepsilon(\alpha)}\,. \tag{C.22}$$

Comparing Eqs. (C.11) and (C.19) we conclude that, for any $Q \in [0, 1]$

$$\lim_{t\to\infty} \mathcal{F}_t(Q) = \mathcal{F}_{\alpha,\varepsilon}(Q)\,. \tag{C.23}$$

Further the convergence is uniform, since the functions $\mathcal{F}_t$ are uniformly Lipschitz (see proof of Lemma 10 above).

Consider now the sequence $\{Q_t\}_{t\geq 0}$ and let $Q_* = \liminf_{t\to\infty} Q_t$. Since $Q_t \in [0, 1]$ for all t, we have $Q_* \in [0, 1]$ as well. We claim that in fact $Q_* = 1$ and therefore $\lim_{t\to\infty} Q_t = 1$, which implies the thesis.

In order to prove the claim, let $\{Q_{t(k)}\}_{k\in\mathbb{N}}$ be a subsequence that converges to Q_*. Then

$$Q_* = \lim_{k\to\infty} \mathcal{F}_{t(k)-1}(Q_{t(k)-1}) = \lim_{k\to\infty} \mathcal{F}_{\alpha,\varepsilon}(Q_{t(k)-1}) \geq \mathcal{F}_{\alpha,\varepsilon}(\lim_{k\to\infty} \inf Q_{t(k)-1}) \geq \mathcal{F}_{\alpha,\varepsilon}(Q_*)\,, \quad (\text{C.24})$$

where, in the last step, we used the fact that $\mathcal{F}_{\alpha,\varepsilon}(\cdot)$ is monotone increasing. Since $\mathcal{F}_{\alpha,\varepsilon}(q) > q$ for all $q \in [0, 1)$ by Lemma 10, we conclude that $Q_* = 1$. $\square$

Before proving $(a3)$ of Lemma 6, we establish one more technical result.

Lemma 11. *Let p_X be a probability measure on the real line such that $p_X(\{0\}) = 1 - \varepsilon$ and $\mathbb{E}_{p_X}\{X^2\} < \infty$, Assume p_X to be such that $\max(p_X((0, a)), p_X((-a, 0))) \leq Ba^b$ for some $B, b > 0$. Then, letting $X_\infty \sim (1 - \varepsilon)\delta_0 + \varepsilon_+\delta_{+\infty} + \varepsilon_-\delta_{-\infty}$ (with the notation introduced above, namely, $\varepsilon_+ = p_X(0, +\infty)$ and $\varepsilon_- = p_X(-\infty, 0)$):*

$$\left| \mathbb{E}\left\{ \left[\eta\left(\frac{X}{\sigma_t} + Z_1; \alpha\right) - \frac{X}{\sigma_t} \right] \left[\eta\left(\frac{X}{\sigma_{t-1}} + Z_2; \alpha\right) - \frac{X}{\sigma_{t-1}} \right] \right\} \right. \tag{C.25}$$

$$\left. - \mathbb{E}\left\{ \left[\eta\left(X_\infty + Z_1; \alpha\right) - X_\infty \right] \left[\eta\left(X_\infty + Z_2; \alpha\right) - X_\infty \right] \right\} \right| \leq B'(\sigma_t^b + \sigma_{t-1}^b)\,,$$

for an eventually different constant B'. Here expectation is taken with respect to $X \sim p_X$ and the independent Gaussian random vector (Z_1, Z_2) with zero mean and covariance $\mathbb{E}\{Z_1^2\} = 1$, $\mathbb{E}\{Z_2^2\} = 1$, $\mathbb{E}\{Z_1 Z_2\} = Q_t$, and

$$\mathsf{F}(\sigma^2, \theta) = \left.\frac{\mathrm{d}\mathsf{F}}{\mathrm{d}(\sigma^2)}(\sigma^2; \alpha\sigma)\right|_{\sigma=0} \sigma^2 + O(\sigma^{2+b})\,. \tag{C.26}$$

Proof. By triangular inequality, the left hand side of Eq. (C.25) can be upper bounded as $D_1 + D_2$ whereby

$$D_1 \equiv \mathbb{E}\left\{ \left[\eta\left(\frac{X}{\sigma_t} + Z_1; \alpha\right) - \frac{X}{\sigma_t} - \eta\left(X_\infty + Z_1; \alpha\right) + X_\infty \right] \left[\eta\left(\frac{X}{\sigma_{t-1}} + Z_2; \alpha\right) - \frac{X}{\sigma_{t-1}} \right] \right\}\,,$$

$$D_2 \equiv \mathbb{E}\left\{ \left[\eta\left(X_\infty + Z_1; \alpha\right) - X_\infty \right] \left[\eta\left(\frac{X}{\sigma_{t-1}} + Z_2; \alpha\right) - \frac{X}{\sigma_{t-1}} - \eta\left(X_\infty + Z_2; \alpha\right) + X_\infty \right] \right\}\,.$$

50

Here X and X_∞ are coupled in such a way that $X = 0$ if and only if $X_\infty = 0$ and the two variables have the same sign in the other case. We focus on bounding D_1 since D_2 can be treated along the same lines. Letting $R(x; \theta) = \eta(x; \theta) - x$, we have

$$D_1 = \mathbb{E}\left\{\left[R\left(\frac{X}{\sigma_t} + Z_1; \alpha\right) - R(X_\infty + Z_1; \alpha)\right]\left[R\left(\frac{X}{\sigma_{t-1}} + Z_2; \alpha\right) + Z_2\right]\right\} = D_{1,a} + D_{1,b},$$

$$D_{1,a} = \mathbb{E}\left\{\left[R\left(\frac{X}{\sigma_t} + Z_1; \alpha\right) - R(X_\infty + Z_1; \alpha)\right]R\left(\frac{X}{\sigma_{t-1}} + Z_2; \alpha\right)\right\},$$

$$D_{1,b} = Q_t\,\mathbb{E}\left\{\left[R'\left(\frac{X}{\sigma_t} + Z_1; \alpha\right) - R'(X_\infty + Z_1; \alpha)\right]\right\},$$

where in the last line we used Stein's lemma to integrate over Z_2, and R' denotes derivative with respect to the first argument. Once again the two terms are treated along the same lines, and we will only consider $D_{1,a}$. We have

$$|D_{1,a}| \leq \alpha\mathbb{E}\left\{\left|R\left(\frac{X}{\sigma_t} + Z_1; \alpha\right) - R(X_\infty + Z_1; \alpha)\right|\right\}$$

$$\leq \alpha\varepsilon_+\mathbb{E}\left\{\left|R\left(\frac{X_+}{\sigma_t} + Z_1; \alpha\right) - R(+\infty; \alpha)\right|\right\} + \alpha\varepsilon_-\mathbb{E}\left\{\left|R\left(\frac{X_-}{\sigma_t} + Z_1; \alpha\right) - R(-\infty; \alpha)\right|\right\}, \quad \text{(C.27)}$$

where X_+ (resp. X_-) is distributed as X conditioned on $X > 0$ (resp. $X < 0$). The function $x \mapsto R(x; \alpha) - R(\infty; \alpha)$ is monotone decreasing, equal to 2α for $x \leq -\alpha$ and to 0 for $x \geq \alpha$. Hence $\widetilde{R}(x) \equiv \mathbb{E}_{Z_1}\{|R(x + Z_1; \alpha) - R(+\infty; \alpha)|\}$ is monotone decreasing, takes values in $(0, 2\alpha)$ and upper bounded by $Ce^{-x^2/4}$ for $x \geq 0$. Denoting by F_+ the distribution of X_+, we have

$$\mathbb{E}\left\{\left|R\left(\frac{X_+}{\sigma_t} + Z_1; \alpha\right) - R(+\infty; \alpha)\right|\right\} = \mathbb{E}\widetilde{R}(X_+/\sigma_t) = \int_0^\infty |\widetilde{R}'(x)|\,F(x\sigma_t)\,\mathrm{d}x \leq B'\sigma_t^b.$$

The other term in Eq. (C.27) is bounded by the same argument. This concludes the proof of Eq. (C.25).

The proof of Eq. (C.26) follows from Eq. (C.25) if we notice that

$$\mathsf{F}(\sigma^2, \alpha\sigma) = \frac{\sigma^2}{\delta}\,\mathbb{E}\left\{\left[\eta\left(\frac{X}{\sigma} + Z; \alpha\right) - X\right]^2\right\},$$

$$\left.\frac{\mathrm{d}\mathsf{F}}{\mathrm{d}(\sigma^2)}(\sigma^2; \alpha\sigma)\right|_{\sigma=0} = \mathbb{E}\left\{\left[\eta(X_\infty + Z; \alpha) - X_\infty\right]^2\right\}.$$

$\square$

The last lemma has a useful consequence that we will exploit in the ensuing proof of Lemma 6.(a3).

Corollary 18. *Let $\mathcal{F}_{\alpha,\varepsilon}(Q)$ be defined as per Eq. (C.11) and $\mathcal{F}_t(Q)$ defined as per Eq. (C.19) with p_X, α, ε satisfying the conditions of Lemma 6.(a3). Then there exists a constants $B, B', b > 0$ depending on p_X such that*

$$\sup_{Q\in[0,1]}\left|\mathcal{F}_t(Q) - \mathcal{F}_{\alpha,\varepsilon}(Q)\right| \leq B\,\sigma_t^b \leq B'\omega^{bt/2}.$$

Proof. The second inequality follows from the first one using Lemma 6.(*a*1). Using Eq. (C.26), we have

$$\frac{\sigma_{t-1}^2}{\sigma_{t+1}^2} = \frac{\sigma_t^2}{\mathsf{F}(\sigma_t^2; \alpha\sigma_t)} \cdot \frac{\sigma_{t-1}^2}{\mathsf{F}(\sigma_{t-1}^2; \alpha\sigma_{t-1})} = \frac{\delta^2}{G_\varepsilon(\alpha)^2}\left\{1 + O(\sigma_t^b, \sigma_{t-1}^b)\right\}.$$

The proof of the corollary is obtained by noting that $\sigma_t = \Theta(\sigma_{t-1})$ and applying Eq. (C.25) to the expectation in Eq. (C.19). $\qquad\square$

Proof of Lemma 6.(a3). Define, as in the proof of part (*a*2), $Q_t \equiv R_{t,t-1}/(\sigma_t\sigma_{t-1})$, and recall that

$$Q_{t+1} = \mathcal{F}_t(Q_t).$$

By Corollary 18, and Lemma 10, it follows that $Q_t \geq 1 - A\overline{\omega}^{2t}$ for some constants $A > 0$, $\overline{\omega} \in (0,1)$. Indeed

$$Q_{t+1} \geq \mathcal{F}_{\alpha,\varepsilon}(Q_t) - B'\omega^{bt/2} \geq 1 - B'\omega^{bt/2} - \mathcal{F}'_{\alpha,\varepsilon}(1)(1 - Q_t).$$

and the claim follows by noting that $\mathcal{F}'_{\alpha,\varepsilon}(1) \in (0,1)$ by Lemma 10.

Next, consider a sequence of centered Gaussian random variables $(Z_t)_{t\geq 0}$ with covariance $\mathbb{E}\{Z_t Z_s\} = R_{t,s}$. By triangular inequality, we have, for any $t < s$,

$$\left(2 - 2\frac{R_{t,s}}{\sigma_t\sigma_s}\right)^{1/2} = \mathbb{E}\left\{\left(\frac{Z_t}{\sigma_t} - \frac{Z_s}{\sigma_s}\right)^2\right\}^{1/2} \leq \sum_{k=t+1}^{s} \mathbb{E}\left\{\left(\frac{Z_k}{\sigma_k} - \frac{Z_{k-1}}{\sigma_{k-1}}\right)^2\right\}^{1/2} = \sum_{k=t+1}^{s}(2 - 2Q_k)^{1/2} \leq A'\overline{\omega}^t.$$
(C.28)

Next consider the quantity in Eq. (5.19). We have

$$\sup_{t,s\geq t_0} \mathbb{P}\left\{|X + Z_s| \geq c\,\sigma_s\,;\ |X + Z_t| < c\,\sigma_t\right\}$$

$$\leq \sup_{t\geq t_0} \mathbb{P}\left\{|X + Z_t| < c\,\sigma_t;\ X \neq 0\right\} + \sup_{t,s\geq t_0} \mathbb{P}\left\{\left|Z_s/\sigma_s\right| \geq c\,;\ \left|Z_t/\sigma_t\right| < c\,;\ X = 0\right\}$$

$$= \sup_{t\geq t_0} \mathbb{P}\left\{|X/\sigma_t + \widetilde{Z}_t| < c\,;\ X \neq 0\right\} + \sup_{t,s\geq t_0} \mathbb{P}\left\{\left|\widetilde{Z}_s\right| \geq c\,;\ \left|\widetilde{Z}_t\right| < c\,\right\},$$ (C.29)

where $(\widetilde{Z}_s, \widetilde{Z}_t)$ are Gaussian with $\mathbb{E}\{\widetilde{Z}_t^2\} = \mathbb{E}\{\widetilde{Z}_s^2\} = 1$, and $\mathbb{E}\{\widetilde{Z}_s\widetilde{Z}_t\} = R_{t,s}/(\sigma_t\sigma_s)$. The first term in Eq. (C.29) vanishes as $t_0 \to \infty$ since $\sigma_t \to 0$ as $t \to \infty$, and the second vanishes by Eq. (C.28). $\quad\square$

C.2 Proof of Lemma 6.(*b*): $\rho > \rho_*(\delta)$

Proof of Lemma 6.(b1), (b2). First notice that, with the definitions given in the previous section

$$\lim_{\sigma^2\to\infty} \frac{\mathrm{d}}{\mathrm{d}(\sigma^2)}\mathsf{F}(\sigma^2, \alpha_*\sigma) = \frac{2}{\delta}\mathbb{E}\left\{(Z - \alpha)_+^2\right\}$$

$$= \frac{2}{\delta}\left\{(1 + \alpha^2)\Phi(-\alpha) - \alpha\phi(\alpha)\right\}.$$

Notice that the right hand side is equal to $2/\delta$ for $\alpha = 0$, monotonically decreasing in α, and vanishing as $\alpha \to \infty$. Hence there exists $\alpha_{\min}(\varepsilon, \delta)$ such that the right hand side is smaller than 1 if and only if

$\alpha > \alpha_{\min}(\varepsilon, \delta)$. Further, $\sigma^2 \mapsto \mathsf{F}(\sigma^2, \alpha\sigma)$ is concave with $\mathsf{F}(0,0) = 0$ and first derivative larger than 1 at $\sigma^2 = 0$ (cf. Lemma 7). It follows that for $\alpha > \alpha_{\min}(\varepsilon, \delta)$ there exists a unique $\sigma_*(\delta, p_X)$ such that $\mathsf{F}(\sigma^2, \alpha\sigma) > \sigma^2$ for all $\sigma \in (0, \sigma_*)$ and $\mathsf{F}(\sigma^2, \alpha\sigma) < \sigma^2$ for $\sigma \in (\sigma_*, \infty)$. It follows that $\sigma_t^2 \to \sigma_*$ for any $\sigma_0^2 \neq 0$. This proves the first part of claim $(b1)$.

Letting $\sigma_*^2 = \sigma_*^2(\alpha)$, it is easy to check that $\alpha \mapsto \sigma_*^2(\alpha)$ is continuous for $\alpha \in (\alpha_{\min}, \infty)$ with $\lim_{\alpha \to \alpha_{\min}} \sigma_*^2(\alpha) = +\infty$ (the limit being taken from the left), and $\lim_{\alpha \to \infty} \sigma_*^2(\alpha) = +\mathbb{E}\{X^2\}/\delta > 0$. As a consequence

$$\lim_{\alpha \to \alpha_{\min}} \mathbb{P}\{|X + \sigma_* Z| \geq \alpha\sigma_*\} = 2\Phi(-\alpha_{\min}), \tag{C.30}$$

$$\lim_{\alpha \to \infty} \mathbb{P}\{|X + \sigma_* Z| \geq \alpha\sigma_*\} = 0. \tag{C.31}$$

Notice that by the definition of $\alpha_{\min}$ given above, we have

$$2\Phi(-\alpha_{\min}) - 2\alpha_{\min}\{\phi(\alpha_{\min}) - \alpha_{\min}\Phi(-\alpha_{\min})\} = \delta.$$

Since $\phi(z) > z\Phi(-z)$ for $z > 0$, it follows that $\lim_{\alpha \to \alpha_{\min}} \mathbb{P}\{|X + \sigma_* Z| \geq \alpha\sigma_*\} > \delta$. We define

$$\alpha_0(\delta, p_X) \equiv \sup\{\alpha > \alpha_{\min}(\varepsilon, \delta) : \mathbb{P}\{|X + \sigma_* Z| \geq \alpha\sigma_*\} \geq \delta\}. \tag{C.32}$$

By the above $\alpha_0 \in (\alpha_{\min}, \infty)$. Further, by continuity, for $\alpha = \alpha_0$, $\mathbb{P}\{|X + \sigma_* Z| \geq \alpha\sigma_*\} = \delta$. We thus proved claim $(b2)$.

In order to prove the second statement in $(b1)$, we proceed analogously to part $(a2)$, and define $Q_t \equiv R_t/(\sigma_t \sigma_{t-1})$. This sequence satisfies the recursion (C.18) with $\mathcal{F}_t$ defined as per Eq. (C.19). As $t \to \infty$ we have $\sigma_t \to \sigma_*$ and hence $\mathcal{F}_t$ converges uniformly to a limit that we denote by an abuse of notation $\mathcal{F}_{\alpha, \delta, p_X}$, where

$$\mathcal{F}_{\alpha, \delta, p_X}(Q) \equiv \frac{1}{\delta}\mathbb{E}\left\{\left[\eta\left(\frac{X}{\sigma_*} + Z_1; \alpha\right) - \frac{X}{\sigma_*}\right]\left[\eta\left(\frac{X}{\sigma_*} + Z_2; \alpha\right) - \frac{X}{\sigma_*}\right]\right\} \tag{C.33}$$

Proceeding as in the proof of Lemma 10, we conclude that $Q \mapsto \mathcal{F}_{\alpha, \delta, p_X}(Q)$ is increasing and convex on $[0, 1]$. Further (for $Z \sim \mathsf{N}(0, 1)$)

$$\mathcal{F}_{\alpha, \delta, p_X}(1) = \frac{1}{\delta}\mathbb{E}\left\{\left[\eta\left(\frac{X}{\sigma_*} + Z_1; \alpha\right) - \frac{X}{\sigma_*}\right]^2\right\} = \frac{1}{\sigma_*^2}\mathsf{F}(\sigma_*^2, \alpha\sigma_*) = 1. \tag{C.34}$$

Finally, for $\alpha \geq \alpha_0(\delta, p_X)$,

$$\frac{\mathrm{d}}{\mathrm{d}Q}\mathcal{F}_{\alpha, \delta, p_X}(Q)\bigg|_{Q=1} = \frac{1}{\delta}\mathbb{P}\left\{\left|\frac{X}{\sigma_*} + Z_1\right| > \alpha\right\} \leq 1, \tag{C.35}$$

and therefore $\mathcal{F}_{\alpha, \delta, p_X}(Q) > Q$ for all $Q \in [0, 1)$. Hence, proceeding again as in the proof of part $(a2)$ we conclude that $\lim_{t \to \infty} Q_t = 1$ and therefore $\lim_{t \to \infty} R_{t, t-1} = \sigma_*^2$ as claimed. $\square$

Proof of Lemma 6.(b3). Throughout this proof we fix $p_X = (1 - \varepsilon)\delta_0 + \varepsilon\gamma$, $\delta \in (\varepsilon, \delta_*(\varepsilon))$. By part $(b1)$, we have $\lim_{t \to \infty} \mathbb{E}\{|\eta(X + \sigma_t X; \alpha\sigma_t)|\} = \mathbb{E}\{|\eta(X + \sigma_* Z; \alpha\sigma_*)|\}$. It is therefore sufficient to prove that $\mathbb{E}\{|\eta(X + \sigma_* Z; \alpha\sigma_*)|\} < \mathbb{E}\{|X|\}$.

Consider the function $\mathcal{E} : (\sigma^2, \theta) \mapsto \mathcal{E}(\sigma^2, \theta)$ defined on $\mathbb{R}_+ \times \mathbb{R}_+)$ by

$$\mathcal{E}(\sigma^2, \theta) \equiv -\frac{1}{2}(1 - \delta)\frac{\sigma^2}{\theta} + \mathbb{E}\min_{s \in \mathbb{R}}\left\{\frac{1}{2\theta}(s - X - \sigma Z)^2 + |s|\right\}, \tag{C.36}$$

where expectation is taken with respect to $X \sim p_X$ and $Z \sim \mathsf{N}(0,1)$. Notice that the minimum over $s \in \mathbb{R}$ is uniquely achieved at $s = \eta(X + \sigma Z; \theta)$. It is not hard to compute the partial derivatives

$$\frac{\partial \mathcal{E}}{\partial \theta}(\sigma^2, \theta) = -\frac{\delta}{2\theta^2}\Big\{\Big(1 - \frac{2}{\delta}\mathbb{P}\{|X + \sigma Z| \geq \theta\}\Big)\sigma^2 + \mathsf{F}(\sigma^2, \theta)\Big\}, \tag{C.37}$$

$$\frac{\partial \mathcal{E}}{\partial \sigma^2}(\sigma^2, \theta) = \frac{\delta}{2\theta}\Big\{1 - \frac{1}{\delta}\mathbb{P}\{|X + \sigma Z| \geq \theta\}\Big\}, \tag{C.38}$$

where $\mathsf{F}(\sigma^2, \theta)$ is defined as per Eq. (C.1). Using these expressions in Eq. (C.36) we conclude that

$$\frac{\partial \mathcal{E}}{\partial \theta}(\sigma^2, \theta) = \frac{\partial \mathcal{E}}{\partial \sigma^2}(\sigma^2, \theta) = 0 \Rightarrow \mathcal{E}(\sigma^2, \theta) = \mathbb{E}\{|\eta(X + \sigma Z; \theta)|\} \tag{C.39}$$

In particular, one can check from Eqs. (C.37), (C.37) that a stationary point[4] is given by setting $\sigma = \sigma_*(\delta, p_X)$ and $\theta = \theta_*(\delta, p_X) \equiv \alpha_0(\delta, p_X)\sigma_*(\delta, p_X)$.

Define $E(\sigma^2) = \mathcal{E}(\sigma^2, \alpha_0(\delta, p_X)\sigma)$. Using again Eqs. (C.37), (C.38) we get

$$\frac{\mathrm{d}E}{\mathrm{d}\sigma^2}(\sigma^2) = \frac{\delta}{4\alpha\sigma^3}\big\{\sigma^2 - \mathsf{F}(\sigma^2, \alpha_0\sigma)\big\}. \tag{C.40}$$

In particular, as a consequence of Lemma 7, and of the analysis at point $(b1)$, we have $\frac{\mathrm{d}E}{\mathrm{d}\sigma^2} < 0$ for $\sigma^2 \in (0, \sigma_*^2)$ (C.37). Therefore, setting $\alpha = \alpha_0(\delta, p_X)$, we have

$$\mathbb{E}\{|\eta(X + \sigma_* Z; \alpha\sigma_*)|\} = E(\sigma_*^2) < \lim_{\sigma \to 0} E(\sigma^2)$$

$$= -\lim_{\sigma \to 0}\frac{1}{2\alpha}\sigma(1 - \delta) + \lim_{\sigma \to 0}\frac{\sigma}{2\alpha}\mathbb{E}\Big\{\Big[\eta\Big(\frac{X}{\sigma} + Z; \alpha\Big) - \frac{X}{\sigma} - Z\Big]^2\Big\} + \lim_{\sigma \to 0}\mathbb{E}\{|\eta(X + \sigma Z; \alpha\sigma)|\}$$

$$= \lim_{\sigma \to 0}\frac{\sigma}{2\alpha}\alpha^2 + \mathbb{E}\{|X|\} = \mathbb{E}\{|X|\}.$$

This concludes the proof. $\qquad\square$

D   Reference results

The following calculus fact is used in the main text.

Lemma 12. *For all $s, x > 0$ we have $x^s \leq \big(\frac{s}{e}\big)^s e^x$.*

Proof. Since $f(x) = \ln(x)$ for $x > 0$ is concave, when $x \geq s$ then

$$\frac{\ln(x) - \ln(s)}{x - s} \leq f'(s) = \frac{1}{s} \tag{D.1}$$

This is equivalent to $(x/s)^s \leq e^{x-s}$ which proves the result. The case of $x < s$ is proved similarly. $\quad\square$

We also use an estimate on the minimum singular value of perturbed rectangular matrices, which was proved in [BC10, Theorem 1.1].

Theorem 10. *For $M, N \in \mathbb{N}$, $N \leq (1 - a)M$, let $B \in \mathbb{R}^{M \times N}$, $\|B\|_2 \leq 1/a$ be any deterministic matrix and $G \in \mathbb{R}^{M \times N}$ be a matrix with i.i.d. entries $G_{ij} \sim \mathsf{N}(0, 1/M)$. Then there exist constants a_1, a_2 depending only on a and bounded for $a > 0$ such that, for all $z < a_2$,*

$$\mathbb{P}\Big\{\sigma_N(A + \nu G) \leq \nu z\Big\} \leq (a_1 z)^{M-N+1}. \tag{D.2}$$

---

[4]Indeed this is the unique saddle point of the function $(\theta^{-1}, \sigma^2) \mapsto \mathcal{E}(\theta, \sigma^2)$ as it can be proved by the general minimax theorem.

References

[AGZ09]   G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge University Press, 2009.

[ALPTJ11] R. Adamczak, A.E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, *Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling*, Constructive Approximation (2011), 61–88.

[AS92]   R. Affentranger and R. Schneider, *Random projections of regular simplices*, Discr. and Comput. Geometry **7** (1992), 219–226.

[BC10]   P. Buergisser and F. Cucker, *Smoothed analysis of moore-penrose inversion*, SIAM J. Matr. Anal. and Appl. (2010), no. 31, 2769–2783.

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

[BS98]   Z. Bai and J. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab. **26** (1998), 316–345.

[BS05]   ———, *Spectral Analysis of Large Dimensional Random Matrices*, Springer, 2005.

[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.

[Don05a] D. L. Donoho, *High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension*, Discrete Comput. Geom. (2005), 617652.

[Don05b] ———, *Neighborly polytopes and sparse solution of underdetermined linear equations*, Technical Report, Statistics Department, Stanford University, 2005.

[DT05a]   D. L. Donoho and J. Tanner, *Neighborliness of randomly-projected simplices in high dimensions*, Proceedings of the National Academy of Sciences **102** (2005), no. 27, 9452–9457.

[DT05b]   ———, *Sparse nonnegative solution of underdetermined linear equations by linear programming*, Proceedings of the National Academy of Sciences **102** (2005), no. 27, 9446–9451.

[DT09]   ———, *Counting faces of randomly projected polytopes when the projection radically lowers dimension*, Journal of American Mathematical Society **22** (2009), 1–53.

[DT11]   D. L. Donoho and J. Tanner, *Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing*, Phil. Trans. R. Soc. A (2011), 4273–4293.

[KWT09]   Y. Kabashima, T. Wadayama, and T. Tanaka, *A typical reconstruction limit for compressed sensing based on lp-norm minimization*, J.Stat. Mech. (2009), L09003.

[Lub07]   D.S. Lubinsky, *A survey of weighted polynomial approximation with exponential weights*, Approcimation Theory **3** (1007), 1–105.

[MAYB11]   A. Maleki, L. Anitori, A. Yang, and R. Baraniuk, *Asymptotic Analysis of Complex LASSO via Complex Approximate Message Passing (CAMP)*, arXiv:1108.0477, 2011.

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

[RFG09]   S. Rangan, A. K. Fletcher, and V. K. Goyal, *Asymptotic analysis of map estimation via the replica method and applications to compressed sensing*, Neural Information Processing Systems (NIPS) (Vancouver), 2009.

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

[TV12]   T. Tao and V. Vu, *Random matrices: The Universality phenomenon for Wigner ensembles*, arXiv:1202.0068, 2012.

[VS92]   A. M. Vershik and P. V. Sporyshev, *Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem*, Selecta Math. Soviet. **11** (1992), 181201.