Approximate message passing with spectral initialization for generalized linear models*

OPEN ACCESS

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: ML 2022

Approximate message passing with spectral initialization for generalized linear models*

Marco Mondelli[1,**] **and Ramji Venkataramanan**[2]

[1] Institute of Science and Technology (IST), Austria
[2] Department of Engineering, University of Cambridge, United Kingdom
E-mail: marco.mondelli@ist.ac.at and ramji.v@eng.cam.ac.uk

Received 4 June 2022
Accepted for publication 5 June 2022
Published 24 November 2022

Online at stacks.iop.org/JSTAT/2022/114003
https://doi.org/10.1088/1742-5468/ac9828

Abstract. We consider the problem of estimating a signal from measurements obtained via a generalized linear model. We focus on estimators based on approximate message passing (AMP), a family of iterative algorithms with many appealing features: the performance of AMP in the high-dimensional limit can be succinctly characterized under suitable model assumptions; AMP can also be tailored to the empirical distribution of the signal entries, and for a wide class of estimation problems, AMP is conjectured to be optimal among all polynomial-time algorithms. However, a major issue of AMP is that in many models (such as phase retrieval), it requires an initialization correlated with the ground-truth signal and independent from the measurement matrix. Assuming that such an initialization is available is typically not realistic. In this paper, we solve this problem by proposing an AMP algorithm initialized with a spectral estimator. With such an initialization, the standard AMP analysis fails since the spectral estimator depends in a complicated way on the design matrix. Our main contribution is a rigorous characterization of the performance of AMP with spectral

*This article is an updated version of: Mondelli M and Venkataramanan R 2021 Approximate message passing with spectral initialization for generalized linear models *Proc. 24th Int. Conf. Artificial Intelligence and Statistics* vol 130, ed A Banerjee and K Fukumizu pp 397–405.
**Author to whom any correspondence should be addressed.

1742-5468/22/114003+41$33.00

initialization in the high-dimensional limit. The key technical idea is to define and analyze a two-phase artificial AMP algorithm that first produces the spectral estimator, and then closely approximates the iterates of the true AMP. We also provide numerical results that demonstrate the validity of the proposed approach.

Keywords: machine learning, message-passing algorithms, statistical inference, learning theory

Contents

1. Introduction

We consider the problem of estimating a d-dimensional signal $\boldsymbol{x} \in \mathbb{R}^d$ from n i.i.d. measurements $y_i \sim p(y|\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle)$, $i \in \{1, \dots, n\}$, where $\langle \cdot, \cdot \rangle$ is the scalar product, $\{\boldsymbol{a}_i\}_{1 \leqslant i \leqslant n}$ are given sensing vectors, and the (stochastic) output function $p(\cdot|\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle)$ is a given probability distribution. This is known as a *generalized linear model (GLM)* [McC18], and it encompasses many settings of interest in statistical estimation and signal processing [RG01, BB08, YLSV12, EK12]. One notable example is the problem of phase retrieval [Fie82, SEC+15], where $y_i = |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle|^2 + w_i$, with w_i being noise. Phase retrieval

J. Stat. Mech. (2022) 114003

appears in several areas of science and engineering, see e.g. [FD87, Mil90, DJ17], and the last few years have witnessed a surge of interest in the design and analysis of efficient algorithms; see the review [FS20] and the discussion at the end of this section.

Here, we consider GLMs in the high-dimensional setting where $n, d \to \infty$, with their ratio tending to a fixed constant, i.e. $n/d \to \delta \in \mathbb{R}$. We focus on a family of iterative algorithms known as approximate message passing (AMP). AMP algorithms are derived via approximations of belief propagation on the factor graph representing the estimation problem. AMP algorithms were first proposed for estimation in linear models [DMM09, BM11], and for estimation in GLMs [Ran11]. AMP has since been applied to a wide range of high-dimensional statistical estimation problems including compressed sensing [KMS+12, BM12, MAYB13], low rank matrix estimation [RF12, DM14, KKM+16], group synchronization [PWBM18], and specific instances of GLMs such as logistic regression [SC19] and phase retrieval [SR14, MXM19, MLKZ20].

Starting from an initialization $\boldsymbol{x}^0 \in \mathbb{R}^d$, the AMP algorithm for GLMs produces iteratively refined estimates of the signal, denoted by $\boldsymbol{x}^t$, for $t \geqslant 1$. An appealing feature of AMP is that, under suitable model assumptions, its performance in the high-dimensional limit can be precisely characterized by a succinct deterministic recursion called *state evolution* [BM11, Bol14, JM13]. Specifically, in the high-dimensional limit, the empirical distribution of the estimate $\boldsymbol{x}^t$ converges to the law of the random variable $\mu_t X + \sigma_t W_t$, for $t \geqslant 1$. Here $X \sim P_X$ (the signal prior), and $W_t \sim \mathsf{N}(0,1)$ is independent of X. The state evolution recursion specifies how the constants (μ_t, σ_t) can be computed from $(\mu_{t-1}, \sigma_{t-1})$ (see section 3 for details).

Using the state evolution analysis, it has been shown that AMP provably achieves Bayes-optimal performance in some special cases [DJM13, DM14, MV21]. Indeed, a conjecture from statistical physics posits that AMP is optimal among all polynomial-time algorithms. The optimality of AMP for GLMs is discussed in [BKM+19].

However, when used for estimation in GLMs, a major issue in current AMP theory is that in many problems (including phase retrieval) we require an initialization $\boldsymbol{x}^0$ that is correlated with the unknown signal $\boldsymbol{x}$ but independent of the sensing vectors $\{\boldsymbol{a}_i\}$. It is often not realistic to assume that such a realization is available. For such GLMs, without a correlated initialization, asymptotic state evolution analysis predicts that the AMP estimates will be uninformative, i.e. their normalized correlation with the signal vanishes in the large system limit. Thus, developing an AMP theory that does not rely on unrealistic assumptions about the initialization is an important open problem.

In this paper, we solve this open problem for a wide class of GLMs by rigorously analyzing the AMP algorithm with a *spectral estimator*. The idea of using a spectral estimator for GLMs was introduced in [Li92], and its performance in the high-dimensional limit was recently characterized in [LL19, MM19]. It was shown that the normalized correlation of the spectral estimator with the signal undergoes a phase transition, and for the special case of phase retrieval, the threshold for strictly positive correlation with the signal matches the information-theoretic threshold [MM19].

Our main technical contribution is a novel analysis of AMP with spectral initialization for GLMs, under the assumption that the sensing vectors $\{\boldsymbol{a}_i\}$ are i.i.d. Gaussian. This yields a rigorous characterization of the performance in the high-dimensional limit

J. Stat. Mech. (2022) 114003

(theorem 1). The analysis of AMP with spectral initialization is far from obvious since the spectral estimator depends in a non-trivial way on the sensing vectors $\{\boldsymbol{a}_i\}$. The existing state evolution analysis for GLMs [Ran11, JM13] crucially depends on the AMP initialization being independent of the sensing vectors, and therefore cannot be directly applied.

At the center of our approach is the design and analysis of an *artificial AMP* algorithm. The artificial AMP operates in two phases: in the first phase, it performs a power method, so that its iterates approach the spectral initialization of the true AMP; in the second phase, its iterates are designed to remain close to the iterates of the true AMP. The initialization of the artificial AMP is correlated with $\boldsymbol{x}$, but independent of the sensing vectors $\{\boldsymbol{a}_i\}$, which allows us to apply the standard state evolution analysis. Note that the initialization of the artificial AMP is impractical (it requires the knowledge of the unknown signal $\boldsymbol{x}$!). However, this is not an issue, since the artificial AMP is employed solely as a proof technique: we prove a state evolution result for the true AMP by showing that its iterates are close to those in the second phase of the artificial AMP.

Initializing AMP with a (different) spectral method has been recently shown to be effective for low-rank matrix estimation [MV21]. However, our proof technique for analyzing spectral initialization for GLMs is different from the approach in [MV21]. The argument in that paper is specific to the spiked random matrix model and relies on a delicate decoupling argument between the outlier eigenvectors and the bulk. Here, we follow an approach developed in [MTV20], where a specially designed AMP is used to establish the joint empirical distribution of the signal, the spectral estimator, and the linear estimator.

For the case of phase retrieval, in [MXM18] it is provided a heuristic argument for the validity of spectral initialization, and it is stated that establishing a rigorous proof is an open problem. Our paper not only solves this open problem, but it also gives a provable initialization method valid for a class of GLMs.

We note that, for some GLMs, AMP does not require a special initialization that is correlated with the signal $\boldsymbol{x}$. In section 3, we give a condition on the GLM output function that specifies precisely when such a correlated initialization is required (see (3.13)). This condition is satisfied by several popular GLMs, including phase retrieval. It is in these cases that AMP with spectral initialization is most useful.

Other related work. For the problem of phase retrieval, several algorithmic solutions have been proposed and analyzed in recent years. An inevitably non-exhaustive list includes semi-definite programming relaxations [CSV13, CESV15, CLS15a, WdM15], a convex relaxation operating in the natural domain of the signal [GS18, BR17], alternating minimization [NJS13], Wirtinger flow [CLS15b, CC17, MWCC20], iterative projections [LGL15], the Kaczmarz method [Wei15, TV19], and mirror descent [WR20]. A generalized AMP (GAMP) algorithm was introduced in [SR14], and an AMP to solve the non-convex problem with ℓ_2 regularization was proposed and analyzed in [MXM19]. Most of the algorithms mentioned above require an initialization correlated with the signal $\boldsymbol{x}$ and, to obtain such an initialization, spectral methods are widely employed.

J. Stat. Mech. (2022) 114003

Beyond the Gaussian setting, spectral methods for phase retrieval with random orthogonal matrices are analyzed in [DBMM20]. Statistical and computational phase transitions in phase retrieval for a large class of correlated real and complex random sensing matrices are investigated in [MLKZ20], and a general AMP algorithm for rotationally invariant matrices is studied in [Fan20]. In [ESAP+20], it is characterized the generalization error of GLMs via the multi-layer vector AMP of [FRS18, PSAR+20]. Thus, the extension of our techniques to more general sensing models represents an interesting avenue for future research.

2. Preliminaries

Notation and definitions. Given $n \in \mathbb{N}$, we use the shorthand $[n] = \{1, \dots, n\}$. Given a vector $\boldsymbol{x}$, we denote by $\|\boldsymbol{x}\|_2$ its Euclidean norm. The *empirical distribution* of a vector $\boldsymbol{x} = (x_1, \dots, x_d)^\mathsf{T}$ is given by $\frac{1}{d}\sum_{i=1}^d \delta_{x_i}$, where δ_{x_i} denotes a Dirac delta mass on x_i. Similarly, the empirical joint distribution of vectors $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$ is $\frac{1}{d}\sum_{i=1}^d \delta_{(x_i, x_i')}$.

GLMs. Let $\boldsymbol{x} \in \mathbb{R}^d$ be the signal of interest, and assume that $\|\boldsymbol{x}\|_2^2 = d$. The signal is observed via inner products with n sensing vectors $(\boldsymbol{a}_i)_{i\in[n]}$, with each $\boldsymbol{a}_i \in \mathbb{R}^d$ having independent Gaussian entries with mean zero and variance $1/d$, i.e. $(\boldsymbol{a}_i) \sim_{\text{i.i.d.}} \mathsf{N}(0, \boldsymbol{I}_d/d)$. Given $g_i = \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$, the components of the observed vector $\boldsymbol{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ are independently generated according to a conditional distribution $p_{Y|G}$, i.e. $y_i \sim p_{Y|G}(y_i|g_i)$. We stack the sensing vectors as rows to define the $n \times d$ sensing matrix $\boldsymbol{A}$, i.e. $\boldsymbol{A} = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_n]^\mathsf{T}$. For the special case of phase retrieval, the model is $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 + \boldsymbol{w}$, where $\boldsymbol{w}$ is a noise vector with independent entries. We consider a sequence of problems of growing dimension d, and assume that, as $d \to \infty$, the sampling ratio $n/d \to \delta$, for some constant $\delta \in (0, \infty)$. We remark that, as $d \to \infty$, the empirical distribution of $\boldsymbol{g} = (g_1, \dots, g_n)$ converges in Wasserstein distance (W_2) to $G \sim \mathsf{N}(0, 1)$.

Spectral initialization. The spectral estimator $\hat{\boldsymbol{x}}^s$ is the principal eigenvector of the $d \times d$ matrix $\boldsymbol{D}_n$, defined as

$$\boldsymbol{D}_n = \boldsymbol{A}^\mathsf{T} \boldsymbol{Z}_s \boldsymbol{A}, \quad \text{with } \boldsymbol{Z}_s = \operatorname{diag}(\mathcal{T}_s(y_1), \dots, \mathcal{T}_s(y_n)), \tag{2.1}$$

where $\mathcal{T}_s : \mathbb{R} \to \mathbb{R}$ is a preprocessing function. We now review some results from [MM19, LL19] on the performance of the spectral estimator in the high-dimensional limit.

Let $G \sim \mathsf{N}(0, 1)$, $Y \sim p(\cdot|G)$, and $Z_s = \mathcal{T}_s(Y)$. We will make the following assumptions on Z_s.

(A1) $\mathbb{P}(Z_s = 0) < 1$.
(A2) Z_s has bounded support and τ is the supremum of this support:

$$\tau = \inf\{z : \mathbb{P}(Z_s \leqslant z) = 1\}. \tag{2.2}$$

(A3) As λ approaches τ from the right, we have

$$\lim_{\lambda \to \tau^+} \mathbb{E}\left\{ \frac{Z_s}{(\lambda - Z_s)^2} \right\} = \lim_{\lambda \to \tau^+} \mathbb{E}\left\{ \frac{Z_s \cdot G^2}{\lambda - Z_s} \right\} = \infty. \tag{2.3}$$

J. Stat. Mech. (2022) 114003

For $\lambda \in (\tau, \infty)$ and $\delta \in (0, \infty)$, define

$$\phi(\lambda) = \lambda \mathbb{E}\left\{ \frac{Z_s \cdot G^2}{\lambda - Z_s} \right\}, \qquad \psi_\delta(\lambda) = \frac{\lambda}{\delta} + \lambda \mathbb{E}\left\{ \frac{Z_s}{\lambda - Z_s} \right\}. \tag{2.4}$$

Note that $\phi(\lambda)$ is a monotone non-increasing function and that $\psi_\delta(\lambda)$ is a convex function. Let $\bar{\lambda}_\delta$ be the point at which ψ_δ attains its minimum, i.e. $\bar{\lambda}_\delta = \arg\min_{\lambda \geqslant \tau} \psi_\delta(\lambda)$. For $\lambda \in (\tau, \infty)$, also define

$$\zeta_\delta(\lambda) = \psi_\delta(\max(\lambda, \bar{\lambda}_\delta)). \tag{2.5}$$

We remark that ζ_δ is an increasing function and, from lemma 2 in [MM19], we have that the equation $\zeta_\delta(\lambda) = \phi(\lambda)$ admits a unique solution for $\lambda > \tau$.

The following result characterizes the performance of the spectral estimator $\hat{\boldsymbol{x}}^s$. Its proof follows directly from lemma 2 in [MM19].

Lemma 2.1. *Let $\boldsymbol{x}$ be such that $\|\boldsymbol{x}\|_2^2 = d$, $\{\boldsymbol{a}_i\}_{i\in[n]} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, and $\boldsymbol{y} = (y_1, \ldots, y_n)$ with $\{y_i\}_{i\in[n]} \sim_{\text{i.i.d.}} p_{Y|G}$. Let $n/d \to \delta$, $G \sim \mathsf{N}(0,1)$ and define $Z_s = \mathcal{T}_s(Y)$ for $Y \sim p_{Y|G}$. Assume that Z_s satisfies the assumptions **(A1)**, **(A2)** and **(A3)**. Let $\hat{\boldsymbol{x}}^s$ be the principal eigenvector of the matrix $\boldsymbol{D}_n$ defined in (2.1), and let λ_δ^* be the unique solution of $\zeta_\delta(\lambda) = \phi(\lambda)$ for $\lambda > \tau$. Then, as $n \to \infty$,*

$$\frac{|\langle \hat{\boldsymbol{x}}^s, \boldsymbol{x} \rangle|^2}{\|\hat{\boldsymbol{x}}^s\|_2^2 \, \|\boldsymbol{x}\|_2^2} \xrightarrow{\text{a.s.}} a^2 \triangleq \begin{cases} 0, & \text{if } \psi_\delta'(\lambda_\delta^*) \leqslant 0, \\ \dfrac{\psi_\delta'(\lambda_\delta^*)}{\psi_\delta'(\lambda_\delta^*) - \phi'(\lambda_\delta^*)}, & \text{if } \psi_\delta'(\lambda_\delta^*) > 0, \end{cases} \tag{2.6}$$

where ψ_δ' and ϕ' are the derivatives of the respective functions.

Remark 2.1 (equivalent characterization). Using the definitions (2.4) and (2.5), the conditions $\zeta_\delta(\lambda_\delta^*) = \phi(\lambda_\delta^*)$ and $\psi_\delta'(\lambda_\delta^*) > 0$ are equivalent to

$$\mathbb{E}\left\{ \frac{Z_s(G^2 - 1)}{\lambda_\delta^* - Z_s} \right\} = \frac{1}{\delta}, \quad \text{and} \quad \mathbb{E}\left\{ \frac{Z_s^2}{(\lambda_\delta^* - Z_s)^2} \right\} < \frac{1}{\delta}. \tag{2.7}$$

When these conditions are satisfied, the limit of the normalized correlation in (2.6) can be expressed as

$$a^2 = \frac{\frac{1}{\delta} - \mathbb{E}\left\{ \frac{Z_s^2}{(\lambda_\delta^* - Z_s)^2} \right\}}{\frac{1}{\delta} + \mathbb{E}\left\{ \frac{Z_s^2 (G^2 - 1)}{(\lambda_\delta^* - Z_s)^2} \right\}}. \tag{2.8}$$

Remark 2.2 (optimal preprocessing function). In [MM19] it is derived the preprocessing function minimizing the value of δ necessary to achieve weak recovery, i.e. a strictly positive correlation between $\hat{\boldsymbol{x}}^s$ and $\boldsymbol{x}$. In particular, let δ_{u} be defined as

J. Stat. Mech. (2022) 114003

$$\delta_{\mathrm{u}} = \left(\int_{\mathbb{R}} \frac{\left(\mathbb{E}_G\{p(y|G)(G^2-1)\}\right)^2}{\mathbb{E}_G\{p(y|G)\}} \,\mathrm{d}y \right)^{-1}, \tag{2.9}$$

with $G \sim \mathsf{N}(0,1)$. Furthermore, let us also define

$$\bar{\mathcal{T}}(y) = \frac{\sqrt{\delta_{\mathrm{u}}} \cdot \mathcal{T}^*(y)}{\sqrt{\delta} - (\sqrt{\delta} - \sqrt{\delta_{\mathrm{u}}})\mathcal{T}^*(y)}, \tag{2.10}$$

where

$$\mathcal{T}^*(y) = 1 - \frac{\mathbb{E}_G\{p(y|G)\}}{\mathbb{E}_G\{p(y|G)\cdot G^2\}}. \tag{2.11}$$

Then, by taking $\mathcal{T}_s = \bar{\mathcal{T}}$, for any $\delta > \delta_{\mathrm{u}}$, we almost surely have

$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}^{\mathrm{s}}, \boldsymbol{x}\rangle|}{\|\hat{\boldsymbol{x}}^{\mathrm{s}}\|_2 \, \|\boldsymbol{x}\|_2} > \epsilon, \tag{2.12}$$

for some $\epsilon > 0$. Furthermore, for any $\delta < \delta_{\mathrm{u}}$, there is no pre-processing function $\mathcal{T}$ such that, almost surely, (2.12) holds. For a more formal statement of this result, see theorem 4 of [MM19]. The preprocessing function that, at a given $\delta > \delta_{\mathrm{u}}$, maximizes the correlation between $\hat{\boldsymbol{x}}^{\mathrm{s}}$ and $\boldsymbol{x}$ is also related to $\mathcal{T}^*(y)$ as defined in (2.11), and it is derived in [LAL19].

3. GAMP with spectral initialization

We make the following additional assumptions on the signal $\boldsymbol{x}$, the output distribution $p_{Y|G}$, and the preprocessing function $\mathcal{T}_s$ used for the spectral estimator.

(B1) Let $\hat{P}_{X,d}$ denote the *empirical distribution* of $\boldsymbol{x} \in \mathbb{R}^d$. As $d \to \infty$, $\hat{P}_{X,d}$ converges weakly to a distribution P_X such that $\lim_{d\to\infty} \mathbb{E}_{\hat{P}_{X,d}}\{|X|^2\} = \mathbb{E}_{P_X}\{|X|^2\}$. We note that $\mathbb{E}_{P_X}\{|X|^2\} = 1$, since we assume $\|\boldsymbol{x}\|_2^2 = d$.

(B2) We have $\mathbb{E}\{|Y|^2\} < \infty$, for $Y \sim p_{Y|G}(\cdot|G)$ and $G \sim \mathsf{N}(0,1)$. Furthermore, there exists a function $q : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ and a random variable V independent of G such that $Y = q(G, V)$. More precisely, for any measurable set $A \subseteq \mathcal{Y}$ and almost every g, we have $\mathbb{P}(Y \in A | G = g) = \mathbb{P}(q(g, V) \in A)$. We also assume that $\mathbb{E}\{|V|^2\} < \infty$. This is without loss of generality due to the functional representation lemma, see page 626 of [EGK11].

(B3) The function $\mathcal{T}_s : \mathbb{R} \to \mathbb{R}$ is bounded and Lipschitz.

Following the terminology of [Ran11], we refer to the AMP for GLMs as GAMP. In each iteration t, the proposed GAMP algorithm produces an estimate $\boldsymbol{x}^t$ of the signal $\boldsymbol{x}$. The algorithm is defined in terms of a sequence of Lipschitz functions $f_t : \mathbb{R} \to \mathbb{R}$ and $h_t : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, for $t \geqslant 0$. We initialize using the spectral estimator $\hat{\boldsymbol{x}}^{\mathrm{s}}$:

J. Stat. Mech. (2022) 114003

$$\boldsymbol{x}^0 = \sqrt{d}\,\frac{1}{\sqrt{\delta}}\hat{\boldsymbol{x}}^s, \tag{3.1}$$

$$\boldsymbol{u}^0 = \frac{1}{\sqrt{\delta}}\boldsymbol{A}f_0(\boldsymbol{x}^0) - \mathsf{b}_0\frac{\sqrt{\delta}}{\lambda_\delta^*}\boldsymbol{Z}_s\boldsymbol{A}\boldsymbol{x}^0, \tag{3.2}$$

where $\mathsf{b}_0 = \frac{1}{n}\sum_{i=1}^d f_0'(x_i^0)$, the diagonal matrix $\boldsymbol{Z}_s$ is defined in (2.1), and λ_δ^* is given by (2.7). Then, for $t \geqslant 0$, the algorithm computes:

$$\boldsymbol{x}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}^{\mathsf{T}}h_t(\boldsymbol{u}^t\,;\boldsymbol{y}) - \mathsf{c}_t f_t(\boldsymbol{x}^t), \tag{3.3}$$

$$\boldsymbol{u}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}f_{t+1}(\boldsymbol{x}^{t+1}) - \mathsf{b}_{t+1}h_t(\boldsymbol{u}^t\,;\boldsymbol{y}). \tag{3.4}$$

Here the functions f_t and h_t are understood to be applied component-wise, i.e. $f_t(\boldsymbol{x}^t) = (f_t(x_1^t), \ldots, f_t(x_d^t))$ and $h_t(\boldsymbol{u}^t\,;\boldsymbol{y}) = (h_t(u_1^t\,;y_1), \ldots, h_t(u_n^t\,;y_n))$. The scalars $\mathsf{b}_t, \mathsf{c}_t$ are defined as

$$\mathsf{c}_t = \frac{1}{n}\sum_{i=1}^n h_t'(u_i^t\,;y_i), \qquad \mathsf{b}_{t+1} = \frac{1}{n}\sum_{i=1}^d f_{t+1}'(x_i^{t+1}), \tag{3.5}$$

where $h_t'(\cdot,\cdot)$ denotes the derivative with respect to the first argument.

The asymptotic empirical distribution of the GAMP iterates $\boldsymbol{x}^t, \boldsymbol{u}^t$, for $t \geqslant 0$, can be succinctly characterized via a deterministic recursion, called *state evolution*. Our main result, theorem 1, shows that for $t \geqslant 0$, the empirical distributions of $\boldsymbol{u}^t$ and $\boldsymbol{x}^t$ converge in Wasserstein distance W_2 to the laws of the random variables U_t and X_t, respectively, with

$$X_t \equiv \mu_{X,t}X + \sigma_{X,t}W_{X,t}, \tag{3.6}$$

$$U_t \equiv \mu_{U,t}G + \sigma_{U,t}W_{U,t}, \tag{3.7}$$

where $(G, W_{U,t}) \sim_{\text{i.i.d.}} \mathsf{N}(0,1)$. Similarly, $X \sim P_X$ and $W_{X,t} \sim \mathsf{N}(0,1)$ are independent. The deterministic parameters $(\mu_{U,t}, \sigma_{U,t},\ \mu_{X,t}, \sigma_{X,t})$ are recursively computed as follows, for $t \geqslant 0$:

$$\begin{aligned}
\mu_{U,t} &= \frac{1}{\sqrt{\delta}}\mathbb{E}\{X f_t(X_t)\},\\
\sigma_{U,t}^2 &= \frac{1}{\delta}\mathbb{E}\big\{f_t(X_t)^2\big\} - \mu_{U,t}^2,\\
\mu_{X,t+1} &= \sqrt{\delta}\mathbb{E}\{G h_t(U_t\,;Y)\} - \mathbb{E}\{h_t'(U_t\,;Y)\}\mathbb{E}\{X f_t(X_t)\},\\
\sigma_{X,t+1}^2 &= \mathbb{E}\{h_t(U_t\,;Y)^2\}.
\end{aligned} \tag{3.8}$$

For the spectral initialization in (3.1) and (3.2), with a as defined in (2.6), the recursion is initialized with

$$\mu_{X,0} = a/\sqrt{\delta}, \qquad \sigma_{X,0}^2 = (1-a^2)/\delta. \tag{3.9}$$

J. Stat. Mech. (2022) 114003

We state the main result in terms of *pseudo-Lipschitz* test functions. A function $\psi : \mathbb{R}^m \to \mathbb{R}$ is pseudo-Lipschitz of order 2, i.e. $\psi \in \mathrm{PL}(2)$, if there is a constant $C > 0$ such that

$$\|\psi(\boldsymbol{x}) - \psi(\boldsymbol{y})\|_2 \leqslant C(1 + \|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2)\|\boldsymbol{x} - \boldsymbol{y}\|_2, \tag{3.10}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$. Examples of test functions in PL(2) with $m = 2$ include $\psi(a, b) = (a - b)^2$, $\psi(a, b) = ab$.

Theorem 1. *Let $\boldsymbol{x}$ be such that $\|\boldsymbol{x}\|_2^2 = d$, $\{\boldsymbol{a}_i\}_{i\in[n]} \sim_{\text{i.i.d.}} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, and $\boldsymbol{y} = (y_1, \ldots, y_n)$ with $\{y_i\}_{i\in[n]} \sim_{\text{i.i.d.}} p_{Y|G}$. Let $n/d \to \delta$, $G \sim \mathsf{N}(0, 1)$, and $Z_s = \mathcal{T}_s(Y)$ for $Y \sim p_{Y|G}(\cdot|G)$. Assume that **(A1)**, **(A2)**, **(A3)** and **(B1)**, **(B2)**, **(B3)** hold. Assume further that $\psi'_\delta(\lambda^*_\delta) > 0$, and let $\hat{\boldsymbol{x}}^s$ be the principal eigenvector of $\boldsymbol{D}_n$, defined as in (2.1), with the sign of $\hat{\boldsymbol{x}}^s$ chosen so that $\langle \hat{\boldsymbol{x}}^s, \boldsymbol{x}\rangle \geqslant 0$.*

Consider the GAMP iteration in equations (3.3) and (3.4) with initialization in equations (3.1) and (3.2). Assume that for $t \geqslant 0$, the functions f_t, h_t are Lipschitz with derivatives that are continuous almost everywhere. Then, the following limits hold almost surely for any PL(2) function $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and t such that $\sigma^2_{X,k}$ is strictly positive for $0 \leqslant k \leqslant t$:

$$\lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \psi(x_i, x_i^{t+1}) = \mathbb{E}\{\psi(X,\, \mu_{X,t+1}X + \sigma_{X,t+1}W_{X,t+1})\}, \tag{3.11}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(y_i, u_i^{t}) = \mathbb{E}\{\psi(Y,\, \mu_{U,t}G + \sigma_{U,t}W_{U,t})\}. \tag{3.12}$$

The result (3.11) also holds for $(t+1) = 0$. In (3.11) (resp. (3.12)), the expectation is over the independent random variables $X \sim P_X$ and $W_{X,t} \sim \mathsf{N}(0, 1)$ (resp. $(G, W_{U,t}) \sim_{\text{i.i.d.}} \mathsf{N}(0, 1)$). The scalars $(\mu_{X,t}, \mu_{U,t}, \sigma^2_{X,t}, \sigma^2_{U,t})_{t\geqslant 0}$ are given by the recursion (3.8) with the initialization (3.9).

We give a sketch of the proof in section 5 and defer the technical details to the appendices.

We now comment on some of the assumptions in the theorem. The assumption $\psi'_\delta(\lambda^*_\delta) > 0$ is required to ensure that the spectral initialization $\boldsymbol{x}^0$ has non-zero correlation with the signal $\boldsymbol{x}$ (lemma 2.1). From remark 2.2, we also know that for any sampling ratio $\delta > \delta_{\mathrm{u}}$ there exists a choice of $\mathcal{T}_s$ such that $\psi'_\delta(\lambda^*_\delta) > 0$. We also note that, for $\delta < \delta_{\mathrm{u}}$, GAMP converges to the 'un-informative fixed point' (where the estimate has vanishing correlation with signal) even if the initial condition has non-zero correlation with the signal, see theorem 5 of [MM19].

There is no loss of generality in assuming the sign of $\hat{\boldsymbol{x}}^s$ to be such that $\langle \hat{\boldsymbol{x}}^s, \boldsymbol{x}\rangle \geqslant 0$. Indeed, if the sign were chosen otherwise, the theorem would hold with the state evolution initialization in (3.9) being $\mu_{X,0} = -a/\sqrt{\delta}$, $\sigma^2_{X,0} = (1 - a^2)/\delta$.

The assumption that $\sigma^2_{X,k}$ is positive for $k \leqslant t$ is natural. Indeed, if $\sigma^2_{X,k} = 0$, then the state evolution result for iteration k implies that $\left\|\boldsymbol{x} - \mu^{-1}_{X,k}\boldsymbol{x}^k\right\|^2/d \to 0$ as $d \to \infty$. That is, we can perfectly estimate $\boldsymbol{x}$ from $\boldsymbol{x}^k$, and thus terminate the algorithm after iteration k.

J. Stat. Mech. (2022) 114003

Let us finally remark that the result in (3.11) is equivalent to the statement that the empirical joint distribution of $(\boldsymbol{x}, \boldsymbol{x}^{t+1})$ converges almost surely in Wasserstein distance (W_2) to the joint law of $(X,\ \mu_{X,t+1}X + \sigma_{X,t+1}W)$. This follows from the fact that a sequence of distributions P_n with finite second moment converges in W_2 to P if and only if P_n converges weakly to P and $\int \|a\|_2^2 \,\mathrm{d}P_n(a) \to \int \|a\|_2^2 \,\mathrm{d}P(a)$, see definition 6.7 and theorem 6.8 of [Vil08].

When does GAMP require spectral initialization? For the GAMP to give meaningful estimates, we need either $\boldsymbol{x}^0$ or $\boldsymbol{x}^1$ to have strictly non-zero asymptotic correlation with $\boldsymbol{x}$. To see when this can be arranged without a special initialization, consider the *linear* estimator $\hat{\boldsymbol{x}}^{\mathrm{L}}(\xi) := \boldsymbol{A}^{\mathsf{T}}\xi(\boldsymbol{y})$, for some function $\xi : \mathbb{R} \to \mathbb{R}$ that acts component-wise on $\boldsymbol{y}$. If there exists a function ξ such that the asymptotic normalized correlation between $\hat{\boldsymbol{x}}^{\mathrm{L}}(\xi)$ and $\boldsymbol{x}$ is strictly non-zero, then AMP does not require a special initialization (spectral or otherwise) that is correlated with $\boldsymbol{x}$. Indeed, in this case we can replace the initialization in (3.1) and (3.2) by $\boldsymbol{x}^0 = \boldsymbol{0}$, $\boldsymbol{u}^0 = \boldsymbol{0}$ (by taking $f_0 = 0$), and let $h_0(\boldsymbol{u}^0; \boldsymbol{y}) = \sqrt{\delta}\xi(\boldsymbol{y})$. This gives $\boldsymbol{x}^1 = \boldsymbol{A}^{\mathsf{T}}\xi(\boldsymbol{y}) = \hat{\boldsymbol{x}}^{\mathrm{L}}(\xi)$, which has strictly non-zero asymptotic correlation with $\boldsymbol{x}$. This ensures that $|\mu_{X,1}| > 0$, and the standard AMP analysis [JM13] directly yields a state evolution result similar to theorem 1.

The output function $p_{Y|G}$ determines whether a non-trivial linear estimator exists for the GLM. From appendix C.1 in [MTV20], we have that, if

$$\int_{\mathbb{R}} \frac{\left(\mathbb{E}_{G\sim\mathsf{N}(0,1)}\{G\, p_{Y|G}(y|G)\}\right)^2}{\mathbb{E}_{G\sim\mathsf{N}(0,1)}\{p_{Y|G}(y|G)\}}\,\mathrm{d}y = 0, \tag{3.13}$$

then the correlation between $\boldsymbol{A}^{\mathsf{T}}\xi(\boldsymbol{y})$ and $\boldsymbol{x}$ will asymptotically vanish for any choice of ξ. The condition (3.13) holds for many output functions of interest, including all distributions $p_{Y|G}$ that are even in G (and, therefore, including phase retrieval). It is for these models that spectral initialization is particularly useful.

We remark that our analysis covers not only the (Wirtinger flow) phase retrieval model $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$, but also the amplitude flow phase retrieval model $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|$. In fact, one can analyze the approximate model $\boldsymbol{y} = \sqrt{|\boldsymbol{A}\boldsymbol{x}|^2 + \boldsymbol{\epsilon}}$ and then let $\|\boldsymbol{\epsilon}\|_2 \to 0$. This is similar to the approach taken, e.g. by [MXM18] and [LWL20]. Since the functions used in each AMP iteration are Lipschitz, state evolution holds as $\|\boldsymbol{\epsilon}\|_2 \to 0$. For other GLMs with non-differentiable output functions, we can use a similar approach to construct a smooth approximation to the output function and obtain the state evolution result.

Bayes-optimal GAMP. Applying theorem 1 to the PL(2) function $\psi(x, y) = (x - f_t(y))^2$, we obtain the asymptotic mean-squared error (MSE) of the GAMP estimate $f_t(\boldsymbol{x}^t)$. In formulas, for $t \geqslant 0$, almost surely,

$$\lim_{d\to\infty} \frac{1}{d}\left\|\boldsymbol{x} - f_t(\boldsymbol{x}^t)\right\|_2^2 = \mathbb{E}\{(X - f_t(\mu_{X,t}X + \sigma_{X,t}W))^2\}. \tag{3.14}$$

If the limiting empirical distribution P_X of the signal is known, then the choice of f_t that minimizes the MSE in (3.14) is

$$f_t^*(s) = \mathbb{E}\{X|\mu_{X,t}X + \sigma_{X,t}W = s\}. \tag{3.15}$$

J. Stat. Mech. (2022) 114003

Similarly, applying the theorem to the PL(2) functions $\psi(x, y) = xf_t(y)$ and $\psi(x, y) = f_t(y)^2$, we obtain the asymptotic normalized correlation with the signal. In formulas, for $t \geqslant 0$, almost surely

$$\lim_{d\to\infty} \frac{|\langle \boldsymbol{x}, f_t(\boldsymbol{x}^t)\rangle|}{\|\boldsymbol{x}\|_2 \|f_t(\boldsymbol{x}^t)\|_2} = \frac{|\mathbb{E}\{X f_t(\mu_{X,t}X + \sigma_{X,t}W)\}|}{\sqrt{\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W)^2\}}}. \tag{3.16}$$

For fixed $(\mu_{X,t}, \sigma_{X,t}^2)$, the normalized correlation in (3.16) is maximized by taking $f_t = cf_t^*$ for any $c \neq 0$. This choice also maximizes the ratio $\mu_{U,t}^2/\sigma_{U,t}^2$ in (3.8). For $f_t = cf_t^*$, from (3.8) we have

$$\mu_{U,t} = \frac{c}{\sqrt{\delta}}\mathbb{E}\{f_t^*(X_t)^2\}, \qquad \sigma_{U,t}^2 = \frac{c}{\sqrt{\delta}}\mu_{U,t} - \mu_{U,t}^2. \tag{3.17}$$

We now specify the choice of $h_t(u\,;\, y)$ that maximizes the ratio $\mu_{X,t+1}^2/\sigma_{X,t+1}^2$ for fixed $(\mu_{U,t}, \sigma_{U,t}^2)$.

Proposition 3.1. *Assume the setting of theorem 1. For a given $(\mu_{U,t}, \sigma_{U,t}^2)$, the ratio $\mu_{X,t+1}^2/\sigma_{X,t+1}^2$ is maximized when $h_t(u\,;\, y) = c\, h_t^*(u\,;\, y)$ where $c \neq 0$ is any constant, and*

$$h_t^*(u\,;\, y) \triangleq \frac{\mathbb{E}\{G|U_t = u,\, Y = y\} - \mathbb{E}\{G|U_t = u\}}{\mathsf{Var}\,(G|U_t = u)} \tag{3.18}$$

$$= \frac{\mathbb{E}_W\{W p_{Y|G}(y\,|\,\rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\,W)\}}{\sqrt{1-\rho_t\,\mu_{U,t}}\;\mathbb{E}_W\{p_{Y|G}(y\,|\,\rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\,W)\}}, \tag{3.19}$$

where $\rho_t = \mu_{U,t}/(\mu_{U,t}^2 + \sigma_{U,t}^2)$ and $W \sim \mathsf{N}(0,1)$. In (3.18), the random variables U_t and Y are conditionally independent given G with

$$\begin{aligned} &Y \sim p_{Y|G}(\,\cdot\,|\,G), \qquad U_t = \mu_{U,t}G + \sigma_{U,t}W_{U,t}, \\ &(G, W_{U,t}) \sim_{\text{i.i.d.}} \mathsf{N}(0,1). \end{aligned} \tag{3.20}$$

The optimal choice for h_t^* in proposition 3.1 was derived by [Ran11] by approximating the belief propagation equations. For completeness, we provide a self-contained proof in appendix A. The proof also shows that with $h_t = ch_t^*$,

$$\mu_{X,t+1} = c\sqrt{\delta}\,\mathbb{E}\{|h_t^*(U_t\,;Y)|^2\}, \qquad \sigma_{X,t+1}^2 = c\,\frac{\mu_{X,t+1}}{\sqrt{\delta}}.$$

As the choices f_t^*, h_t^* maximize the signal-to-noise ratios $\mu_{U,t}^2/\sigma_{U,t}^2$ and $\mu_{X,t+1}^2/\sigma_{X,t+1}^2$, respectively, we refer to this algorithm as Bayes-optimal GAMP. We note that to apply theorem 1 to the Bayes-optimal GAMP, we need f_t^*, h_t^* to be Lipschitz. This holds under relatively mild conditions on P_X and $p_{Y|G}$, see lemma F.1 in [MV21].

4. Numerical simulations

We now illustrate the performance of the GAMP algorithm with spectral initialization via numerical examples. For concreteness, we focus on noiseless phase retrieval, where $y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2$, $i \in [n]$.

J. Stat. Mech. (2022) 114003

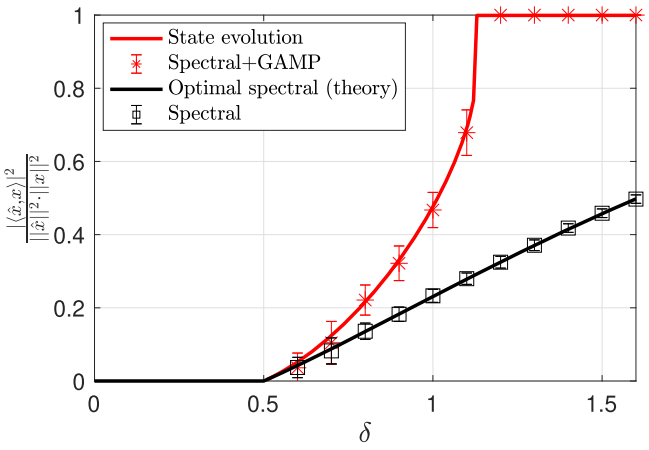


Figure 1. Performance comparison between GAMP with spectral initialization (in red) and the spectral method alone (in black) for a Gaussian prior $P_X \sim \mathsf{N}(0,1)$. The solid lines are the theoretical predictions of theorem 1 for GAMP with spectral initialization, and of lemma 2.1 for the spectral method. Error bars indicate one standard deviation around the empirical mean.

Gaussian prior. In figure 1, $\boldsymbol{x}$ is chosen uniformly at random on the d-dimensional sphere with radius $\sqrt{d}$ and $\{\boldsymbol{a}_i\}_{i\in[n]} \sim_{\text{i.i.d.}} \mathsf{N}(0, \boldsymbol{I}_d/d)$. Note that, as $d \to \infty$, the limiting empirical distribution P_X of $\boldsymbol{x}$ is a standard Gaussian. We take $d = 8000$, and the numerical simulations are averaged over $n_{\text{sample}} = 50$ independent trials. The performance of an estimate $\hat{\boldsymbol{x}}$ is measured via its normalized squared scalar product with the signal $\boldsymbol{x}$. The black points are obtained by estimating $\boldsymbol{x}$ via the spectral method, using the optimal pre-processing function $\mathcal{T}_s$ reported in equation (137) of [MM19]. The empirical results match the black curve, which gives the best possible squared correlation in the high-dimensional limit, as given by theorem 1 of [LAL19]. The red points are obtained by running the GAMP algorithm (3.3) and (3.4) with the spectral initialization (3.1) and (3.2). The function f_t is chosen to be the identity, and $h_t = \sqrt{\delta} h_t^*$, for h_t^* given by proposition 3.1. The algorithm is run until the normalized squared difference between successive iterates is small. As predicted by theorem 1, the numerical simulations agree well with the state evolution curve in red, which is obtained by computing the fixed point of the recursion (3.8) initialized with (3.9). We also remark that the threshold for exact recovery can be obtained from the fixed points of state evolution, see e.g. [BKM+19].

Bayes-optimal GAMP for a binary-valued prior. Assume now that each entry of the signal $\boldsymbol{x}$ takes value in $\{-1, 1\}$, with $P_X(1) = 1 - p_X(-1) = \mathsf{p}$. In figure 2, we take $\mathsf{p} = \frac{1}{2}$, and compare the performance of the GAMP algorithm with spectral initialization for two different choices of the function f_t: f_t equal to identity (in blue) and $f_t = f_t^*$ (in red), where f_t^* is the Bayes-optimal choice (3.15). By computing the conditional

J. Stat. Mech. (2022) 114003

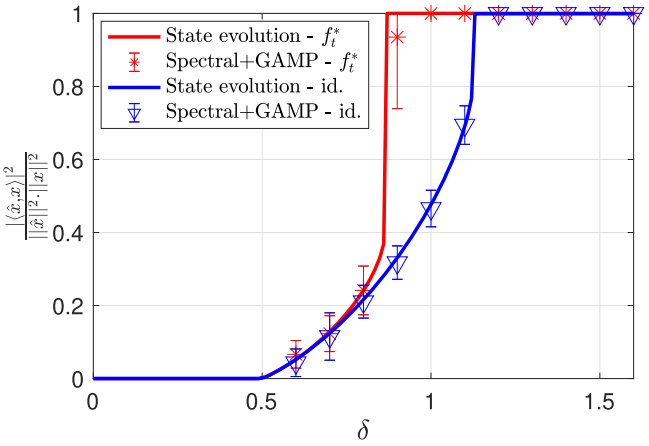


Figure 2. Performance comparison between two different choices of f_t for a binary prior $P_X(1) = P_X(-1) = \frac{1}{2}$. The Bayes-optimal choice $f_t = f_t^*$ (in red) has a lower threshold compared to f_t equal to identity (in blue).

expectation, we have

$$f_t^*(s) = 2\mathbb{P}(X = 1|\mu_{X,t}X + \sigma_{X,t}W = s) - 1 = \frac{2}{1 + \frac{1-\mathsf{p}}{\mathsf{p}} \exp\left(\frac{-2s\mu_{X,t}}{\sigma_{X,t}^2}\right)} - 1. \tag{4.1}$$

The rest of the setting is analogous to that of figure 1. There is a significant performance gap between the Bayes-optimal choice $f_t = f_t^*$ and the choice $f_t(x) = x$. As in the previous experiment, we observe very good agreement between the GAMP algorithm and the state evolution prediction of theorem 1. We remark that for this setting, the information-theoretically optimal overlap (computed using the formula in [BKM+19]) is 1 for all $\delta > 0$. Since the components of $\boldsymbol{x}$ are in $\{-1, 1\}$, there are 2^d choices for $\boldsymbol{x}$. The information-theoretically optimal estimator picks the choice that is consistent with $y_i = \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$, $i \in [n]$. (Since $\boldsymbol{A}$ is Gaussian, with high probability this solution is unique.)

Coded diffraction patterns. We consider the model of coded diffraction patterns described in section 7.2 of [MM19]. Here the signal $\boldsymbol{x}$ is the image of figure 3(a), and it can be viewed as a $d_1 \times d_2 \times 3$ array with $d_1 = 820$ and $d_2 = 1280$. The sensing vectors are given by

$$a_r(t_1, t_2) = d_\ell(t_1, t_2) \cdot \mathrm{e}^{\mathrm{i}2\pi k_1 t_1/d_1} \cdot \mathrm{e}^{\mathrm{i}2\pi k_2 t_2/d_2}, \tag{4.2}$$

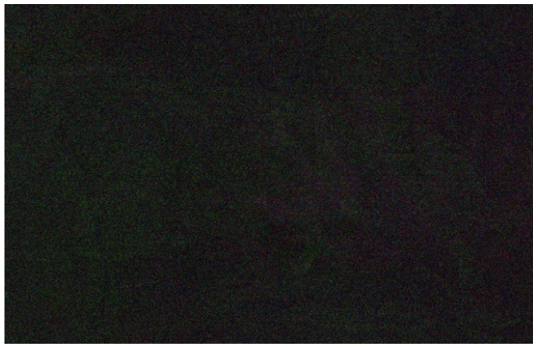
where $r \in [n]$, $t_1 \in [d_1]$, $t_2 \in [d_2]$, i denotes the imaginary unit, $a_r(t_1, t_2)$ is the (t_1, t_2)th component of $\boldsymbol{a}_r \in \mathbb{C}^d$, and the $(d_\ell(t_1, t_2))$'s are i.i.d. and uniform in $\{1, -1, i, -i\}$. The index $r \in [n]$ is associated to a pair (ℓ, k), with $\ell \in [L]$; the index $k \in [d]$ is associated to a pair (k_1, k_2) with $k_1 \in [d_1]$ and $k_2 \in [d_2]$. Thus, $n = L \cdot d$ and, therefore, $\delta = L \in \mathbb{N}$. To obtain non-integer values of δ, we set to 0 a suitable fraction of the vectors $\boldsymbol{a}_r$, chosen uniformly at random.

In this model, the scalar product $\langle \boldsymbol{x}_j, \boldsymbol{a}_r \rangle$ can be computed with an FFT algorithm. Furthermore, in order to evaluate the principal eigenvector for the spectral initialization, we use a power method which stops if either the number of iterations reaches the

J. Stat. Mech. (2022) 114003

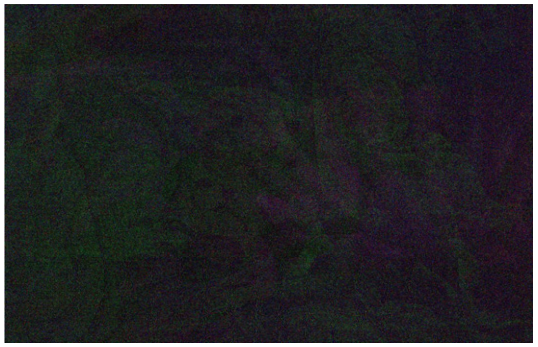
(a) Original image.

(b) Proposed, $\delta = 2.2$.



(c) Spectral, $\delta = 2.2$.

(d) Proposed, $\delta = 2.4$.



(e) Spectral, $\delta = 2.4$.

Figure 3. Visual comparison between the reconstruction of the GAMP algorithm with spectral initialization and that of the spectral method alone for measurements given by coded diffraction patterns.

maximum value of 100 000 or the modulus of the scalar product between the estimate at the current iteration T and at the iteration $T-10$ is larger than $1–10^{-7}$.

The GAMP algorithm with spectral initialization for the complex-valued setting is described in appendix D. Figure 3 shows a visual representation of the results. The improvement achieved by the GAMP algorithm over the spectral estimator is impressive, with GAMP achieving full recovery already at $\delta = 2.4$. A numerical comparison of the performance of the two methods is given in figure 5 in appendix D. We emphasize that the state evolution result of theorem 1 is only valid for Gaussian sensing matrices.

J. Stat. Mech. (2022) 114003

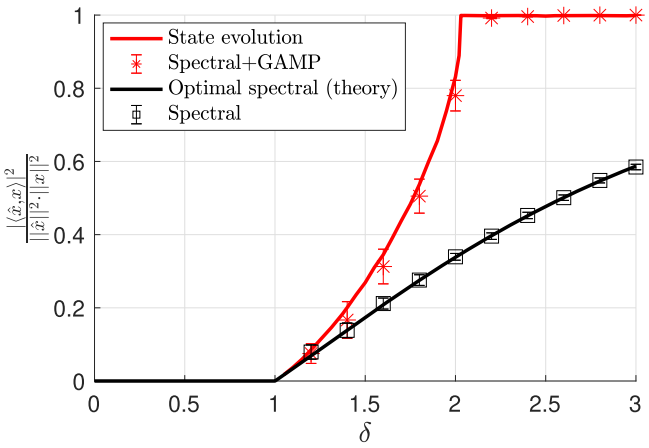


Figure 4. Performance comparison between complex GAMP with spectral initialization (in red) and the spectral method alone (in black) for a Gaussian prior $P_X \sim \mathsf{CN}(0,1)$. On the x-axis, we have the sampling ratio $\delta = n/d$; on the y-axis, we have the normalized squared scalar product between the signal and the estimate. The experimental results (* and $\square$ markers) are in excellent agreement with the theoretical predictions (solid lines) given by state evolution for GAMP and lemma 2.1 for the spectral method. Error bars indicate one standard deviation around the empirical mean.

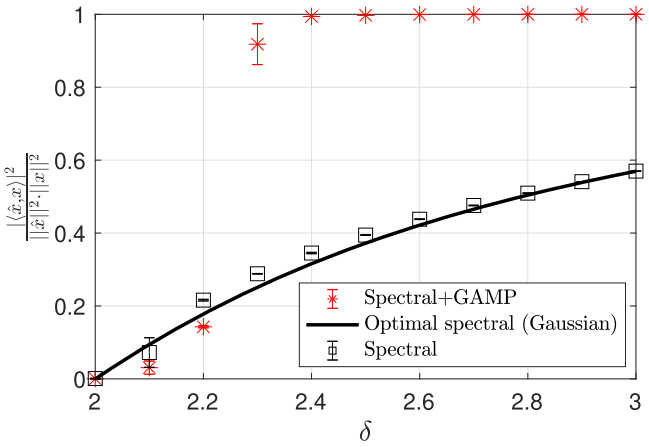


Figure 5. Performance comparison between complex GAMP with spectral initialization (in red) and the spectral method alone (in black) for a model of coded diffraction patterns.

Extending it to structured matrices such as coded diffraction patterns is an interesting direction for future work.

J. Stat. Mech. (2022) 114003

5. Sketch of the proof of theorem 1

We give an outline of the proof here, and provide the technical details in the appendices.

The artificial GAMP algorithm. We construct an artificial GAMP algorithm, whose iterates are denoted by $\tilde{\boldsymbol{x}}^t, \tilde{\boldsymbol{u}}^t$, for $t \geqslant 0$. Starting from an initialization $(\tilde{\boldsymbol{x}}^0, \tilde{\boldsymbol{u}}^0)$, for $t \geqslant 0$ we iteratively compute:

$$\tilde{\boldsymbol{x}}^{t+1} = \frac{1}{\sqrt{\delta}} \boldsymbol{A}^{\mathsf{T}} \tilde{h}_t(\tilde{\boldsymbol{u}}^t\,; \boldsymbol{y}) - \tilde{\mathsf{c}}_t \tilde{f}_t(\tilde{\boldsymbol{x}}^t), \tag{5.1}$$

$$\tilde{\boldsymbol{u}}^{t+1} = \frac{1}{\sqrt{\delta}} \boldsymbol{A} \tilde{f}_{t+1}(\tilde{\boldsymbol{x}}^{t+1}) - \tilde{\mathsf{b}}_{t+1} \tilde{h}_t(\tilde{\boldsymbol{u}}^t\,; \boldsymbol{y}). \tag{5.2}$$

For $t \geqslant 0$, the functions $\tilde{f}_t : \mathbb{R} \to \mathbb{R}$ and $\tilde{h}_t : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are Lipschitz, and will be specified below. The scalars $\tilde{\mathsf{c}}_t$ and $\tilde{\mathsf{b}}_{t+1}$ are defined as

$$\tilde{\mathsf{c}}_t = \frac{1}{n} \sum_{i=1}^{n} \tilde{h}_t'(\tilde{u}_i^t\,; y_i), \qquad \tilde{\mathsf{b}}_{t+1} = \frac{1}{n} \sum_{i=1}^{d} \tilde{f}_{t+1}'(\tilde{x}_i^{t+1}), \tag{5.3}$$

where $\tilde{h}_t'$ denotes the derivative with respect to the first argument. The iteration is initialized as follows. Choose any $\alpha \in (0, 1)$, and a standard Gaussian vector $\boldsymbol{n} \sim \mathsf{N}(0, \boldsymbol{I}_d)$ that is independent of $\boldsymbol{x}$ and $\boldsymbol{A}$. Then,

$$\tilde{\boldsymbol{x}}^0 = \alpha\, \boldsymbol{x} + \sqrt{1 - \alpha^2}\, \boldsymbol{n}, \qquad \tilde{\boldsymbol{u}}^0 = \frac{1}{\sqrt{\delta}} \boldsymbol{A} \tilde{f}_0(\tilde{\boldsymbol{x}}^0). \tag{5.4}$$

The artificial GAMP is divided into two phases. In the first phase, which lasts up to iteration T, the functions $\tilde{f}_t, \tilde{h}_t$ for $0 \leqslant t \leqslant (T-1)$, are chosen such that as $T \to \infty$, the iterate $\tilde{\boldsymbol{x}}^T$ approaches the initialization $\boldsymbol{x}^0$ of the true GAMP algorithm defined in (3.1). In the second phase, the functions $\tilde{f}_t, \tilde{h}_t$ for $t \geqslant T$, are chosen to match those of the true GAMP. The key observation is that a state evolution result for the artificial GAMP follows directly from the standard analysis of GAMP [JM13] since the initialization $\tilde{\boldsymbol{x}}^0$ is independent of $\boldsymbol{A}$. By showing that as $T \to \infty$, the iterates and the state evolution parameters of the artificial GAMP approach the corresponding quantities of the true GAMP, we prove that the state evolution result of theorem 1 holds.

We now specify the functions used in the artificial GAMP. For $0 \leqslant t \leqslant (T-1)$, we set

$$\tilde{f}_t(x) = \frac{x}{\beta_t}, \qquad \tilde{h}_t(x\,; y) = \sqrt{\delta}\, x\, \frac{\mathcal{T}_s(y)}{\lambda_\delta^* - \mathcal{T}_s(y)}, \tag{5.5}$$

where $\mathcal{T}_s$ is the pre-processing function used for the spectral estimator, λ_δ^* is the unique solution of $\zeta_\delta(\lambda) = \phi(\lambda)$ for $\lambda > \tau$ (also given by (2.7)), and $(\beta_t)_{t \geqslant 0}$ are constants coming from the state evolution recursion defined below. Furthermore, for $t \geqslant T$, we set

$$\tilde{f}_t(x) = f_{t-T}(x), \qquad \tilde{h}_t(x\,; y) = h_{t-T}(x\,; y). \tag{5.6}$$

J. Stat. Mech. (2022) 114003

With these choices of $\tilde{f}_t, \tilde{h}_t$, the coefficients $\tilde{\mathsf{c}}_t$ and $\tilde{\mathsf{b}}_t$ in (5.3) become:

$$\begin{aligned}\tilde{\mathsf{c}}_t &= \frac{\sqrt{\delta}}{n}\sum_{i=1}^{n}\frac{\mathcal{T}_s(y_i)}{\lambda_\delta^* - \mathcal{T}_s(y_i)}, \qquad \tilde{\mathsf{b}}_t = \frac{1}{\delta\beta_t}, \quad 0 \leqslant t \leqslant (T-1),\\ \tilde{\mathsf{c}}_t &= \frac{1}{n}\sum_{i=1}^{n} h'_{t-T}(\tilde{u}_i^t;\, y_i), \qquad \tilde{\mathsf{b}}_t = \frac{1}{n}\sum_{i=1}^{d} f'_{t-T}(\tilde{x}_i^t), \quad t \geqslant T.\end{aligned} \tag{5.7}$$

Since the initialization $\tilde{\boldsymbol{x}}^0$ in (5.4) is independent of $\boldsymbol{A}$, the state evolution result of [JM13] can be applied to the artificial GAMP. This result, formally stated in proposition B.1 in appendix B.1, implies that for $t \geqslant 0$, the empirical distributions of $\tilde{\boldsymbol{x}}^t$ and $\tilde{\boldsymbol{u}}^t$ converge in W_2 distance to the laws of the random variables $\tilde{X}_t$ and $\tilde{U}_t$, respectively, with

$$\tilde{X}_t \equiv \mu_{\tilde{X},t} X + \sigma_{\tilde{X},t} W_{\tilde{X},t}, \qquad \tilde{U}_t \equiv \mu_{\tilde{U},t} G + \sigma_{\tilde{U},t} W_{\tilde{U},t}. \tag{5.8}$$

Here $W_{\tilde{X},t}, W_{\tilde{U},t}$ are standard normal and independent of X and G, respectively. The state evolution recursion defining the parameters $(\mu_{\tilde{X},t}, \sigma_{\tilde{X},t}, \mu_{\tilde{U},t}, \sigma_{\tilde{U},t}, \beta_t)$ has the same form as (3.8), except that we use the functions defined in (5.5) for $0 \leqslant t \leqslant (T-1)$, and the functions in (5.6) for $t \geqslant T$. The detailed expressions are given in appendix B.1.

Analysis of the first phase. The first phase of the artificial GAMP is designed so that its output vectors after T iterations $(\tilde{\boldsymbol{x}}^T, \tilde{\boldsymbol{u}}^T)$ are close to the initialization $(\boldsymbol{x}^0, \boldsymbol{u}^0)$ of the true GAMP algorithm given by (3.1) and (3.2). This part of the algorithm is similar to the GAMP used in [MTV20] to approximate the spectral estimator $\hat{\boldsymbol{x}}^s$. In particular, the state evolution recursion of the first phase (given in (B.2)) converges as $T \to \infty$ to the following fixed point:

$$\lim_{T\to\infty} \mu_{\tilde{X},T} = \frac{a}{\sqrt{\delta}}, \qquad \lim_{T\to\infty} \sigma^2_{\tilde{X},T} = \frac{1-a^2}{\delta}, \tag{5.9}$$

where a is the limit (normalized) correlation between the spectral estimator $\hat{\boldsymbol{x}}^s$ and the signal, see (2.8). Furthermore, the GAMP iterate $\tilde{\boldsymbol{x}}^T$ approaches $\hat{\boldsymbol{x}}^s$, i.e.

$$\lim_{T\to\infty}\lim_{d\to\infty} \frac{\left\|\sqrt{d}\,\hat{\boldsymbol{x}}^s - \sqrt{\delta}\,\tilde{\boldsymbol{x}}^T\right\|^2}{d} = 0 \ \text{ a.s.} \tag{5.10}$$

These results are formally stated in lemmas B.2 and B.3, respectively, contained in appendix B.2.

Analysis of the second phase. The second phase of the artificial GAMP is designed so that its iterates $\tilde{\boldsymbol{x}}^{T+t}, \tilde{\boldsymbol{u}}^{T+t}$ are close to $\boldsymbol{x}^t, \boldsymbol{u}^t$, respectively for $t \geqslant 0$, and the corresponding state evolution parameters are also close. In particular, in order to prove theorem 1, we first analyze a slightly modified version of the true GAMP algorithm in (3.3) and (3.4) where the 'memory' coefficients b_t and c_t in (3.5) are replaced by deterministic values obtained from state evolution. The iterates of this modified GAMP, denoted by $\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{u}}^t$, are as follows. Start with the initialization

$$\hat{\boldsymbol{x}}^0 = \boldsymbol{x}^0 = \sqrt{d}\,\frac{1}{\sqrt{\delta}}\hat{\boldsymbol{x}}^s, \tag{5.11}$$

J. Stat. Mech. (2022) 114003

$$\hat{\boldsymbol{u}}^0 = \frac{1}{\sqrt{\delta}} \boldsymbol{A} f_0(\hat{\boldsymbol{x}}^0) - \bar{\mathsf{b}}_0 \frac{\sqrt{\delta}}{\lambda_\delta^*} \boldsymbol{Z}_s \boldsymbol{A} \hat{\boldsymbol{x}}^0, \tag{5.12}$$

where $\bar{\mathsf{b}}_0 = \frac{1}{\delta}\mathbb{E}\{f_0'(X_0)\}$. Then, for $t \geqslant 0$:

$$\hat{\boldsymbol{x}}^{t+1} = \frac{1}{\sqrt{\delta}} \boldsymbol{A}^{\mathsf{T}} h_t(\hat{\boldsymbol{u}}^t; \boldsymbol{y}) - \bar{\mathsf{c}}_t f_t(\hat{\boldsymbol{x}}^t), \tag{5.13}$$

$$\hat{\boldsymbol{u}}^{t+1} = \frac{1}{\sqrt{\delta}} \boldsymbol{A} f_{t+1}(\hat{\boldsymbol{x}}^{t+1}) - \bar{\mathsf{b}}_{t+1} h_t(\hat{\boldsymbol{u}}^t; \boldsymbol{y}). \tag{5.14}$$

Here, for $t \geqslant 0$, the deterministic memory coefficients $\bar{\mathsf{b}}_t$ and $\bar{\mathsf{c}}_t$ are

$$\bar{\mathsf{c}}_t = \mathbb{E}\{h_t'(U_t; Y)\}, \qquad \bar{\mathsf{b}}_t = \mathbb{E}\{f_t'(X_t)\}/\delta, \tag{5.15}$$

where X_t, U_t are defined in (3.6) and (3.7).

Let us now summarize our approach. We have defined three different GAMP iterations: (i) the *true GAMP* with iterates $(\boldsymbol{x}^t, \boldsymbol{u}^t)$ given by (3.3) and (3.4) and initialization $(\boldsymbol{x}^0, \boldsymbol{u}^0)$ given by (3.1) and (3.2), (ii) the *modified GAMP* with iterates $(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{u}}^t)$ given by (5.13) and (5.14) and initialization $(\hat{\boldsymbol{x}}^0, \hat{\boldsymbol{u}}^0)$ given by (5.11) and (5.12), and (iii) the *artificial GAMP* with iterates $(\tilde{\boldsymbol{x}}^t, \tilde{\boldsymbol{u}}^t)$ given by (5.1) and (5.2) and initialization $(\tilde{\boldsymbol{x}}^0, \tilde{\boldsymbol{u}}^0)$ given by (5.4). We recall that the *true GAMP* is the algorithm with spectral initialization that is actually implemented and whose performance we want to study. As the true GAMP is initialized with the spectral estimator $\hat{\boldsymbol{x}}^s$ which depends on $\boldsymbol{A}$, its performance cannot be characterized using the existing theory. To solve this problem, we introduce the *artificial GAMP* purely as a proof technique. In fact, the initialization of the artificial GAMP assumes knowledge of the signal, which makes it impractical. Finally, the *modified GAMP* is a slight modification of the true GAMP to simplify the proof.

Lemma B.5 in appendix B.3 proves that, for each $t \geqslant 0$, (i) the iterates $(\tilde{\boldsymbol{x}}^{t+T}, \tilde{\boldsymbol{u}}^{t+T})$ are close to $(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{u}}^t)$ for sufficiently large T, and (ii) the corresponding state evolution parameters are also close. We then use this lemma to prove theorem 1 by showing that the iterates of the *true GAMP* have the same asymptotic empirical distribution as those of the *modified GAMP*. In particular, we show in appendix B.4 that, almost surely,

$$\begin{aligned}\lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \psi(x_i, x_i^t) &= \lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \hat{x}_i^t) \\ &= \mathbb{E}\{\psi(X, \mu_{X,t} X + \sigma_{X,t} W)\}.\end{aligned} \tag{5.16}$$

6. Discussion

A major shortcoming in existing AMP theory for GLMs like phase retrieval is the unrealistic assumption that the initialization of the algorithm is correlated with the ground-truth signal and, at the same time, independent of the measurement matrix. This paper solves this problem by providing a rigorous analysis of AMP with a spectral initialization. Spectral initializations have been widely studied in recent years,

J. Stat. Mech. (2022) 114003

and have two attractive features. First, for phase retrieval, they meet the information theoretic threshold for weak recovery [MM19]. This means that, when the spectral initialization fails, no other method can work. Second, for a large class of GLMs, if the spectral method is unsuccessful, then AMP has an attractive fixed point at 0, see theorem 5 in [MM19]. This is a strong indication that, when the spectral initialization fails, the problem is computationally hard. An interesting future direction is to analyze the fixed points of AMP with spectral initialization, and compare with those of other algorithms that can be initialized with a spectral estimator, e.g. gradient descent.

Our analysis is based on an artificial AMP that first closely approximates the spectral estimator and then the true AMP algorithm. This technical tool is versatile and could be used beyond GLMs with Gaussian sensing matrices. Examples include more general measurement models [Fan20, ESAP+20], other message passing algorithms, e.g. vector AMP [SRF16], or the design of an artificial AMP that leads to a different estimator. We also highlight that the AMP analyzed here is rather general, and it includes as special cases both the Bayes-optimal AMP for GLMs and AMPs designed to optimize objective functions tailored to the signal prior.

Acknowledgments

The authors would like to thank Andrea Montanari for helpful discussions. M Mondelli was partially supported by the 2019 Lopez-Loreta Prize. R Venkataramanan was partially supported by the Alan Turing Institute under the EPSRC Grant EP/N510129/1.

Appendix A. Proof of proposition 3.1

From assumption **(B2)** on page 6, we recall that $h_t(u;\ y) = h_t(u;\ q(g,v))$. We write $\partial_g h_t(u;\ q(g,v))$ for the partial derivative with respect to g. We will show that $\mu_{X,t+1}$ in (3.8) can be written as:

$$\mu_{X,t+1} = \sqrt{\delta}\mathbb{E}\{\partial_g h_t(U_t\,;\, q(G,V))\}, \tag{A.1}$$

$$= \sqrt{\delta}\mathbb{E}\left\{h_t(U_t\,;Y)\left(\frac{\mathbb{E}\{G|U_t,Y\} - \mathbb{E}\{G|U_t\}}{\mathsf{Var}\ \{G|U_t\}}\right)\right\}. \tag{A.2}$$

From (A.2), we have that

$$\frac{\mu_{X,t+1}}{\sigma_{X,t+1}} = \frac{\sqrt{\delta}}{\sqrt{\mathbb{E}\{h_t(U_t\,;Y)^2\}}}\mathbb{E}\left\{h_t(U_t\,;Y)\left(\frac{\mathbb{E}\{G|U_t,Y\} - \mathbb{E}\{G|U_t\}}{\mathsf{Var}\ \{G|U_t\}}\right)\right\}. \tag{A.3}$$

The absolute value of the rhs is maximized when $h_t = c\,h_t^*$, for $c \neq 0$ and h_t^* is given in (3.18). To obtain the alternative expression in (3.19) from (3.18) we recall that U_t

J. Stat. Mech. (2022) 114003

is Gaussian with zero mean and variance $(\mu_{U,t}^2 + \sigma_{U,t}^2)$. Furthermore, the conditional distribution of G given $U_t = u$ is Gaussian with $\mathbb{E}\{G|U_t = u\} = \rho_t u$ and $\mathsf{Var}\,(G|U_t = u) = (1 - \rho_t \mu_{U,t})$. Therefore, with $W \sim \mathsf{N}(0,1)$ we have

$$\begin{aligned}\mathbb{E}\{G|U_t = u,\, Y = y\} &= \frac{\mathbb{E}_W\{(\rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\,W)\,p_{Y|G}(y\,|\,\rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\,W)\}}{\mathbb{E}_W\{p_{Y|G}(y\,|\,\rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\,W)\}} \\ &= \rho_t u \;+\; \sqrt{1-\rho_t\,\mu_{U,t}}\,\frac{\mathbb{E}\{W p_{Y|G}(y\,|\,\rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\,W)\}}{\mathbb{E}_W\{p_{Y|G}(y\,|\,\rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\,W)\}}. \qquad \text{(A.4)}\end{aligned}$$

Substituting (A.4) in (3.18) yields (3.19).

It remains to show (A.2), which we do by first showing (A.1). Define $e_t : \mathbb{R}^3 \to \mathbb{R}$ by

$$e_t(g, w, v) = h_t(\mu_{U,t} g + \sigma_{U,t} w\,;\, q(g, v)). \qquad \text{(A.5)}$$

Then, using the chain rule, the partial derivative of $e_t(g, w, v)$ with respect to g is

$$\partial_g e_t(g, w, v) = \mu_{U,t} h_t'(\mu_{U,t} g + \sigma_{U,t} w\,;\, q(g, v)) \;+\; \partial_g h_t(u\,;\, q(g, v)). \qquad \text{(A.6)}$$

The parameter $\mu_{X,t+1}$ in (3.8) can be written as

$$\begin{aligned}\mu_{X,t+1} &= \sqrt{\delta}[\mathbb{E}\{G\, e_t(G, W_{U,t}, V)\} - \mu_{U,t}\mathbb{E}\{h_t'(\mu_{U,t}G + \sigma_{U,t}W_{U,t}\,;\, Y)\}] \\ &\overset{\text{(i)}}{=} \sqrt{\delta}[\mathbb{E}\{\partial_g e_t(G, W_{U,t}, V)\} - \mu_{U,t}\mathbb{E}\{h_t'(\mu_{U,t}G + \sigma_{U,t}W_{U,t}\,;\, Y)\}] \\ &= \sqrt{\delta}\,\mathbb{E}\{\partial_g h_t(U_t\,;\, q(G, V))\}, \qquad \text{(A.7)}\end{aligned}$$

where the last equality is due to (A.6), and (i) holds due to Stein's lemma. Finally, we obtain (A.2) from (A.1) as follows:

$$\begin{aligned}\mathbb{E}\{\partial_g h_t(U_t\,;\, q(G, V))\} &= \mathbb{E}\big\{\mathbb{E}_{G|U_t}[\partial_g h_t(U_t\,;\, q(G, V))|U_t]\big\} \\ &\overset{\text{(ii)}}{=} \mathbb{E}\big\{\mathbb{E}_{G|U_t}[h_t(U_t\,;\, q(G, V)) \cdot (G - \mathbb{E}\{G|U_t\})/\mathsf{Var}\,\{G|U_t\}|U_t]\big\} \\ &= \mathbb{E}\big\{\mathbb{E}_{G|U_t,Y}[h_t(U_t\,;\, Y) \cdot (G - \mathbb{E}\{G|U_t\})/\mathsf{Var}\,\{G|U_t\}|U_t, Y]\big\} \\ &= \mathbb{E}\{h_t(U_t\,;Y) \cdot ((\mathbb{E}\{G|U_t, Y\} - \mathbb{E}\{G|U_t\})/\mathsf{Var}\,\{G|U_t\})\}. \qquad \text{(A.8)}\end{aligned}$$

Here step (ii) holds due to Stein's lemma. This completes the proof of the proposition. □

J. Stat. Mech. (2022) 114003

Appendix B. Proof of the main result

B.1. The artificial GAMP algorithm

The state evolution parameters for the artificial GAMP are recursively defined as follows. Recall from (5.8) that $\tilde{X}_t = \mu_{\tilde{X},t} X + \sigma_{\tilde{X},t} W_{\tilde{X},t}$ and $\tilde{U}_t \equiv \mu_{\tilde{U},t} G + \sigma_{\tilde{U},t} W_{\tilde{U},t}$. Using (5.4), the state evolution initialization is

$$\mu_{\tilde{X},0} = \alpha, \qquad \sigma^2_{\tilde{X},0} = 1 - \alpha^2, \qquad \beta_0 = \sqrt{\mu^2_{\tilde{X},0} + \sigma^2_{\tilde{X},0}} = 1. \tag{B.1}$$

For $0 \leqslant t \leqslant (T-1)$, the state evolution parameters are iteratively computed by using the functions defined in (5.5) in (3.8):

$$\begin{aligned}
\mu_{\tilde{U},t} &= \frac{\mu_{\tilde{X},t}}{\sqrt{\delta}\beta_t}, \qquad \sigma^2_{\tilde{U},t} = \frac{\sigma^2_{\tilde{X},t}}{\delta\,\beta_t^2}, \\
\mu_{\tilde{X},t+1} &= \frac{\mu_{\tilde{X},t}}{\sqrt{\delta}\beta_t}, \qquad \sigma^2_{\tilde{X},t+1} = \frac{1}{\beta_t^2}\,\mathbb{E}\left\{\frac{Z_s^2(G^2\mu^2_{\tilde{X},t} + \sigma^2_{\tilde{X},t})}{(\lambda^*_\delta - Z_s)^2}\right\}, \\
\beta_{t+1} &= \sqrt{\mu^2_{\tilde{X},t+1} + \sigma^2_{\tilde{X},t+1}}.
\end{aligned} \tag{B.2}$$

Here we recall that $G \sim \mathsf{N}(0,1)$, $Y \sim p_{Y|G}(\cdot|G)$, $Z_s = \mathcal{T}_s(Y)$, and the equality in (2.7) which is used to obtain the expression for $\mu_{\tilde{X},t+1}$. For $t \geqslant T$, the state evolution parameters are:

$$\begin{aligned}
\mu_{\tilde{U},t} &= \frac{1}{\sqrt{\delta}}\mathbb{E}\{X f_{t-T}(\tilde{X}_t)\}, \\
\sigma^2_{\tilde{U},t} &= \frac{1}{\delta}\mathbb{E}\left\{f_{t-T}(\tilde{X}_t)^2\right\} - \mu^2_{\tilde{U},t}, \\
\mu_{\tilde{X},t+1} &= \sqrt{\delta}\mathbb{E}\{G h_{t-T}(\tilde{U}_t; Y)\} - \mathbb{E}\{h'_{t-T}(\tilde{U}_t; Y)\}\mathbb{E}\{X f_{t-T}(\tilde{X}_t)\}, \\
\sigma^2_{\tilde{X},t+1} &= \mathbb{E}\{h_{t-T}(\tilde{U}_t; Y)^2\}.
\end{aligned} \tag{B.3}$$

Proposition B.1 (state evolution for artificial GAMP). *Consider the setting of theorem 1, the artificial GAMP iteration described in (5.1)–(5.7), and the corresponding state evolution parameters defined in (B.1)–(B.3).*

For any PL(2) function $\psi : \mathbb{R}^2 \to \mathbb{R}$, the following holds almost surely for $t \geqslant 1$:

$$\lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \tilde{x}_i^t) = \mathbb{E}\{\psi(X, \tilde{X}_t)\}, \tag{B.4}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(y_i, \tilde{u}_i^t) = \mathbb{E}\{\psi(Y, \tilde{U}_t)\}. \tag{B.5}$$

Here $X \sim P_X$ and $Y \sim P_{Y|G}$, with $G \sim \mathsf{N}(0,1)$. The random variables $\tilde{X}_t$, $\tilde{U}_t$ are defined in (5.8).

J. Stat. Mech. (2022) 114003

The proposition follows directly from the state evolution result of [JM13] since the initialization $\tilde{\boldsymbol{x}}^0$ of the artificial GAMP is independent of $\boldsymbol{A}$.

B.2. Analysis of the first phase

Lemma B.2 (fixed point of state evolution for first phase). *Consider the setting of theorem 1. Then, the state evolution recursion for the first phase, given by (B.1) and (B.2), converges as $T \to \infty$ to the following fixed point:*

$$\mu_{\tilde{X}} \triangleq \lim_{T\to\infty} \mu_{\tilde{X},T} = \frac{a}{\sqrt{\delta}}, \qquad \sigma^2_{\tilde{X}} \triangleq \lim_{T\to\infty} \sigma^2_{\tilde{X},T} = \frac{1-a^2}{\delta}, \tag{B.6}$$

where a is defined in (2.8).

Proof. Recall that λ^*_δ denotes the unique solution of $\zeta_\delta(\lambda) = \phi(\lambda)$ for $\lambda > \tau$ (also given by (2.7)), and define $Z = Z_s/(\lambda^*_\delta - Z_s)$, where $Z_s = \mathcal{T}_s(Y)$. Note that

$$\mathbb{E}\{Z(G^2-1)\} = \mathbb{E}\left\{\frac{Z_s(G^2-1)}{\lambda^*_\delta - Z_s}\right\} = \frac{1}{\delta}, \tag{B.7}$$

where the second equality follows from the equality in (2.7). Moreover, the inequality in (2.7) implies that

$$\frac{\mathbb{E}\{Z^2\}}{(\mathbb{E}\{Z(G^2-1)\})^2} = \delta^2 \mathbb{E}\left\{\frac{Z_s^2}{(\lambda^*_\delta - Z_s)^2}\right\} < \delta. \tag{B.8}$$

Thus, by recalling that the state evolution initialization $\mu_{\tilde{X},0} = \alpha$ is strictly positive, the result follows from lemma 5.2 in [MTV20]. □

Lemma B.3 (convergence to spectral estimator). *Consider the setting of theorem 1, and consider the first phase of the artificial GAMP iteration, given by (5.1) and (5.2) with $\tilde{f}_t$ and $\tilde{h}_t$ defined in (5.5). Then,*

$$\lim_{T\to\infty}\lim_{d\to\infty} \frac{\left\|\sqrt{d}\,\hat{\boldsymbol{x}}^{\mathrm{s}} - \sqrt{\delta}\,\tilde{\boldsymbol{x}}^T\right\|^2}{d} = 0 \ \text{ a.s..} \tag{B.9}$$

Furthermore, for any PL(2) function $\psi : \mathbb{R}\times\mathbb{R}\to\mathbb{R}$, almost surely we have:

$$\lim_{d\to\infty}\frac{1}{d}\sum_{i=1}^{d}\psi(x_i, \sqrt{d}\,\hat{x}^{\mathrm{s}}_i) = \lim_{T\to\infty}\lim_{d\to\infty}\frac{1}{d}\sum_{i=1}^{d}\psi(x_i, \sqrt{\delta}\,\tilde{x}^T_i) = \mathbb{E}\{\psi(X, \sqrt{\delta}\,(\mu_{\tilde{X}}X + \sigma_{\tilde{X}}W))\}. \tag{B.10}$$

Here $X \sim P_X$ and $W \sim \mathsf{N}(0,1)$ are independent.

Proof. As in the proof of the previous result, let $Z = Z_s/(\lambda^*_\delta - Z_s)$ and note that (B.7) and (B.8) hold. Also define

J. Stat. Mech. (2022) 114003

$$Z' \triangleq \frac{Z}{Z + \delta\mathbb{E}\{Z(G^2-1)\}} = \frac{Z}{Z+1} = \frac{Z}{\lambda_\delta^*}. \tag{B.11}$$

Then, the assumptions of lemma 5.4 in [MTV20] are satisfied, with the only difference of the initialization of the GAMP iteration (cf (5.4) in this paper and (5.4) in [MTV20]). However, it is straightforward to verify that the difference in the initialization does not affect the proof of lemma 5.4 in [MTV20]. Thus, (B.9) follows from (5.87) of [MTV20], and (B.10) follows by taking $k=2$ in (5.31) of [MTV20]. □

We will also need the following result on the convergence of the GAMP iterates.

Lemma B.4 (convergence of GAMP iterates). *Consider the first phase of the artificial GAMP iteration, given by (5.1) and (5.2) with $\tilde{f}_t$ and $\tilde{h}_t$ defined in (5.5). Then, the following limits hold almost surely:*

$$\lim_{T\to\infty}\lim_{n\to\infty}\frac{1}{n}\big\|\tilde{\boldsymbol{u}}^{T-1}-\tilde{\boldsymbol{u}}^{T-2}\big\|_2^2=0, \qquad \lim_{T\to\infty}\lim_{d\to\infty}\frac{1}{d}\big\|\tilde{\boldsymbol{x}}^{T}-\tilde{\boldsymbol{x}}^{T-1}\big\|_2^2=0. \tag{B.12}$$

Though the initialization of the GAMP in [MTV20] is different from (5.4), the proof of lemma B.4 is the same as that of lemma 5.3 in [MTV20] since it only relies on $\mu_{\tilde{X},0}=\alpha$ being strictly non-zero.

B.3. Analysis of the second phase

Lemma B.5. *Assume the setting of theorem 1. Consider the artificial GAMP algorithm (5.1) and (5.2) with the related state evolution recursion (B.2) and (B.3), and the modified version of the true GAMP algorithm (5.13) and (5.14). Fix any $\varepsilon>0$. Then, for $t\geqslant 0$ such that $\sigma^2_{X,k}>0$ for $0\leqslant k\leqslant t$, the following statements hold:*

(a)

$$\lim_{T\to\infty}\left|\mu_{\tilde{U},t+T}-\mu_{U,t}\right|=0, \qquad \lim_{T\to\infty}\left|\sigma^2_{\tilde{U},t+T}-\sigma^2_{U,t}\right|=0, \tag{B.13}$$

$$\lim_{T\to\infty}\left|\mu_{\tilde{X},T+t+1}-\mu_{X,t+1}\right|=0, \qquad \lim_{T\to\infty}\left|\sigma^2_{\tilde{X},T+t+1}-\sigma^2_{X,t+1}\right|=0. \tag{B.14}$$

(b) Let $\psi:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ be a PL(2) function. Then, almost surely,

$$\lim_{T\to\infty}\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\,\tilde{u}_i^{\,T+t})-\frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\,\hat{u}_i^{\,t})\right|=0, \tag{B.15}$$

$$\lim_{T\to\infty}\lim_{d\to\infty}\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\,\tilde{x}_i^{\,T+t+1})-\frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\,\hat{x}_i^{\,t+1})\right|=0. \tag{B.16}$$

The limits in (B.14) and (B.16) also hold for $t+1=0$.

Proof. We will use $\kappa_t,\kappa_t',c_t,\gamma_t$ to denote generic positive constants which depend on t, but not on n, d, or ε. The values of these constants may change throughout the proof.

Proof of (B.13) and (B.14). We prove the result by induction, starting from the base case $\left|\mu_{\tilde{X},T} - \mu_{X,0}\right|$, $|\sigma^2_{\tilde{X},T} - \sigma^2_{X,0}|$. From lemma B.2, we have

$$\lim_{T\to\infty} \mu_{\tilde{X},T} = \mu_{\tilde{X}} = \frac{a}{\sqrt{\delta}}, \qquad \lim_{T\to\infty} \sigma^2_{\tilde{X},T} = \sigma^2_{\tilde{X}} = \frac{1-a^2}{\delta}. \tag{B.17}$$

Recalling from (3.9) that $\mu_{X,0} = \frac{a}{\sqrt{\delta}}$, $\sigma^2_{X,0} = \frac{1-a^2}{\delta}$, (B.17) implies that

$$\lim_{T\to\infty} \left|\mu_{\tilde{X},T} - \mu_{X,0}\right| = 0, \qquad \lim_{T\to\infty} \left|\sigma^2_{\tilde{X},T} - \sigma^2_{X,0}\right| = 0. \tag{B.18}$$

Assume toward induction that (B.14) holds with $(t+1)$ replaced by t, and that $\sigma^2_{X,k} > 0$ for $0 \leqslant k \leqslant t$. We will show that (B.13) holds, and then that (B.14) holds.

For brevity, we write $\Delta_{\mu,t}, \Delta_{\sigma,t}$ for $(\mu_{X,t} - \mu_{\tilde{X},t+T})$ and $(\sigma_{X,t} - \sigma_{\tilde{X},t+T})$, respectively. By the induction hypothesis, given any $\varepsilon > 0$, for T sufficiently large we have

$$|\Delta_{\mu,t}| < \kappa_t \varepsilon, \qquad |\Delta_{\sigma,t}| < \frac{\kappa_t}{\sigma_{X,t} + \sigma_{\tilde{X},t+T}}\varepsilon = \kappa'_t \varepsilon. \tag{B.19}$$

Since $\sigma_{X,t}$ is strictly positive, κ'_t is finite and bounded above.

From (3.8) we have

$$\begin{aligned}\mu_{U,t} &= \frac{1}{\sqrt{\delta}}\mathbb{E}\{X f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})\} \\ &= \frac{1}{\sqrt{\delta}}\mathbb{E}\{X f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t} + \Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}).\end{aligned} \tag{B.20}$$

Recalling that f_t is Lipschitz and letting L_t denote its Lipschitz constant, we have

$$\begin{aligned}&\left|f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t} + \Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}) - f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})\right| \\ &\quad \leqslant L_t|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|.\end{aligned} \tag{B.21}$$

Using (B.21) in (B.20), we obtain

$$\begin{aligned}\sqrt{\delta}\mu_{U,t} \geqslant &\mathbb{E}\{X f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})\} - L_t\mathbb{E}\{|X||\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|\}, \\ \sqrt{\delta}\mu_{U,t} \leqslant &\mathbb{E}\{X f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})\} + L_t\mathbb{E}\{|X||\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|\}.\end{aligned} \tag{B.22}$$

Since $W_{X,t} \overset{\mathrm{d}}{=} W_{\tilde{X},t+T}$ and independent of X, we have that $\mathbb{E}\{X f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})\} = \sqrt{\delta}\mu_{\tilde{U},t+T}$. Therefore, (B.22) implies

$$\sqrt{\delta}\left|\mu_{U,t} - \mu_{\tilde{U},t+T}\right| \leqslant L_t(\Delta_{\mu,t} + \Delta_{\sigma,t}\mathbb{E}\{|W_{X,t}|\}), \tag{B.23}$$

where we have used $\mathbb{E}\{|X|^2\} < \sqrt{\mathbb{E}\{X^2\}} = 1$. Noting that $\mathbb{E}\{|W_{X,t}|\} = \sqrt{2/\pi}$, from (B.19) it follows that for sufficiently large T:

$$\left|\mu_{U,t} - \mu_{\tilde{U},t+T}\right| \leqslant \frac{L_t}{\sqrt{\delta}}(\kappa_t + \kappa'_t\sqrt{2/\pi})\,\varepsilon < \gamma_t\,\varepsilon. \tag{B.24}$$

J. Stat. Mech. (2022) 114003

Next consider $\sigma_{U,t}^2$. From (3.8), we have

$$\sigma_{U,t}^2 = \frac{1}{\delta}\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\} - \mu_{U,t}^2. \tag{B.25}$$

Furthermore, as $W_{X,t} \overset{\mathrm{d}}{=} W_{\tilde{X},t+T}$ and independent of X, we also have that

$$\sigma_{\tilde{U},t+T}^2 = \frac{1}{\delta}\mathbb{E}\{f_t(\mu_{\tilde{X},t+T}X + \sigma_{\tilde{X},t+T}W_{X,t})^2\} - \mu_{\tilde{U},t+T}^2. \tag{B.26}$$

Using the reverse triangle inequality, we have

$$\begin{aligned}
&\left|f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t} + \Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t})\right| \\
&\quad \geqslant |f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})| \\
&\qquad - \left|f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t} + \Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}) - f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})\right| \\
&\quad \geqslant |f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})| - L_t|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|,
\end{aligned} \tag{B.27}$$

where the last inequality follows from (B.21). Similarly,

$$\begin{aligned}
&\left|f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t} + \Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t})\right| \\
&\qquad \leqslant |f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})| + L_t|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|.
\end{aligned} \tag{B.28}$$

Using (B.27), we obtain the bound

$$\begin{aligned}
&\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\} \\
&\qquad \geqslant \mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})^2\} - L_t^2\mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\} \\
&\qquad\quad - 2L_t\sqrt{\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\} \cdot \mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\}}.
\end{aligned} \tag{B.29}$$

Similarly, using (B.28) we get

$$\begin{aligned}
&\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\} \\
&\qquad \leqslant \mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})^2\} + L_t^2\mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\} \\
&\qquad + 2L_t\sqrt{\mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})^2\} \cdot \mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\}}.
\end{aligned} \tag{B.30}$$

Furthermore,

$$\mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\} \leqslant 2|\Delta_{\mu,t}|^2\mathbb{E}\{X^2\} + 2|\Delta_{\sigma,t}|^2\mathbb{E}\{W_{X,t}\}^2 = 2(|\Delta_{\mu,t}|^2 + |\Delta_{\sigma,t}|^2).$$

From (3.8) and (B.3), we note that

$$\begin{aligned}
\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})\}^2 &= \delta(\mu_{U,t}^2 + \sigma_{U,t}^2), \\
\mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{\tilde{X},T+t})\}^2 &= \delta(\mu_{\tilde{U},T+t}^2 + \sigma_{\tilde{U},T+t}^2).
\end{aligned} \tag{B.31}$$

Therefore, equations (B.29) and (B.30) imply that

J. Stat. Mech. (2022) 114003

$$\begin{aligned}&\left|\mathbb{E}\{f_t(\mu_{X,t}X+\sigma_{X,t}W_{X,t})^2\}-\mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X+\sigma_{\tilde{X},T+t}W_{X,t})^2\}\right|\\&\quad\leqslant 2L_t^2(|\Delta_{\mu,t}|^2+|\Delta_{\sigma,t}|^2)+2L_t\sqrt{2\delta(\mu_{U,t}^2+\sigma_{U,t}^2+\mu_{\tilde{U},T+t}^2+\sigma_{\tilde{U},T+t}^2)(|\Delta_{\mu,t}|^2+|\Delta_{\sigma,t}|^2)}.\end{aligned}\tag{B.32}$$

Using this in (B.25) and (B.26), we have

$$\begin{aligned}\left|\sigma_{U,t}^2-\sigma_{\tilde{U},t+T}^2\right|\leqslant{}&|\mu_{\tilde{U},T+t}-\mu_{U,t}|\cdot|\mu_{\tilde{U},T+t}+\mu_{U,t}|+\left(\frac{2}{\delta}L_t^2(|\Delta_{\mu,t}|^2+|\Delta_{\sigma,t}|^2)\right.\\&\left.+\frac{2}{\sqrt{\delta}}L_t\sqrt{2(\mu_{U,t}^2+\sigma_{U,t}^2+\mu_{\tilde{U},T+t}^2+\sigma_{\tilde{U},T+t}^2)(|\Delta_{\mu,t}|^2+|\Delta_{\sigma,t}|^2)}\right).\end{aligned}\tag{B.33}$$

From (B.19), we obtain

$$|\Delta_{\mu,t}|^2+|\Delta_{\sigma,t}|^2<\left(\kappa_t^2+(\kappa_t')^2\right)\varepsilon^2.\tag{B.34}$$

Furthermore, as f_t is Lipschitz, from (B.31) and the induction hypothesis we have

$$|\mu_{\tilde{U},T+t}|+|\mu_{U,t}|+\sigma_{U,t}+\sigma_{\tilde{U},T+t}\leqslant c_t,\tag{B.35}$$

for some constant c_t. Using (B.24), (B.34) and (B.35) in (B.33), we conclude that for sufficiently large T:

$$\left|\sigma_{U,t}^2-\sigma_{\tilde{U},T+t}^2\right|<\gamma_t\varepsilon.\tag{B.36}$$

Next, we show that if (B.13) holds for some $t\geqslant 0$ and $\sigma_{X,k}^2>0$ for $k\leqslant t$, then:

$$\lim_{T\to\infty}\left|\mu_{\tilde{X},T+t+1}-\mu_{X,t+1}\right|=0,\qquad\lim_{T\to\infty}\left|\sigma_{\tilde{X},T+t+1}^2-\sigma_{X,t+1}^2\right|=0.\tag{B.37}$$

We denote the Lipschitz constant of h_t by $\bar{L}_t$, and write $\bar{\Delta}_{\mu,t}$, $\bar{\Delta}_{\sigma,t}$ for $(\mu_{U,t}-\mu_{\tilde{U},t+T})$ and $(\sigma_{U,t}-\sigma_{\tilde{U},t+T})$, respectively. Using this notation, we have

$$\begin{aligned}&\left|h_t(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t}+\bar{\Delta}_{\mu,t}G+\bar{\Delta}_{\sigma,t}W_{U,t};Y)-h_t(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t};Y)\right|\\&\quad\leqslant\bar{L}_t\left|\bar{\Delta}_{\mu,t}G+\bar{\Delta}_{\sigma,t}W_{U,t}\right|.\end{aligned}\tag{B.38}$$

The induction hypothesis (B.13) implies that for sufficiently large T:

$$\left|\bar{\Delta}_{\mu,t}\right|<\gamma_t\varepsilon,\qquad\left|\bar{\Delta}_{\sigma,t}\right|<\frac{\gamma_t}{\sigma_{U,t}+\sigma_{\tilde{U},t+T}}\varepsilon=\gamma_t\varepsilon.\tag{B.39}$$

We note that $\sigma_{U,t}>0$ since $\sigma_{X,t}>0$. Indeed, from the discussion leading to (3.17), for a fixed $\mu_{X,t},\sigma_{X,t}$ the smallest possible ratio $\sigma_{U,t}^2/\mu_{U,t}^2$ is achieved by the Bayes-optimal choice $f_t=cf_t^*$, where $f_t^*(X_t)=E\{X|X_t\}$. Furthermore, from (3.17), in order for $\sigma_{U,t}=0$, we need $\mathbb{E}\{\mathbb{E}\{X|X_t\}^2\}=1$. From Jensen's inequality, we also have $\mathbb{E}\{\mathbb{E}\{X|X_t\}^2\}\leqslant\mathbb{E}\{\mathbb{E}\{X^2|X_t\}\}=1$. Therefore, $\mathbb{E}\{\mathbb{E}\{X|X_t\}^2\}=1$ only if X is a deterministic function of $X_t=\mu_{X,t}X+\sigma_{X,t}W$. But this is impossible when $\sigma_{X,t}>0$. Therefore $\sigma_{U,t}>0$, and γ_t in (B.39) is strictly positive.

J. Stat. Mech. (2022) 114003

From (B.38), we obtain

$$\begin{aligned}\mathbb{E}\{Gh_t(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t}\,;\,Y)\}-\bar{L}_t\mathbb{E}\{\big|\bar{\Delta}_{\mu,t}\big|G^2+\big|\bar{\Delta}_{\sigma,t}\big|\cdot|G|\cdot|W_{U,t}|\}\\ \leqslant\mathbb{E}\{Gh_t(\mu_{U,t}G+\sigma_{U,t}W_{U,t}\,;\,Y)\}\\ \leqslant\mathbb{E}\{Gh_t(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t}\,;\,Y)\}+\bar{L}_t\mathbb{E}\{\big|\bar{\Delta}_{\mu,t}\big|G^2+\big|\bar{\Delta}_{\sigma,t}\big|\cdot|G|\cdot|W_{U,t}|\}.\end{aligned} \tag{B.40}$$

Now, using (3.8) and (B.3), we have:

$$\begin{aligned}\frac{1}{\sqrt{\delta}}\big|\mu_{\tilde{X},T+t+1}-\mu_{X,t+1}\big|&=\Big|\mathbb{E}\{G(h_t(\tilde{U}_{T+t}\,;\,Y)-h_t(U_t\,;\,Y))\}-\mu_{U,t}\Big(\mathbb{E}\{h_t'(\tilde{U}_{T+t}\,;Y)\}-\mathbb{E}\{h_t'(U_t\,;Y)\}\Big)\\ &\quad-\mathbb{E}\{h_t'(\tilde{U}_{T+t}\,;Y)\}(\mu_{\tilde{U},T+t}-\mu_{U,t})\Big|\\ &\leqslant\bar{L}_t\big(\big|\bar{\Delta}_{\mu,t}\big|+\big|\bar{\Delta}_{\sigma,t}\big|(2/\pi)\big)+|\mu_{U,t}|\cdot|\mathbb{E}\{h_t'(\tilde{U}_{T+t}\,;Y)\}-\mathbb{E}\{h_t'(U_t\,;Y)\}|+\bar{L}_t\big|\bar{\Delta}_{\mu,t}\big|.\end{aligned} \tag{B.41}$$

For the inequality above, we used (B.40) (noting that $\mathbb{E}\{|W_{U,t}|\}=\mathbb{E}\{|G|\}=\sqrt{2/\pi}$ and $\mathbb{E}\{G^2\}=1$), and the fact that $|h_t'|$ is bounded by $\bar{L}_t$, the Lipschitz constant of h_t. Now,

$$\begin{aligned}&\Big|\mathbb{E}\{h_t'(U_t\,;Y)\;-\mathbb{E}\{h_t'(\tilde{U}_{T+t}\,;Y)\}\Big|\\ &\qquad=\big|\mathbb{E}\{h_t'(\mu_{U,t}G+\sigma_{U,t}W_{U,t}\,;Y)\}-\mathbb{E}\{h_t'(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t}\,;Y)\}\big|.\end{aligned} \tag{B.42}$$

By the induction hypothesis (B.13), we have

$$\lim_{T\to\infty}\mu_{\tilde{U},T+t}=\mu_{U,t},\qquad\lim_{T\to\infty}\sigma_{\tilde{U},T+t}=\sigma_{U,t}. \tag{B.43}$$

Thus, as $T\to\infty$, the random variable $(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t})$ converges in distribution to $\mu_{U,t}G+\sigma_{U,t}W_{U,t}$. Then, lemma C.1 in appendix C implies that

$$\lim_{T\to\infty}\Big|\mathbb{E}\{h_t'(U_t\,;Y)\;-\mathbb{E}\{h_t'(\tilde{U}_{T+t}\,;Y)\}\Big|=0. \tag{B.44}$$

Using (B.44), (B.39) and (B.35) in (B.41) proves that the first limit in (B.37) holds.

Finally, we prove the second limit in (B.37). From (3.8), (B.3) and arguments along the same lines as (B.29)–(B.32), we obtain the bound

$$\begin{aligned}\Big|\sigma_{X,t+1}^2-\sigma_{\tilde{X},T+t+1}^2\Big|&=\Big|\mathbb{E}\{h_t(U_t\,;Y)^2\}\;-\;\mathbb{E}\{h_t(\tilde{U}_{t+T}\,;Y)^2\}\Big|\\ &\leqslant 2\bar{L}_t^2\big(\big|\bar{\Delta}_{\mu,t}\big|^2+\big|\bar{\Delta}_{\sigma,t}\big|^2\big)\\ &\quad+2\bar{L}_t\sqrt{(\sigma_{X,t+1}^2+\sigma_{\tilde{X},T+t+1}^2)\big(\big|\bar{\Delta}_{\mu,t}\big|^2+\big|\bar{\Delta}_{\sigma,t}\big|^2\big)}.\end{aligned} \tag{B.45}$$

Furthermore, as h_t is Lipschitz, the formulas for $\sigma_{X,t+1}^2$ and $\sigma_{\tilde{X},T+t+1}$ (in (3.8) and (B.3)) along with the induction hypothesis (B.43) imply that

$$\sigma_{X,t+1}^2+\sigma_{\tilde{X},T+t+1}^2\leqslant c_t, \tag{B.46}$$

J. Stat. Mech. (2022) 114003

for some constant c_t. By using (B.46) and (B.39), we can upper bound the rhs of (B.45) with $\kappa_{t+1}\varepsilon$, for sufficiently large T. This completes the proof of the second limit in (B.37).

Proof of (B.15) and (B.16). Since $\psi \in \mathrm{PL}(2)$, for $i \in [d]$ we have

$$\left|\psi(x_i, \tilde{x}_i^{T+t+1}) - \psi(x_i, \hat{x}_i^{t+1})\right| \leqslant C\big(1 + |x_i| + |\tilde{x}_i^{T+t+1}| + |\hat{x}_i^{t+1}|\big)\left|\tilde{x}_i^{T+t+1} - \hat{x}_i^{t+1}\right|, \tag{B.47}$$

for a universal constant $C > 0$. Therefore,

$$\begin{aligned}
&\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i, \tilde{x}_i^{T+t+1}) - \frac{1}{d}\sum_{i=1}^{d}\psi(x_i, \hat{x}_i^{t+1})\right| \\
&\quad\leqslant \frac{C}{d}\sum_{i=1}^{d}\big(1 + |x_i| + |\tilde{x}_i^{T+t+1}| + |\hat{x}_i^{t+1}|\big)\left|\tilde{x}_i^{T+t+1} - \hat{x}_i^{t+1}\right| \\
&\quad\leqslant 4C\left[1 + \frac{1}{d}\sum_{i=1}^{d}\Big(|x_i|^2 + |\tilde{x}_i^{T+t+1}|^2 + |\hat{x}_i^{t+1}|^2\Big)\right]^{1/2}\frac{\big\|\tilde{\boldsymbol{x}}^{T+t+1} - \hat{\boldsymbol{x}}^{t+1}\big\|_2}{\sqrt{d}},
\end{aligned} \tag{B.48}$$

where the second inequality follows from Cauchy–Schwarz. By the same argument,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_i, \tilde{u}_i^{T+t}) - \frac{1}{n}\sum_{i=1}^{n}\psi(y_i, \hat{u}_i^{t})\right| \\
&\quad\leqslant 4C\left[1 + \frac{1}{n}\sum_{i=1}^{n}\Big(|y_i|^2 + |\tilde{u}_i^{T+t}|^2 + |\hat{u}_i^{t}|^2\Big)\right]^{\frac{1}{2}}\frac{\big\|\tilde{\boldsymbol{u}}^{T+t} - \hat{\boldsymbol{u}}^{t}\big\|_2}{\sqrt{n}}.
\end{aligned} \tag{B.49}$$

We will show via induction that as $d \to \infty$: (i) the terms inside the square brackets in (B.48) and (B.49) converge almost surely to finite deterministic values, and (ii) as $T \to \infty$ (with the limit in T taken after the limit in d), the terms $\frac{\|\tilde{\boldsymbol{x}}^{T+t} - \hat{\boldsymbol{x}}^{t}\|_2}{\sqrt{d}}$ and $\frac{\|\tilde{\boldsymbol{u}}^{T+t+1} - \hat{\boldsymbol{u}}^{t+1}\|_2}{\sqrt{d}}$ converge to 0 almost surely.

Base case $t = 0$: the result (B.16) for $t + 1 = 0$ directly follows from lemma B.3. Next, using (B.49), the lhs of (B.15) for $t = 0$ can be bounded as

$$\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_i, \tilde{u}_i^{T}) - \frac{1}{n}\sum_{i=1}^{n}\psi(y_i, \hat{u}_i^{0})\right| \leqslant 4C\left[1 + \frac{\|\boldsymbol{y}\|_2^2}{n} + \frac{\|\tilde{\boldsymbol{u}}^{T}\|_2^2}{n} + \frac{\|\hat{\boldsymbol{u}}^{0}\|_2^2}{n}\right]^{\frac{1}{2}}\frac{\|\tilde{\boldsymbol{u}}^{T} - \hat{\boldsymbol{u}}^{0}\|_2}{\sqrt{n}}. \tag{B.50}$$

J. Stat. Mech. (2022) 114003

From the definition of the artificial GAMP (5.1)–(5.6), we have

$$\tilde{\boldsymbol{u}}^T = \frac{1}{\sqrt{\delta}}\boldsymbol{A} f_0(\tilde{\boldsymbol{x}}^T) - \sqrt{\delta}\,\tilde{\mathsf{b}}_T\,\boldsymbol{Z}\tilde{\boldsymbol{u}}^{T-1}, \tag{B.51}$$

where we define

$$\boldsymbol{Z} = \boldsymbol{Z}_s(\lambda_\delta^* \boldsymbol{I} - \boldsymbol{Z}_s)^{-1}, \tag{B.52}$$

with $\boldsymbol{Z}_s = \mathrm{diag}(\mathcal{T}_s(y_1), \ldots, \mathcal{T}_s(y_n))$. Similarly, defining

$$\boldsymbol{e}_1 := \tilde{\boldsymbol{u}}^{T-1} - \tilde{\boldsymbol{u}}^{T-2}, \tag{B.53}$$

we obtain $\tilde{\boldsymbol{u}}^{T-1} = \frac{1}{\sqrt{\delta}\beta_{T-1}}\big[\boldsymbol{A}\,\tilde{\boldsymbol{x}}^{T-1} - \boldsymbol{Z}\,\tilde{\boldsymbol{u}}^{T-1} + \boldsymbol{Z}\boldsymbol{e}_1\big]$, or

$$\tilde{\boldsymbol{u}}^{T-1} = \frac{1}{\sqrt{\delta}\beta_{T-1}}\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\big[\boldsymbol{A}\,\tilde{\boldsymbol{x}}^{T-1} + \boldsymbol{Z}\boldsymbol{e}_1\big]. \tag{B.54}$$

Substituting (B.54) in (B.51), we obtain

$$\begin{aligned}\tilde{\boldsymbol{u}}^T = \frac{1}{\sqrt{\delta}}\boldsymbol{A} f_0(\tilde{\boldsymbol{x}}^T) - \frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\boldsymbol{Z}\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\boldsymbol{A}\,\tilde{\boldsymbol{x}}^{T-1} \\ - \frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\boldsymbol{Z}^2\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\boldsymbol{e}_1.\end{aligned} \tag{B.55}$$

Using (B.55) and the expression for $\hat{\boldsymbol{u}}^0$ from (5.12), we have

$$\begin{aligned}\frac{1}{d}\big\|\tilde{\boldsymbol{u}}^T - \hat{\boldsymbol{u}}^0\big\|_2^2 &\leqslant 3\,\frac{\big\|\boldsymbol{A} f_0(\tilde{\boldsymbol{x}}^T) - \boldsymbol{A} f_0(\hat{\boldsymbol{x}}^0)\big\|_2^2}{\delta\, d} + 3\left\|\frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\boldsymbol{Z}^2\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\frac{\boldsymbol{e}_1}{\sqrt{d}}\right\|_2^2 \\ &\quad + \frac{3}{d}\left\|\frac{\bar{\mathsf{b}}_0\sqrt{\delta}}{\lambda_\delta^*}\boldsymbol{Z}_s\boldsymbol{A}\,\hat{\boldsymbol{x}}^0 - \frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\boldsymbol{Z}\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\boldsymbol{A}\,\tilde{\boldsymbol{x}}^{T-1}\right\|_2^2 \\ &:= 3(S_1 + S_2 + S_3).\end{aligned} \tag{B.56}$$

We now bound each of the three terms. By Cauchy–Schwarz inequality,

$$S_1 \leqslant \|\boldsymbol{A}\|_{\mathrm{op}}^2\,\frac{\big\|f_0(\tilde{\boldsymbol{x}}^T) - f_0(\hat{\boldsymbol{x}}^0)\big\|_2^2}{\delta\, d} \leqslant \|\boldsymbol{A}\|_{\mathrm{op}}^2\,\frac{L_0^2}{\delta}\cdot\frac{\big\|\tilde{\boldsymbol{x}}^T - \hat{\boldsymbol{x}}^0\big\|_2^2}{d}, \tag{B.57}$$

where L_0 is the Lipschitz constant of f_0. Since the entries of $\boldsymbol{A}$ are i.i.d. $\mathsf{N}(0, 1/d)$, almost surely the operator norm of $\boldsymbol{A}$ is bounded by a universal constant for sufficiently large d [AGZ09]. From lemma B.3 and the definition of $\hat{\boldsymbol{x}}^0$ in (5.11), we also have

$$\lim_{T\to\infty}\lim_{d\to\infty}\frac{\big\|\tilde{\boldsymbol{x}}^T - \hat{\boldsymbol{x}}^0\big\|_2^2}{d} = \frac{1}{\delta}\cdot\frac{\big\|\sqrt{\delta}\,\tilde{\boldsymbol{x}}^T - \sqrt{d}\,\hat{\boldsymbol{x}}^{\mathrm{s}}\big\|_2^2}{d} = 0 \quad \text{a.s.}. \tag{B.58}$$

Therefore,

$$\lim_{T\to\infty}\lim_{d\to\infty} S_1 = 0 \ \text{ a.s.}\,. \tag{B.59}$$

Next, recalling the definition of $\boldsymbol{e}_1$ from (B.53) we bound S_2 as follows:

$$S_2 \leqslant \frac{\tilde{\mathsf{b}}_T^{\,2}}{\beta_{T-1}^2}\left\|\boldsymbol{Z}^2\Big(\boldsymbol{I}+\boldsymbol{Z}/(\sqrt{\delta}\beta_{T-1})\Big)^{-1}\right\|_{\text{op}}^2 \cdot \frac{\left\|\tilde{\boldsymbol{u}}^{T-1}-\tilde{\boldsymbol{u}}^{T-2}\right\|_2^2}{d}. \tag{B.60}$$

From lemma B.4, we know that $\lim_{T\to\infty}\lim_{d\to\infty}\frac{\left\|\tilde{\boldsymbol{u}}^{T-1}-\tilde{\boldsymbol{u}}^{T-2}\right\|_2^2}{d}=0$ almost surely. We now show that the other terms on the rhs of (B.60) are bounded almost surely. Recall from (5.7) that $\tilde{\mathsf{b}}_T=\frac{1}{n}\sum_{i=1}^d f_0'(\tilde{x}_i^{\,T})$. Proposition B.1 guarantees that the empirical distribution of $\tilde{\boldsymbol{x}}^t$ converges to the law of $\tilde{X}_t\equiv\mu_{\tilde{X},t}X+\sigma_{\tilde{X},t}W$. Since f_0 is Lipschitz, lemma C.1 in appendix C therefore implies that almost surely:

$$\lim_{d\to\infty}\tilde{\mathsf{b}}_T=\frac{1}{\delta}\mathbb{E}\{f_0'(\mu_{\tilde{X},T}X+\sigma_{\tilde{X},T}W)\}. \tag{B.61}$$

From lemma B.2, we know that $\lim_{T\to\infty}\mu_{\tilde{X},T}=\frac{a}{\sqrt{\delta}}$ and $\lim_{T\to\infty}\sigma^2_{\tilde{X},T}=\frac{1-a^2}{\delta}$. Therefore, letting $T\to\infty$ and applying lemma C.1 again, we obtain

$$\lim_{T\to\infty}\lim_{d\to\infty}\tilde{\mathsf{b}}_T=\frac{1}{\delta}\,\mathbb{E}\left\{f_0'\left(\frac{a}{\sqrt{\delta}}X+\frac{\sqrt{1-a^2}}{\sqrt{\delta}}W\right)\right\}\ \text{a.s.} \tag{B.62}$$

From lemma B.2, we have $\beta_{T-1}\to 1/\sqrt{\delta}$ as $T\to\infty$. Also recall from assumption **(A2)** on page 4 that τ is the supremum of the support of Z_s, and that $\lambda_\delta^*>\tau$. Therefore, $Z=\frac{Z_s}{\lambda_\delta^*-Z_s}$ has bounded support, due to which $\left\|\boldsymbol{Z}^2\Big(\boldsymbol{I}+\boldsymbol{Z}/(\sqrt{\delta}\beta_{T-1})\Big)^{-1}\right\|_{\text{op}}^2<C$ for a universal constant $C>0$. Hence,

$$\lim_{T\to\infty}\lim_{d\to\infty}S_2=0\ \text{ a.s..} \tag{B.63}$$

To bound S_3, we first write the term inside the norm on the second line of (B.56) as

$$\begin{aligned}&\frac{\sqrt{\delta}}{\lambda_\delta^*}\boldsymbol{Z}_s\boldsymbol{A}\,\hat{\boldsymbol{x}}^0(\bar{\mathsf{b}}_0-\tilde{\mathsf{b}}_T)+\frac{\tilde{\mathsf{b}}_T}{\lambda_\delta^*}\boldsymbol{Z}_s\boldsymbol{A}\left(\sqrt{\delta}\,\hat{\boldsymbol{x}}^0-\frac{\tilde{\boldsymbol{x}}^{T-1}}{\beta_{T-1}}\right)\\&\quad+\frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\left(\frac{\boldsymbol{Z}_s}{\lambda_\delta^*}-\boldsymbol{Z}\left(\boldsymbol{I}+\frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\right)\boldsymbol{A}\,\tilde{\boldsymbol{x}}^{T-1}.\end{aligned}$$

Then, using triangle inequality and Cauchy–Schwarz, we have

J. Stat. Mech. (2022) 114003

$$S_3 \leqslant \frac{3\delta}{(\lambda_\delta^*)^2}\frac{\|\boldsymbol{Z}_s\boldsymbol{A}\,\hat{\boldsymbol{x}}^0\|_2^2}{d}(\bar{\mathsf{b}}_0 - \tilde{\mathsf{b}}_T)^2 + \frac{3\tilde{\mathsf{b}}_T^2}{(\lambda_\delta^*)^2}\|\boldsymbol{Z}_s\boldsymbol{A}\|_{\rm op}^2\frac{\left\|\sqrt{\delta}\,\hat{\boldsymbol{x}}^0 - \tilde{\boldsymbol{x}}^{T-1}/\beta_{T-1}\right\|_2^2}{d}$$
$$+ \frac{3\tilde{\mathsf{b}}_T^2}{\beta_{T-1}^2}\frac{\left\|\boldsymbol{A}\,\tilde{\boldsymbol{x}}^{T-1}\right\|_2^2}{d}\left\|\frac{1}{\lambda_\delta^*}\boldsymbol{Z}_s - \boldsymbol{Z}\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\right\|_{\rm op}^2$$
$$:= 3(S_{3a} + S_{3b} + S_{3c}). \tag{B.64}$$

Using the expression for $\hat{\boldsymbol{x}}^0$ from (5.11) and applying Cauchy–Schwarz, we can bound S_{3a} as:

$$S_{3a} \leqslant \frac{1}{(\lambda_\delta^*)^2}\|\boldsymbol{Z}_s\|_{\rm op}^2\|\boldsymbol{A}\|_{\rm op}^2\|\,\hat{\boldsymbol{x}}^{\rm s}\|_2^2(\bar{\mathsf{b}}_0 - \tilde{\mathsf{b}}_T)^2. \tag{B.65}$$

We note that Z_s is bounded, $\|\,\hat{\boldsymbol{x}}^{\rm s}\|_2 = 1$, and $\|\boldsymbol{A}\|_{\rm op}^2$ is bounded almost surely by a universal constant for sufficiently large d. Moreover, recalling the definitions of $\bar{\mathsf{b}}_0$ and $X_0 = \mu_{X,0}X + \sigma_{X,0}W_{X,0}$ from (5.15) and (3.9), we see that $\bar{\mathsf{b}}_0 = \frac{1}{\delta}\mathbb{E}\{f_0'(X_0)\}$ is the limit of $\tilde{\mathsf{b}}_T$ in (B.62). Therefore $\lim_{T\to\infty}\lim_{d\to\infty} S_{3a} = 0$ almost surely.

Next, we bound S_{3b}. Recalling that $\hat{\boldsymbol{x}}^0 = \sqrt{d}\,\hat{\boldsymbol{x}}^{\rm s}/\sqrt{\delta}$, we have

$$\frac{\left\|\sqrt{\delta}\,\hat{\boldsymbol{x}}^0 - \tilde{\boldsymbol{x}}^{T-1}/\beta_{T-1}\right\|_2^2}{d} = \frac{\left\|\sqrt{d}\,\hat{\boldsymbol{x}}^{\rm s} - \sqrt{\delta}\,\tilde{\boldsymbol{x}}^T + \sqrt{\delta}\,\tilde{\boldsymbol{x}}^T - \sqrt{\delta}\,\tilde{\boldsymbol{x}}^{T-1} + \sqrt{\delta}\,\tilde{\boldsymbol{x}}^{T-1} - \tilde{\boldsymbol{x}}^{T-1}/\beta_{T-1}\right\|_2^2}{d}$$
$$\leqslant \frac{3\left\|\sqrt{d}\,\hat{\boldsymbol{x}}^{\rm s} - \sqrt{\delta}\,\tilde{\boldsymbol{x}}^T\right\|_2^2}{d} + \frac{3\left\|\sqrt{\delta}\,\tilde{\boldsymbol{x}}^T - \sqrt{\delta}\,\tilde{\boldsymbol{x}}^{T-1}\right\|_2^2}{d}$$
$$+ \frac{3\|\,\tilde{\boldsymbol{x}}^{T-1}\|_2^2}{d}(\sqrt{\delta} - 1/\beta_{T-1})^2. \tag{B.66}$$

Lemmas B.3 and B.4 imply that the first two terms on the rhs of (B.66) tend to zero in the iterated limit $T\to\infty$, $d\to\infty$. Furthermore, from lemma B.2, we have $\lim_{T\to\infty}\beta_{T-1} = 1/\sqrt{\delta}$. From proposition B.1, we also have

$$\lim_{d\to\infty}\frac{\left\|\,\tilde{\boldsymbol{x}}^{T-1}\right\|_2^2}{d} = \mu_{\tilde{X},T-1}^2 + \sigma_{\tilde{X},T-1}^2 = \beta_{T-1}^2 \quad \text{a.s.} \tag{B.67}$$

Therefore, $\lim_{T\to\infty}\lim_{d\to\infty} S_{3b} = 0$ almost surely.

To bound S_{3c}, recalling from (B.52) that $Z = \frac{Z_s}{\lambda_\delta^* - Z_s}$, we have

$$\frac{1}{\lambda_\delta^*}\boldsymbol{Z}_s - \boldsymbol{Z}\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1} = \frac{1}{\beta_{T-1}}\boldsymbol{Z}_s^2\left(\lambda_\delta^*\boldsymbol{I} + \boldsymbol{Z}_s\left(\frac{1}{\sqrt{\delta}\beta_{T-1}} - 1\right)\right)^{-1}$$
$$\times\frac{\left(\frac{1}{\sqrt{\delta}} - \beta_{T-1}\right)}{\lambda_\delta^*}. \tag{B.68}$$

J. Stat. Mech. (2022) 114003

Since $\lim_{T\to\infty}\beta_{T-1}=\frac{1}{\sqrt{\delta}}$, almost surely

$$\lim_{T\to\infty}\left\|\frac{1}{\lambda_\delta^*}\boldsymbol{Z}_s-\boldsymbol{Z}\left(\boldsymbol{I}+\frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\right\|_{\text{op}}^2=0. \tag{B.69}$$

Thus $\lim_{T\to\infty}\lim_{d\to\infty}S_{3c}=0$ almost surely. Using the results above in (B.64), we have shown that

$$\lim_{T\to\infty}\lim_{d\to\infty}S_3=0\quad\text{a.s.} \tag{B.70}$$

Using (B.59), (B.63) and (B.70) in (B.56), and recalling that $n/d\to\delta$, we obtain

$$\lim_{T\to\infty}\lim_{n\to\infty}\frac{\left\|\tilde{\boldsymbol{u}}^T-\hat{\boldsymbol{u}}^0\right\|_2}{\sqrt{n}}=0. \tag{B.71}$$

To complete the proof for the base case, we show that the term inside the brackets in (B.50) is finite almost surely as $n\to\infty$. First, by assumption **(B2)** on page 6, we have $\lim_{n\to\infty}\|\boldsymbol{y}\|_2^2/n=\mathbb{E}\{Y^2\}$ almost surely. Furthermore, by proposition B.1, we almost surely have

$$\lim_{n\to\infty}\|\tilde{\boldsymbol{u}}^T\|_2^2/n=\mu_{\tilde{U},T}^2+\sigma_{\tilde{U},T}^2. \tag{B.72}$$

Next, using the triangle inequality, we have

$$\|\tilde{\boldsymbol{u}}^T\|_2-\left\|\tilde{\boldsymbol{u}}^T-\hat{\boldsymbol{u}}^0\right\|_2\leqslant\|\hat{\boldsymbol{u}}^0\|_2\leqslant\|\tilde{\boldsymbol{u}}^T\|_2+\left\|\tilde{\boldsymbol{u}}^T-\hat{\boldsymbol{u}}^0\right\|_2. \tag{B.73}$$

Combining this with (B.71), we obtain

$$\lim_{T\to\infty}\lim_{n\to\infty}\frac{\|\hat{\boldsymbol{u}}^0\|_2^2}{n}=\lim_{T\to\infty}\mu_{\tilde{U},T}^2+\sigma_{\tilde{U},T}^2=\mu_{U,0}^2+\sigma_{U,0}^2\quad\text{a.s.} \tag{B.74}$$

Therefore, using (B.50), we have shown that

$$\lim_{T\to\infty}\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^n\psi(y_i,\,\tilde{u}_i^T)-\frac{1}{n}\sum_{i=1}^n\psi(y_i,\,\hat{u}_i^0)\right|=0\quad\text{a.s.} \tag{B.75}$$

Induction step: assume that (B.15) holds for some t, and that (B.16) holds with $t+1$ replaced by t. Also assume toward induction that almost surely

$$\lim_{T\to\infty}\lim_{d\to\infty}\frac{\left\|\tilde{\boldsymbol{x}}^{T+t}-\hat{\boldsymbol{x}}^t\right\|_2^2}{d}=0,\qquad\lim_{T\to\infty}\lim_{n\to\infty}\frac{\left\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^t\right\|_2^2}{n}=0. \tag{B.76}$$

The limits in (B.76) hold for $t=0$, as established in the proof of the base case (see (B.66) and (B.71)).

From (B.48), we have the bound

J. Stat. Mech. (2022) 114003

$$\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\,\tilde{x}_i^{\,T+t+1})-\frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\,\hat{x}_i^{\,t+1})\right|$$
$$\leqslant 4C\left[1+\frac{\|\boldsymbol{x}\|_2^2}{d}+\frac{\|\tilde{\boldsymbol{x}}^{T+t+1}\|_2^2}{d}+\frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d}\right]^{\frac{1}{2}}\frac{\|\tilde{\boldsymbol{x}}^{T+t+1}-\hat{\boldsymbol{x}}^{t+1}\|_2}{\sqrt{d}}. \quad \text{(B.77)}$$

Using (5.1), (5.6), (5.13) and the triangle inequality, we obtain:

$$\begin{aligned}\frac{\|\tilde{\boldsymbol{x}}^{T+t+1}-\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d} &\leqslant \frac{2}{\delta d}\|\boldsymbol{A}^{\mathsf{T}}h_t(\tilde{\boldsymbol{u}}^{T+t};\boldsymbol{y})-\boldsymbol{A}^{\mathsf{T}}h_t(\hat{\boldsymbol{u}}^{t};\boldsymbol{y})\|_2^2\\ &\quad+2\,\frac{\|\tilde{\mathsf{c}}_{T+t}f_t(\tilde{\boldsymbol{x}}^{T+t})-\bar{\mathsf{c}}_t f_t(\hat{\boldsymbol{x}}^{t})\|_2^2}{d}\\ &\leqslant \frac{2}{\delta d}\|\boldsymbol{A}^{\mathsf{T}}h_t(\tilde{\boldsymbol{u}}^{T+t};\boldsymbol{y})-\boldsymbol{A}^{\mathsf{T}}h_t(\hat{\boldsymbol{u}}^{t};\boldsymbol{y})\|_2^2\\ &\quad+4\,\frac{\|f_t(\tilde{\boldsymbol{x}}^{T+t})\|_2^2}{d}(\tilde{\mathsf{c}}_{T+t}-\bar{\mathsf{c}}_t)^2+4\,\bar{\mathsf{c}}_t^{\,2}\frac{\|f_t(\tilde{\boldsymbol{x}}^{T+t})-f_t(\hat{\boldsymbol{x}}^{t})\|_2^2}{d}\\ &:=2\,S_1+4\,S_2+4\,S_3. \end{aligned}\quad \text{(B.78)}$$

The term S_1 can be bounded as

$$S_1\leqslant\|\boldsymbol{A}\|_{\mathrm{op}}^2\,\frac{\|h_t(\tilde{\boldsymbol{u}}^{T+t};\boldsymbol{y})-h_t(\hat{\boldsymbol{u}}^{t};\boldsymbol{y})\|_2^2}{\delta d}\leqslant\|\boldsymbol{A}\|_{\mathrm{op}}^2\,\bar{L}_t^{\,2}\frac{\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^{t}\|_2^2}{\delta d}, \quad \text{(B.79)}$$

where $\bar{L}_t$ is the Lipschitz constant of the function h_t. Since the operator norm of $\boldsymbol{A}$ is bounded almost surely as $d\to\infty$, by the induction hypothesis (B.76) we have $\lim_{T\to\infty}\lim_{d\to\infty}\frac{\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^{t}\|^2}{\delta d}=0$ almost surely. Therefore,

$$\lim_{T\to\infty}\lim_{d\to\infty}S_1=0\quad\text{a.s.}. \quad \text{(B.80)}$$

To bound S_2, we recall from (5.7) that $\tilde{\mathsf{c}}_{T+t}=\frac{1}{n}\sum_i h_t'(\tilde{u}_i^{\,t};y_i)$. Proposition B.1 guarantees that the joint empirical distribution of $(\tilde{\boldsymbol{u}}^{T+t},\boldsymbol{y})$ converges to the law of $(\tilde{U}_{T+t},Y)\equiv(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,T+t},\,Y)$. Since h_t is Lipschitz, lemma C.1 in appendix C implies that

$$\lim_{n\to\infty}\tilde{\mathsf{c}}_{T+t}=\mathbb{E}\{h_t'(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,T+t},\,Y)\}\quad\text{a.s.}. \quad \text{(B.81)}$$

From (B.13), we know that $\lim_{T\to\infty}\mu_{\tilde{U},T+t}=\mu_{U,t}$ and $\lim_{T\to\infty}\sigma^2_{\tilde{U},T+t}=\sigma^2_{U,t}$. Therefore applying lemma C.1 in appendix C again, we obtain:

$$\lim_{T\to\infty}\lim_{n\to\infty}\tilde{\mathsf{c}}_{T+t}=\mathbb{E}\{h_t'(\mu_{U,t}G+\sigma_{U,t}W_{U,t},\,Y)\}=\bar{\mathsf{c}}_t\quad\text{a.s.} \quad \text{(B.82)}$$

Next, using the result in proposition B.1 with the test function $\psi(x,\tilde{x})=(f_t(\tilde{x}))^2$, we almost surely have

$$\lim_{T\to\infty}\lim_{d\to\infty}\frac{\left\|f_t(\tilde{\boldsymbol{x}}^{T+t})\right\|_2^2}{d}=\lim_{T\to\infty}\mathbb{E}\{f_t(\tilde{X}_{T+t})^2\}=\mathbb{E}\{f_t(X_t)^2\},\tag{B.83}$$

where the last equality follows from (B.13) since f_t is Lipschitz. Combining the above with (B.82), we obtain

$$\lim_{T\to\infty}\lim_{d\to\infty}S_2=0\quad\text{a.s.}.\tag{B.84}$$

For the third term S_3 in (B.78), since f_t is Lipschitz (with Lipschitz constant denoted by L_t), we have the bound:

$$S_3\leqslant\bar{\mathtt{c}}_t^2L_t^2\frac{\left\|\tilde{\boldsymbol{x}}^{T+t}-\boldsymbol{x}^t\right\|_2^2}{d}.\tag{B.85}$$

Thus, by the induction hypothesis (B.76), we obtain

$$\lim_{T\to\infty}\lim_{d\to\infty}S_3=0\quad\text{a.s.}.\tag{B.86}$$

We have therefore shown that

$$\lim_{T\to\infty}\lim_{d\to\infty}\frac{\left\|\tilde{\boldsymbol{x}}^{T+t+1}-\hat{\boldsymbol{x}}^{t+1}\right\|^2}{d}=0\quad\text{a.s.}.\tag{B.87}$$

Next, we show that the terms inside the brackets on the rhs of (B.77) are finite almost surely as $d\to\infty$. Using the pseudo-Lipschitz test function $\psi(x,\tilde{x})=x^2+\tilde{x}^2$, proposition B.1 implies that almost surely

$$\lim_{d\to\infty}\frac{1}{d}\sum_{i=1}^{d}\left(|x_i|^2+|\tilde{x}_i^{T+t+1}|^2\right)=\mathbb{E}\{X^2\}+\mu^2_{\tilde{X},T+t+1}+\sigma^2_{\tilde{X},T+t+1}.\tag{B.88}$$

Moreover, (B.14) implies that $\lim_{T\to\infty}\mu^2_{\tilde{X},T+t+1}+\sigma^2_{\tilde{X},T+t+1}=\mu^2_{X,t+1}+\sigma^2_{X,t+1}$. Using the triangle inequality, we have

$$\|\tilde{\boldsymbol{x}}^{T+t+1}\|_2-\|\tilde{\boldsymbol{x}}^{T+t+1}-\hat{\boldsymbol{x}}^{t+1}\|_2\leqslant\|\hat{\boldsymbol{x}}^{t+1}\|_2\leqslant\|\tilde{\boldsymbol{x}}^{T+t+1}\|_2+\left\|\hat{\boldsymbol{x}}^{t+1}-\tilde{\boldsymbol{x}}^{T+t+1}\right\|_2.\tag{B.89}$$

Hence, using (B.87) and proposition B.1, we almost surely have

$$\begin{aligned}\lim_{T\to\infty}\lim_{d\to\infty}\frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d}&=\lim_{T\to\infty}\lim_{d\to\infty}\frac{\|\tilde{\boldsymbol{x}}^{T+t+1}\|_2^2}{d}\\&=\lim_{T\to\infty}\left(\mu^2_{\tilde{X},T+t+1}+\sigma^2_{\tilde{X},T+t+1}\right)=\mu^2_{X,t+1}+\sigma^2_{X,t+1}.\end{aligned}\tag{B.90}$$

We have thus shown via (B.77) that almost surely

$$\lim_{T\to\infty}\lim_{d\to\infty}\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\tilde{x}_i^{T+t+1})-\frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\hat{x}_i^{t+1})\right|=0.\tag{B.91}$$

J. Stat. Mech. (2022) 114003

To complete the proof via induction, we need to show that if (B.87) and (B.91) hold with $(t+1)$ replaced by t for some $t > 0$, then almost surely

$$\lim_{T\to\infty}\lim_{n\to\infty}\frac{\left\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^{t}\right\|_2^2}{n}=0,\qquad \lim_{T\to\infty}\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\,\tilde{u}_i^{\,T+t})-\frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\,\hat{u}_i^{\,t})\right|=0. \tag{B.92}$$

From (B.49), we have the bound

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\,\tilde{u}_i^{\,T+t})-\frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\,\hat{u}_i^{\,t})\right|\\&\qquad\leqslant 4C\left[1+\frac{\|\boldsymbol{y}\|_2^2}{n}+\frac{\|\tilde{\boldsymbol{u}}^{T+t}\|_2^2}{n}+\frac{\|\hat{\boldsymbol{u}}^{t}\|_2^2}{n}\right]^{\frac{1}{2}}\frac{\left\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^{t}\right\|_2}{\sqrt{n}}.\end{aligned} \tag{B.93}$$

Using (5.2), (5.6), (5.14) and the triangle inequality, we obtain

$$\begin{aligned}\frac{\left\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^{t}\right\|_2^2}{n}&\leqslant\frac{2}{\delta n}\left\|\boldsymbol{A}f_t(\tilde{\boldsymbol{x}}^{T+t})-\boldsymbol{A}f_t(\hat{\boldsymbol{x}}^{t})\right\|_2^2\\&\quad+2\frac{\left\|\tilde{\mathsf{b}}_{T+t}h_{t-1}(\tilde{\boldsymbol{u}}^{T+t-1};\boldsymbol{y})-\bar{\mathsf{b}}_t h_{t-1}(\hat{\boldsymbol{u}}^{t-1};\boldsymbol{y})\right\|_2^2}{n}\\&\leqslant\frac{2}{\delta n}\left\|\boldsymbol{A}f_t(\tilde{\boldsymbol{x}}^{T+t})-\boldsymbol{A}f_t(\hat{\boldsymbol{x}}^{t})\right\|_2^2+4\frac{\|h_{t-1}(\hat{\boldsymbol{u}}^{t-1};\boldsymbol{y})\|_2^2}{n}(\tilde{\mathsf{b}}_{T+t}-\bar{\mathsf{b}}_t)^2\\&\quad+4\bar{\mathsf{b}}_t^{\,2}\frac{\left\|h_{t-1}(\tilde{\boldsymbol{u}}^{T+t-1};\boldsymbol{y})-h_{t-1}(\hat{\boldsymbol{u}}^{t-1};\boldsymbol{y})\right\|_2^2}{n}\\&:=2\,S_1+4\,S_2+4\,S_3.\end{aligned} \tag{B.94}$$

Using arguments along the same lines as (B.80)–(B.86) (omitted for brevity), we can show that almost surely

$$\lim_{T\to\infty}\lim_{n\to\infty}S_1=\lim_{T\to\infty}\lim_{n\to\infty}S_2=\lim_{T\to\infty}\lim_{n\to\infty}S_3=0.$$

Hence $\lim_{T\to\infty}\lim_{n\to\infty}\frac{\left\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^{t}\right\|_2}{\sqrt{n}}=0$ almost surely. Furthermore, using a triangle inequality argument as in (B.89), we obtain $\lim_{T\to\infty}\lim_{n\to\infty}\frac{\|\tilde{\boldsymbol{u}}^{T+t}\|_2^2}{n}=\lim_{T\to\infty}\lim_{n\to\infty}\frac{\|\hat{\boldsymbol{u}}^{t}\|_2^2}{n}$ almost surely. By proposition B.1 and (B.13), the latter limit equals $\mu_{U,t}^2+\sigma_{U,t}^2$. Using these limits in (B.93) yields the result (B.92), and completes the proof of the lemma. □

J. Stat. Mech. (2022) 114003

B.4. Putting everything together: proof of theorem 1

We will first use lemma B.5 to show that the result of the theorem holds for the GAMP iteration $(\hat{\boldsymbol{x}}^t, \hat{\boldsymbol{u}}^t)$, i.e. under the assumptions of theorem 1, we almost surely have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(y_i, \hat{u}_i^t) = \mathbb{E}\{\psi(Y, \mu_{U,t}G + \sigma_{U,t}W_{U,t})\}, \quad t \geqslant 0, \tag{B.95}$$

$$\lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \hat{x}_i^{t+1}) = \mathbb{E}\{\psi(X, \mu_{X,t+1}X + \sigma_{X,t+1}W_{X,t+1}\}, \quad t+1 \geqslant 0. \tag{B.96}$$

Consider the lhs of (B.96). Using the triangle inequality, for any $T > 0$, we have

$$\begin{aligned}
&\left|\frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \hat{x}_i^{t+1}) - \mathbb{E}\{\psi(X, \mu_{X,t+1}X + \sigma_{X,t+1}W_{X,t+1}\}\right| \\
&\quad\leqslant \left|\frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \hat{x}_i^{t+1}) - \frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \tilde{x}_i^{T+t+1})\right| \\
&\qquad + \left|\frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \tilde{x}_i^{T+t+1}) - \mathbb{E}\{\psi(X, \mu_{\tilde{X},T+t+1}X + \sigma_{\tilde{X},T+t+1}W_{\tilde{X},T+t+1})\}\right| \\
&\qquad + \left|\mathbb{E}\{\psi(X, \mu_{\tilde{X},T+t+1}X + \sigma_{\tilde{X},T+t+1}W_{\tilde{X},T+t+1})\}\right. \\
&\qquad \left. - \mathbb{E}\{\psi(X, \mu_{X,t+1}X + \sigma_{X,t+1}W_{X,t+1})\}\right| \\
&\quad := T_1 + T_2 + T_3.
\end{aligned} \tag{B.97}$$

We first bound T_3 using the pseudo-Lipschitz property of ψ, noting that $W_{\tilde{X},T+t}$ and $W_{X,t}$ are both $\sim \mathsf{N}(0,1)$:

$$\begin{aligned}
T_3 &\leqslant \mathbb{E}\{\left|\psi(X, \mu_{\tilde{X},T+t+1}X + \sigma_{\tilde{X},T+t+1}W) - \psi(X, \mu_{X,t+1}X + \sigma_{X,t+1}W)\right|\}, \qquad W \sim \mathsf{N}(0,1) \\
&\leqslant C\mathbb{E}\Big\{\Big(1 + \left[X^2 + \mu_{\tilde{X},T+t+1}^2X^2 + \sigma_{\tilde{X},T+t+1}^2W^2\right]^{1/2} \\
&\quad + \left[X^2 + \mu_{X,t+1}^2X^2 + \sigma_{X,t+1}^2W^2\right]^{1/2}\Big) \\
&\quad \cdot \left(X^2(\mu_{\tilde{X},T+t+1} - \mu_{X,t+1})^2 + W^2(\sigma_{\tilde{X},T+t+1} - \sigma_{X,t+1})^2\right)^{1/2}\Big\} \\
&\leqslant 3C\Big(3 + \mu_{\tilde{X},T+t+1}^2 + \sigma_{\tilde{X},T+t+1}^2 + \mu_{X,t+1}^2 + \sigma_{X,t+1}^2\Big)^{1/2} \\
&\quad \cdot \left((\mu_{\tilde{X},T+t+1} - \mu_{X,t+1})^2 + (\sigma_{\tilde{X},T+t+1} - \sigma_{X,t+1})^2\right)^{1/2},
\end{aligned} \tag{B.98}$$

where we have used Cauchy–Schwarz inequality in the last line. From lemma B.5 (equation (B.14)), we know that $\lim_{T\to\infty}\left|\mu_{\tilde{X},T+t+1} - \mu_{X,t+1}\right| = 0$ and $\lim_{T\to\infty}\left|\sigma_{\tilde{X},T+t+1} - \sigma_{X,t+1}\right| = 0$. Therefore, $\lim_{T\to\infty} T_3 = 0$. Next, from (B.16) we have

J. Stat. Mech. (2022) 114003

that $\lim_{T\to\infty}\lim_{d\to\infty} T_1 = 0$ almost surely. Furthermore, by proposition B.1, for any $T > 0$ we almost surely have $\lim_{d\to\infty} T_2 = 0$. Letting $T, d \to \infty$ (with the limit in d taken first) and noting that the lhs of (B.97) does not depend on T, we obtain that (B.96) holds.

The proof of (B.95) uses a bound similar to (B.97) and arguments along the same lines. It is omitted for brevity.

Next, we prove the main result by showing that under the assumptions of the theorem, almost surely

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_i, u_i^t) - \frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\, \hat{u}_i^{\,t})\right| = 0, \qquad \lim_{n\to\infty}\frac{\|\boldsymbol{u}^t - \hat{\boldsymbol{u}}^{\,t}\|_2^2}{n} = 0, \quad t \geqslant 0 \tag{B.99}$$

$$\lim_{d\to\infty}\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i, x_i^{t+1}) - \frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\, \hat{x}_i^{\,t+1})\right| = 0, \qquad \lim_{d\to\infty}\frac{\|\boldsymbol{x}^{t+1} - \hat{\boldsymbol{x}}^{\,t+1}\|_2^2}{d} = 0, \quad t+1 \geqslant 0. \tag{B.100}$$

Combining (B.99) and (B.100) with (B.95) and (B.96) yields the results in (3.11) and (3.12).

The proof of (B.99) and (B.100) is via induction and uses arguments very similar to those to prove (B.15) and (B.16). To avoid repetition we only provide a few steps. Noting that $\boldsymbol{x}^0 = \hat{\boldsymbol{x}}^0$, we now show (B.100), under the induction hypothesis that (B.99) holds and also that (B.100) holds with $t+1$ replaced by t.

Since $\psi \in \mathrm{PL}(2)$, we have

$$\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i, x_i^{t+1}) - \frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\, \hat{x}_i^{\,t+1})\right| \leqslant 4C\left[1 + \frac{\|\boldsymbol{x}\|_2^2}{d} + \frac{\|\boldsymbol{x}^{t+1}\|_2^2}{d} + \frac{\|\,\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d}\right]^{\frac{1}{2}} \times \frac{\|\boldsymbol{x}^{t+1} - \hat{\boldsymbol{x}}^{\,t+1}\|_2}{\sqrt{d}}. \tag{B.101}$$

Furthermore, using the definitions of $\boldsymbol{x}^{t+1}$ and $\hat{\boldsymbol{x}}^{\,t+1}$, and the triangle inequality we have

$$\begin{aligned}\frac{\|\boldsymbol{x}^{t+1} - \hat{\boldsymbol{x}}^{\,t+1}\|_2^2}{d} &\leqslant \frac{2}{\delta d}\left\|\boldsymbol{A}^{\mathsf{T}} h_t(\boldsymbol{u}^t; \boldsymbol{y}) - \boldsymbol{A}^{\mathsf{T}} h_t(\,\hat{\boldsymbol{u}}^{\,t}; \boldsymbol{y})\right\|_2^2 + 4\,\frac{\|f_t(\boldsymbol{x}^t)\|_2^2}{d}(\mathsf{c}_t - \bar{\mathsf{c}}_{\,t})^2 \\ &\quad + 4\,\bar{\mathsf{c}}_t^{\,2}\frac{\|f_t(\boldsymbol{x}^t) - f_t(\,\hat{\boldsymbol{x}}^t)\|_2^2}{d} \\ &\leqslant \frac{2\,\bar{L}_t^{\,2}}{\delta}\|\boldsymbol{A}\|_{\mathrm{op}}^2\frac{\|\boldsymbol{u}^t - \hat{\boldsymbol{u}}^{\,t}\|_2^2}{d} + 4\,\frac{\|f_t(\boldsymbol{x}^t)\|_2^2}{d}(\mathsf{c}_t - \bar{\mathsf{c}}_{\,t})^2 \\ &\quad + 4\,\bar{\mathsf{c}}_t^{\,2}L_t^2\frac{\|\boldsymbol{x}^t - \hat{\boldsymbol{x}}^{\,t}\|_2^2}{d},\end{aligned} \tag{B.102}$$

J. Stat. Mech. (2022) 114003

where L_t, $\bar{L}_t$ are the Lipschitz constants of f_t, h_t, respectively. By the induction hypothesis and lemma C.1, the terms $\frac{\|\boldsymbol{u}^t-\hat{\boldsymbol{u}}^t\|_2^2}{d}$, $\frac{\|\boldsymbol{x}^t-\hat{\boldsymbol{x}}^t\|_2^2}{d}$, and $(\mathsf{c}_t-\bar{\mathsf{c}}_t)^2$ tend to zero. Furthermore, by the induction hypothesis, we almost surely have $\frac{\|f_t(\boldsymbol{x}^t)\|_2^2}{d}\to\mathbb{E}\{f_t(X_t)^2\}$, and by (B.96), $\frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d}\to(\mu_{X,t+1}^2+\sigma_{X,t+1}^2)$ as $d\to\infty$. Finally, by a triangle inequality argument analogous to (B.89), we also have

$$\lim_{d\to\infty}\frac{\|\boldsymbol{x}^{t+1}\|_2^2}{d}=\lim_{d\to\infty}\frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d}=(\mu_{X,t+1}^2+\sigma_{X,t+1}^2)\quad\text{a.s.}.$$

Using these limits in (B.101) proves (B.100). The proof of (B.99) (under the induction hypothesis that (B.100) holds with $(t+1)$ replaced by t) is along the same lines: we use a bound similar to (B.101) and a decomposition of $\frac{\|\boldsymbol{u}^t-\hat{\boldsymbol{u}}^t\|_2^2}{n}$ similar to (B.102). This completes the proof of the theorem. □

Appendix C. An auxiliary lemma

The following result is proved in [BM11, lemma 6].

Lemma C.1. *Let $F:\mathbb{R}^2\to\mathbb{R}$ be a Lipschitz function, and let $F'(u,v)$ denote its derivative with respect to the first argument at $(u,v)\in\mathbb{R}^2$. Assume that $F'(\cdot,v)$ is continuous almost everywhere in the first argument, for each $v\in\mathbb{R}$. Let (U_m,V_m) be a sequence of random vectors in $\mathbb{R}^2$ converging in distribution to the random vector (U,V) as $m\to\infty$. Furthermore, assume that the distribution of U is absolutely continuous with respect to the Lebesgue measure. Then,*

$$\lim_{m\to\infty}\mathbb{E}\{F'(U_m,V_m)\}=\mathbb{E}\{F'(U,V)\}.$$

Appendix D. Complex-valued GAMP

Consider a complex sensing matrix $\boldsymbol{A}$ with rows distributed as $(\boldsymbol{a}_i)\sim_{\text{i.i.d.}}\mathsf{CN}(0,\boldsymbol{I}_d/d))$, for $i\in[n]$. The output of the GLM $\boldsymbol{y}\in\mathbb{C}^n$ is generated as $p_{Y|G}(\boldsymbol{y}|\boldsymbol{g})$, where $\boldsymbol{g}=\boldsymbol{A}\boldsymbol{x}$. The GAMP algorithm for the complex setting has been studied in the context of phase retrieval by [SR14, MXM19]. Here, we briefly review the complex GAMP and present some numerical results for complex GAMP with spectral initialization.

As in section 4, we take f_t to be the identity function, and $h_t=\sqrt{\delta}h_t^*$, where h_t^* is given in (3.18). To obtain a compact state evolution recursion, we initialize with a scaled version of the spectral estimator $\hat{\boldsymbol{x}}^{\text{s}}$:

$$\boldsymbol{x}^0=\sqrt{d}\,\frac{a}{1-a^2}\,\hat{\boldsymbol{x}}^{\text{s}},\qquad \boldsymbol{u}^0=\frac{1}{\sqrt{\delta}}\boldsymbol{A}\boldsymbol{x}^0-\frac{1}{\sqrt{\delta}\lambda_\delta^*}\boldsymbol{Z}_s\boldsymbol{A}\boldsymbol{x}^0. \tag{D.1}$$

J. Stat. Mech. (2022) 114003

The iterates are then computed as:

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}^{\mathsf{H}} h_t^*(\boldsymbol{u}^t\,;\boldsymbol{y}) - \mathsf{c}_t f_t(\boldsymbol{x}^t), \tag{D.2}$$

$$\boldsymbol{u}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}\boldsymbol{x}^{t+1} - \frac{1}{\sqrt{\delta}} h_t^*(\boldsymbol{u}^t\,;\boldsymbol{y}). \tag{D.3}$$

Here, the Onsager coefficient c_t is given by [SR14]:

$$\mathsf{c}_t = \frac{\sqrt{\delta}}{\mathsf{Var}\,(G|U_t = u)}\left(\frac{\mathsf{Var}\,\{G|U_t = u,\, Y = y\}}{\mathsf{Var}\,(G|U_t = u)} - 1\right). \tag{D.4}$$

For this choice of f_t, h_t, the state evolution iteration can be written in terms of a single parameter $\mu_t \equiv \mu_{X,t}$. For $t \geqslant 0$:

$$\begin{aligned} &\mu_{U,t} = \frac{1}{\sqrt{\delta}}\mu_t, \qquad \sigma_{U,t}^2 = \frac{\mu_t}{\delta}, \qquad \sigma_{X,t}^2 = \mu_{X,t} = \mu_t, \\ &\mu_{t+1} = \sqrt{\delta}\mathbb{E}\Big\{|h_t^*(U_t\,;\,Y)|^2\Big\}. \end{aligned} \tag{D.5}$$

The recursion is initialized with $\mu_0 = \frac{a^2}{1-a^2}$. Moreover, the parameter μ_{t+1} can be consistently estimated from the iterate $\boldsymbol{u}^t$ as $\hat{\mu}_{t+1} = \sqrt{\delta}\|h^*(\boldsymbol{u}^t\,;\boldsymbol{y})\|_2^2/n$. It can also be estimated as the positive solution of the quadratic equation $\hat{\mu}_{t+1}^2 + \hat{\mu}_{t+1} = \|\boldsymbol{x}^{t+1}\|_2^2/d$.

We now discuss some numerical results for noiseless (complex) phase retrieval, where $y_i = |(\boldsymbol{A}\boldsymbol{x})_i|^2$, for $i \in [n]$. For a given measurement matrix $\boldsymbol{A}$, note that replacing $\boldsymbol{x}$ by $\mathrm{e}^{\mathrm{i}\theta}\boldsymbol{x}$ leaves the measurement $\boldsymbol{y}$ unchanged. Therefore the performance of any estimator is measured up to a constant phase rotation:

$$\min_{\theta\in[0,2\pi)} \frac{\left|\langle\,\hat{\boldsymbol{x}},\, \mathrm{e}^{\mathrm{i}\theta}\boldsymbol{x}\rangle\right|^2}{\|\boldsymbol{x}\|_2^2\,\|\,\hat{\boldsymbol{x}}\,\|_2^2}. \tag{D.6}$$

Figure 4 shows the performance of GAMP with spectral initialization when the signal $\boldsymbol{x}$ is uniform on the d-dimensional *complex* sphere with radius $\sqrt{d}$, and the sensing vectors $(\boldsymbol{a}_i) \sim_{\text{i.i.d.}} \mathsf{CN}(0, \boldsymbol{I}_d/d)$.

Figure 5 shows the performance with coded diffraction pattern sensing vectors, given by (4.2). The signal $\boldsymbol{x}$ is the image in figure 3(a), which is a $d_1 \times d_2 \times 3$ array with $d_1 = 820$ and $d_2 = 1280$. The three components $\boldsymbol{x}_j \in \mathbb{R}^d$ ($j \in \{1,2,3\}$ and $d = d_1 \cdot d_2$) are treated separately, and the performance is measured via the average squared normalized scalar product $\frac{1}{3}\sum_{j=1}^3 \frac{|\langle \hat{\boldsymbol{x}}_j, \boldsymbol{x}_j\rangle|^2}{\|\,\hat{\boldsymbol{x}}_j\|_2^2\|\boldsymbol{x}_j\|_2^2}$.

The red points in figure 5 are obtained by running the complex GAMP algorithm with spectral initialization, as given in (D.1)–(D.4). We perform $n_{\text{sample}} = 5$ independent trials and show error bars at one standard deviation. For comparison, the black points correspond to the empirical performance of the spectral method alone, and the black curve gives the theoretical prediction for the optimal squared correlation for Gaussian sensing vectors (see theorem 1 of [LAL19]).

J. Stat. Mech. (2022) 114003

References

[AGZ09] Anderson G W, Guionnet A and Zeitouni O 2009 *An Introduction to Random Matrices* (Cambridge: Cambridge University Press)

[BB08] Boufounos P T and Baraniuk R G 2008 One-bit compressive sensing *Conf. Information Sciences and Systems (CISS)* pp 16–21

[BKM+19] Barbier J, Krzakala F, Macris N, Miolane L and Zdeborová L 2019 Optimal errors and phase transitions in high-dimensional generalized linear models *Proc. Natl Acad. Sci. USA* **116** 5451–60

[BM11] Bayati M and Montanari A 2011 The dynamics of message passing on dense graphs, with applications to compressed sensing *IEEE Trans. Inf. Theory* **57** 764–85

[BM12] Bayati M and Montanari A 2012 The LASSO risk for Gaussian matrices *IEEE Trans. Inf. Theory* **58** 1997–2017

[Bol14] Bolthausen E 2014 An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model *Commun. Math. Phys.* **325** 333–66

[BR17] Bahmani S and Romberg J 2017 Phase retrieval meets statistical learning theory: a flexible convex relaxation *Int. Conf. Artificial Intelligence and Statistics (AISTATS)* pp 252–60

[CC17] Chen Y and Candès J 2017 Solving random quadratic systems of equations is nearly as easy as solving linear systems *Commun. Pure Appl. Math.* **70** 822–83

[CESV15] Candès E J, Eldar Y C, Strohmer T and Voroninski V 2015 Phase retrieval via matrix completion *SIAM Rev.* **57** 225–51

[CLS15a] Candès E J, Li X and Soltanolkotabi M 2015 Phase retrieval from coded diffraction patterns *Appl. Comput. Harmon. Anal.* **39** 277–99

[CLS15b] Candes E J, Li X and Soltanolkotabi M 2015 Phase retrieval via Wirtinger flow: theory and algorithms *IEEE Trans. Inf. Theory* **61** 1985–2007

[CSV13] Candès E J, Strohmer T and Voroninski V 2013 Phaselift: exact and stable signal recovery from magnitude measurements via convex programming *Commun. Pure Appl. Math.* **66** 1241–74

[DBMM20] Dudeja R, Bakhshizadeh M, Ma J and Maleki A 2020 Analysis of spectral methods for phase retrieval with random orthogonal matrices *IEEE Trans. Inf. Theory* **66** 5182–203

[DJ17] Demanet L and Jugnon V 2017 Convex recovery from interferometric measurements *IEEE Trans. Comput. Imaging* **3** 282–95

[DJM13] Donoho D L, Javanmard A and Montanari A 2013 Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing *IEEE Trans. Inf. Theory* **59** 7434–64

[DM14] Deshpande Y and Montanari A 2014 Information-theoretically optimal sparse PCA *IEEE Int. Symp. Information Theory (ISIT)* pp 2197–201

[DMM09] Donoho D L, Maleki A and Montanari A 2009 Message-passing algorithms for compressed sensing *Proc. Natl Acad. Sci. USA* **106** 18914–9

[EGK11] El Gamal A and Kim Y-H 2011 *Network Information Theory* (Cambridge: Cambridge University Press)

[EK12] Eldar Y C and Kutyniok G 2012 *Compressed Sensing: Theory and Applications* (Cambridge: Cambridge University Press)

[ESAP+20] Emami M, Sahraee-Ardakan M, Pandit P, Rangan S and Fletcher A 2020 Generalization error of generalized linear models in high dimensions *Int. Conf. Machine Learning, PMLR* pp 2892–901

[Fan20] Zhou F 2020 Approximate message passing algorithms for rotationally invariant matrices (arXiv:2008.11892)

[FD87] Fienup C and Dainty J 1987 Phase retrieval and image reconstruction for astronomy *Image Recovery: Theory and Application* (New York: Academic) vol 231 p 275

[Fie82] Fienup J R 1982 Phase retrieval algorithms: a comparison *Appl. Opt.* **21** 2758–69

[FRS18] Fletcher A K, Rangan S and Schniter P 2018 Inference in deep networks in high dimensions *2018 IEEE Int. Symp. Information Theory (ISIT)* (Piscataway, NJ: IEEE) pp 1884–8

[FS20] Fannjiang A and Strohmer T 2020 The numerics of phase retrieval (arXiv:2004.05788)

[GS18] Goldstein T and Studer C 2018 PhaseMax: convex phase retrieval via basis pursuit *IEEE Trans. Inf. Theory* **64** 2675–89

[JM13] Javanmard A and Montanari A 2013 State evolution for general approximate message passing algorithms, with applications to spatial coupling *Inf. Inference* **2** 115–44

[KKM+16] Kabashima Y, Krzakala F, Mezard M, Sakata A and Zdeborova L 2016 Phase transitions and sample complexity in Bayes-optimal matrix factorization *IEEE Trans. Inf. Theory* **62** 4228–65

[KMS+12] Krzakala F, Mézard M, Sausset F, Sun Y and Zdeborová L 2012 Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices *J. Stat. Mech.* P08009

[LAL19] Luo W, Alghamdi W and Lu Y M 2019 Optimal spectral initialization for signal recovery with applications to phase retrieval *IEEE Trans. Signal Process.* **67** 2347–56

[LGL15] Li G, Gu Y and Lu Y M 2015 Phase retrieval using iterative projections: dynamics in the large systems limit *Allerton Conf. Communication, Control, and Computing* (Allerton) pp 1114–8

J. Stat. Mech. (2022) 114003

[Li92] Li K-C 1992 On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma *J. Am. Stat. Assoc.* **87** 1025–39
[LL19] Lu Y M and Li G 2019 Phase transitions of spectral initialization for high-dimensional non-convex estimation *Inf. Inference* **9** 507–41
[LWL20] Luo Q, Wang H and Lin S 2020 Phase retrieval via smoothed amplitude flow *Signal Process.* **177** 107719
[MAYB13] Maleki A, Anitori L, Yang Z and Baraniuk R G 2013 Asymptotic analysis of complex lasso via complex approximate message passing (CAMP) *IEEE Trans. Inf. Theory* **59** 4290–308
[McC18] McCullagh P 2018 *Generalized Linear Models* (New York: Routledge)
[Mil90] Millane R P 1990 Phase retrieval in crystallography and optics *J. Opt. Soc. Am.* A **7** 394–411
[MLKZ20] Antoine M, Loureiro B, Krzakala F and Zdeborová L 2020 Phase retrieval in high dimensions: statistical and computational phase transitions (arXiv:2006.05228)
[MM19] Mondelli M and Montanari A 2019 Fundamental limits of weak recovery with applications to phase retrieval *Found. Comput. Math.* **19** 703–73
[MTV20] Mondelli M, Thrampoulidis C and Venkataramanan R 2020 Optimal combination of linear and spectral estimators for generalized linear models (arXiv:2008.03326)
[MV21] Montanari A and Venkataramanan R 2021 Estimation of low-rank matrices via approximate message passing *Ann. Stat.* **45** 321–45
[MWCC20] Ma C, Wang K, Chi Y and Chen Y 2020 Implicit regularization in nonconvex statistical estimation: gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution *Found. Comput. Math.* **20** 451–632
[MXM18] Ma J, Xu J and Maleki A 2018 Approximate message passing for amplitude based optimization *Int. Conf. Machine Learning (ICML)* pp 3371–80
[MXM19] Ma J, Xu J and Maleki A 2019 Optimization-based AMP for phase retrieval: the impact of initialization and ℓ_2 regularization *IEEE Trans. Inf. Theory* **65** 3600–29
[NJS13] Netrapalli P, Jain P and Sanghavi S 2013 Phase retrieval using alternating minimization *Advances in Neural Information Processing Systems (NIPS)* pp 2796–804
[PSAR+20] Pandit P, Sahraee-Ardakan M, Rangan S, Schniter P and Fletcher A K 2020 Inference with deep generative priors in high dimensions *IEEE J. Sel. Areas Inf. Theory* **1** 336–47
[PWBM18] Perry A, Wein S W, Bandeira S B and Moitra A 2018 Message-passing algorithms for synchronization problems over compact groups *Commun. Pure Appl. Math.* **71** 2275–322
[Ran11] Rangan S 2011 Generalized approximate message passing for estimation with random linear mixing *IEEE Int. Symp. Information Theory (ISIT)*
[RF12] Rangan S and Fletcher A K 2012 Iterative estimation of constrained rank-one matrices in noise *IEEE Int. Symp. Information Theory (ISIT)* pp 1246–50
[RG01] Rangan S and Goyal V K 2001 Recursive consistent estimation with bounded noise *IEEE Trans. Inf. Theory* **47** 457–64
[SC19] Sur P and Candès J 2019 A modern maximum-likelihood theory for high-dimensional logistic regression *Proc. Natl Acad. Sci. USA* **116** 14516–25
[SEC+15] Shechtman Y, Eldar Y C, Cohen O, Chapman H N, Miao J and Segev M 2015 Phase retrieval with application to optical imaging: a contemporary overview *IEEE Signal Process. Mag.* **32** 87–109
[SR14] Schniter P and Rangan S 2014 Compressive phase retrieval via generalized approximate message passing *IEEE Trans. Signal Process.* **63** 1043–55
[SRF16] Schniter P, Rangan S and Fletcher A K 2016 Vector approximate message passing for the generalized linear model *50th Asilomar Conf. Signals, Systems and Computers* (Piscataway, NJ: IEEE) pp 1525–9
[TV19] Tan Y S and Vershynin V 2019 Phase retrieval via randomized Kaczmarz: theoretical guarantees *Inf. Inference* **8** 97–123
[Vil08] Villani C 2008 *Optimal Transport: Old and New* vol 338 (Berlin: Springer)
[WdM15] Waldspurger I, d'Aspremont A and Mallat S 2015 Phase recovery, maxcut and complex semidefinite programming *Math. Program.* **149** 47–81
[Wei15] Ke W 2015 Solving systems of phaseless equations via Kaczmarz methods: a proof of concept study *Inverse Problems* **31**
[WR20] Wu F and Rebeschini P 2020 A continuous-time mirror descent approach to sparse phase retrieval *Advances in Neural Information Processing Systems (NeurIPS)* vol 33 pp 20192–203
[YLSV12] Yang F, Lu Y M, Sbaiz L and Vetterli M 2012 Bits from photons: oversampled image acquisition using binary Poisson statistics *IEEE Trans. Image Process.* **21** 1421–36

J. Stat. Mech. (2022) 114003