Journal of Machine Learning Research 24 (2023) 1-49 Submitted 5/21; Revised 2/23; Published 2/23

Fundamental limits and algorithms for sparse linear regression with sublinear sparsity

Lan V. Truong LT407@CAM.AC.UK
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK

Editor: Pierre Alquier

Abstract

We establish exact asymptotic expressions for the normalized mutual information and minimum mean-square-error (MMSE) of sparse linear regression in the sub-linear sparsity regime. Our result is achieved by a generalization of the adaptive interpolation method in Bayesian inference for linear regimes to sub-linear ones. A modification of the well-known approximate message passing algorithm to approach the MMSE fundamental limit is also proposed, and its state evolution is rigorously analysed. Our results show that the traditional linear assumption between the signal dimension and number of observations in the replica and adaptive interpolation methods is not necessary for sparse signals. They also show how to modify the existing well-known AMP algorithms for linear regimes to sub-linear ones.

Keywords: Bayesian Inference, Approximate Message Passing, Replica Method, Interpolation Method.

1. Introduction

The estimation of a signal from linear random observations has a myriad of applications such as compressed sensing, error correction via sparse superposition codes, Boolean group testing, and supervised machine learning. The fitting of linear relationships among variables in a data set is a standard tool in data analysis. More frequently, there exists conditions under which sparse models fit the data quite well. For example, Rosenfeld et al. (Agrawal et al., 1996) used data to mimic heuristics to identify small segments of a population in which a few additional risk factors were highly predictive of certain kinds of cancer, whereas the same risk factors were not significant in the population (Juba, 2016). Estimation errors can be characterized by some standard measures in statistics and information theory such as the minimum mean square error (MMSE) and/or mutual information. These fundamental limits are usually obtained by using the replica or interpolation methods in statistical physics where the number of observations is usually assumed to scale linearly with the signal dimension (Edwards and Anderson, 1975). Accordingly, most of existing approximate message passing algorithms are designed based on the same assumption. However, in many practical applications in machine learning, communications, and signal processing such as medical image recognition and group testing, the number of observations are very small compared with the signal dimension. In this work, we estimate two fundamental limits (MMSE and mutual information) and propose an approximate message passing algorithm to achieve the MMSE for sub-linear regimes where the number of observations scales sub-linearly with the signal dimension.

1.1 Related Papers

In recent years, there has been the progress on a coherent mathematical theory of the replica and interpolation method in statistical physics of spin glasses (Edwards and Anderson, 1975). These methods have been fruitfully extended and adapted to the problems of interest in a wide range of applications in Bayesian inferences, multiuser communications, and theoretical computer science (Tanaka, 2002; Dongning Guo et al., 2005) and (Truong, 2022). The replica method, although very interesting, is based on some non-rigorous assumptions. (Reeves and Pfister, 2016) proved that the replica prediction in (Tanaka, 2002; Dongning Guo et al., 2005) is exact. In more recent years, an adaptive interpolation method has been proposed to prove fundamental limits predicted by replica method in a rigorous way (Barbier and Macris, 2017; Barbier et al., 2018, 2019). Roughly speaking, this method interpolates between the original problem and the mean-field replica solution in small steps, each step involving its own set of trial parameters and Gaussian mean-fields in the sprit of Guerra and Toninelli (Guerra and Toninelli, 2002; Barbier and Macris, 2017). We can adjust the set of trial parameters in various ways so that we get both upper and lower bounds that eventually match.

The "All-or-Nothing" phenomenon for the linear and non-linear models has been characterized in a variety of recent papers. In (Reeves et al., 2019b), Reeves et al. consider a binary k-sparse linear regression problem, where the number of observations m is sub-linear to the signal dimension n, and established an "All-or-Nothing" information-theoretic phase transition at a critical sample size $m^* = 2k \log(n/k)/ \log(1 + k/\Delta_n)$ for two regimes $k/\Delta_n = \Omega(1)$ and $k = o(\sqrt{n})$ with Δ_n being the noise variance. Their results are based on an assumption that the sparse signal is uniformly distributed from the set $\{v \in \{0, 1\}^n : \|v\|_0 = k\}$. (Reeves et al., 2019a) considers a double limit where one first obtains the high-dimensional limit (under linear sparsity) and then considers the limiting behavior of the RS formulas and the AMP state evolution with respect to a family of prior distributions with which allows the prior to scale with the dimensions. However, the analysis reveals that the resulting formulas can simplify dramatically in the sparse regime. Indeed, in certain cases (e.g., Bernoulli prior) the single-letter mutual information function converges to a piecewise linear limit, and this gives rise to an "all-or-nothing" phenomenon. Sharp information-theoretic bounds were established in (Scarlett and Cevher, 2017) and (Truong and Scarlett, 2020) for support recovery problems in linear and phase retrieval models, respectively. In addition, the "All-or-Nothing" phenomenon was also considered for Bernoulli group testing (Truong et al., 2020) or sparse spiked matrix estimation in (Barbier and Macris, 2019), (Luneau et al., 2020), and (Niles-Weed and Zadik, 2020). In (Barbier et al., 2020), this phenomenon was also investigated for the generalized linear models with sub-linear regimes and Bernoulli and Bernoulli-Rademacher distributed vectors.

Although the results achieved by the replica method and the adaptive interpolation counterpart are very interesting, they are mainly constrained to the case where the number of observations scales *linearly* with the signal dimension. (Luneau et al., 2020) considered generalized linear models in regimes where the number of nonzero components of the signal and accessible data points are sublinear with respect to the size of the signal. They obtained a proof of the replica symmetric formula for the linear model for the case $\alpha > 8/9$. In this work, thanks to the development of a new proof technique of the key concentration inequality in (Barbier et al., 2016), we can widen the range of α to all $[0, 1]$ for a similar model. In addition, we develop a variant of the approximate message passing for the sparse linear regression with sub-linear sparsity which can approach the developed fundamental limit for most of simulation cases. Our numerical results show that the weak recovery (and detection) (Reeves et al., 2019b) is possible at various ranges of SNR *under the sparsity in the expected sense* where the *expected* number of nonzero elements, k, in the vector $\mathbf{S}$ is much less than the signal dimension n. For the sparse model in (Reeves et al., 2019b), the number of nonzero elements in each vector $\mathbf{S}$ is always equal to k.

Approximate message passing (AMP) refers to a class of efficient algorithms for statistical estimation in high-dimensional problems such as compressed sensing and low-rank matrix estimation. AMP is initially proposed for sparse signal recovery and compressed sensing (Donoho, 2006; Candès and Wakin, 2008; Metzler et al., 2016). AMP algorithms have been proved to be effective in reconstructing sparse signals from a small number of incoherent linear measurements. Their dynamics are accurately tracked by a simple one-dimensional iteration termed *state evolution* (Bayati and Montanari, 2011). AMP algorithms achieve state-of-the-art performance for several high-dimensional statistical estimation problems, including compressed sensing (Donoho et al., 2009; Bayati and Montanari, 2011; Krzakala et al., 2012) and low-rank matrix estimation (Matsushita and Tanaka, 2013; Deshpande and Montanari, 2014; Kabashima et al., 2016; Montanari and Venkataramanan, 2021). Moreover, these techniques are also popular and practical in a variety of engineering and computer science applications such as imaging (Fletcher and Rangan, 2014; Metzler et al., 2017), communications (Schniter, 2011; Rush et al., 2017), and deep learning (Pandit et al., 2019; Emami et al., 2020; Pandit et al., 2020). See (Feng et al., 2022) for a detailed survey on this research topic. Our results imply that a judicious modification of AMP for linear regimes can work well for sub-linear ones.

1.2 Contributions

In this paper, we consider the same k-sparse linear regression as (Reeves et al., 2019b) but in more general signal domain. However, we assume that the signal is sparse in expected sense as (Barbier et al., 2020) and the number of observation is sub-linear to the signal dimension. Our contributions include:

- We characterize MMSE and mutual information exactly for the sub-linear regimes where $k = O(n^{\alpha})$ and $m = \delta n^{\alpha}$ for some $\alpha \in (0, 1]$. Our result is achieved by a generalization of the adaptive interpolation method in Bayesian inference for linear regimes (Barbier et al., 2016; Barbier and Macris, 2017) to sub-linear ones. Compared with (Barbier et al., 2016; Luneau et al., 2020), the bound (RHS) in the concentration in Lemma 4 is new. We need to develop a new proof to show this concentration inequality for the sub-linear sparsity.

- We design a variant of the classical AMP algorithm (Donoho et al., 2009) for the sub-linear regimes which approaches the information-theoretic fundamental limits for many cases. The state evolution is also rigorously analysed in our work, and we redefine the states in non-asymptotic sense. As a by-product, we generalize a general version of the strong law of large numbers and Hájek-Rényi type maximal inequality, which may be of independent interest.

- We perform some numerical evaluations and show that the gap between MSE achieved by our AMP and the MMSE fundamental limit is very small. Our results also show that the new variant of AMP works well for a wide range of α in $[0,1]$.

1.3 Paper Organization

The problem settings is placed in Section 2, where we introduce the system model and our assumptions. In Section 3, we state some information-theoretic fundamental limits such as the average mutual information and MMSE which are obtained by using a rigorous analysis with the adaptive interpolation method. An approximate message passing is proposed and its performance analysis is given in Section 4. We place some auxiliary but important proofs in the appendices.

2. Problem Settings

2.1 Problem settings

Let $\mathbf{S} \in \mathbb{R}^n$ be a signal observed via a linear model with measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$. Let $\{\Delta_n\}_{n=1}^{\infty}$ be a positive sequence. We consider the same linear model as (Barbier and Macris, 2017):

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{W}\sqrt{\Delta_n}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{S} = (S_1, S_2, \cdots, S_n)^T \in \mathbb{R}^n, \mathbf{W} \in \mathbb{R}^m$, and $\mathbf{Y} \in \mathbb{R}^m$. Instead of assuming that $m = n\delta$ for some $\delta > 0$ as standard literature in replica and adaptive interpolation methods, we assume that $m = \delta n^{\alpha}$ for some $\alpha > 0$ and $0 < \alpha \leq 1$. We also assume:

1. $\mathbf{A}$ is a Gaussian matrix with $A_{ij} \sim \mathcal{N}(0, 1/m)$.
2. $\{S_n\}_{n=1}^{\infty}$ is an i. i. d. sequence with $S_i \sim \tilde{P}_0$, where $\tilde{P}_0(s) = \big(1-k/n\big)\delta(s)+\big(k/n\big)P_0(s)$ for some $k = O(n^{\alpha})$ with $0 < \alpha \leq 1$ and $P_0(s) = \sum_{b=1}^{B} p_b\delta(s-a_b)$ with a finite number B of constant terms such that $s_{\max} := \max_b |a_b| < \infty$.
3. $\mathbf{W} \sim \mathcal{N}(0, \mathbf{I}_m)$.
4. Δ_n can be any function of α, and n.

2.2 Notations

For any $k > 1$, we say a function $\phi : \mathbb{R}^q \to \mathbb{R}$ is *pseudo-Lipschitz* of order k if there exists a constant $L > 0$ such that for all $x, y \in \mathbb{R}^q$:

$$\big|\phi(x) - \phi(y)\big| \leq L\big(1 + \|x\|^{k-1} + \|y\|^{k-1}\big)\|x - y\| \tag{2}$$

where $\|\|$ denotes the Euclidean norm-2. In addition, for any sequence of vectors $\{\mathbf{x}^{(n)}\}_{n=1}^{\infty}$, we denote by $\mathbf{x}_1^n = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}\}$. As standard literature, the mutual information between two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is written as $I(\mathbf{X};\mathbf{Y})$. The transpose of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^*$. The σ-algebra which is generated by the union of two σ-algebras $\mathcal{G}_1$ and $\mathcal{G}_2$ is denoted by $\sigma(\mathcal{G}_1) \cup \sigma(\mathcal{G}_2)$.

3. Information-Theoretic Fundamental Limits

For this case, we assume that the sensing matrix $\mathbf{A}$ has i.i.d. Gaussian components. Similar to (Barbier and Macris, 2017), let

$$\Sigma(u;v)^{-2} := \frac{\delta n^{\alpha-1}}{u+v}, \tag{3}$$

$$\psi(u;v) := \frac{\delta}{2}\bigg[\log\bigg(1+\frac{u}{v}\bigg) - \frac{u}{u+v}\bigg]. \tag{4}$$

Define the following sequence of Replica Symmetric (RS) potentials:

$$f_{n,\mathrm{RS}}(E;\Delta_n) := \psi(E;\Delta_n) + i_{n,\mathrm{den}}(\Sigma(E;\Delta_n)), \tag{5}$$

where $i_{n,\mathrm{den}}(\Sigma) = n^{1-\alpha} I(S; S+\tilde{W}\Sigma)$ is a normalized mutual information of a scalar Gaussian denoising model $Y = S + \tilde{W}\Sigma$ with $S \sim \tilde{P}_0, \tilde{W} \sim \mathcal{N}(0,1)$, and Σ^{-2} an effective signal to noise ratio:

$$i_{n,\mathrm{den}}(\Sigma) := n^{1-\alpha}\mathbb{E}_{S,\tilde{W}}\bigg[\log\int \tilde{P}_0(x)\exp\bigg[-\frac{1}{\Sigma^2}\bigg(\frac{(x-S)^2}{2} - (x-S)\tilde{W}\Sigma\bigg)\bigg]dx. \tag{6}$$

Our information-theoretic fundamental result is the following:

Theorem 1 *Let $\nu_n = n^{\alpha-1}\mathbb{E}_{S\sim P_0}[S^2]$ and $\hat{\mathbf{S}} = \mathbb{E}[\mathbf{S}|\mathbf{Y}]$ be the MMSE estimator. Then, under the condition that $\Delta_n = \Omega_n(1)$[1] and that $\arg\min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n)$ is unique for all Δ_n, in the large system limits, the following holds:*

$$\lim_{n\to\infty}\bigg[\frac{I(\mathbf{S};\mathbf{Y}|\mathbf{A})}{n^{\alpha}} - \min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n)\bigg] = 0, \tag{7}$$

$$\lim_{n\to\infty}\bigg[\frac{1}{n^{\alpha}}\sum_{i=1}^{n}\mathbb{E}[(S_i-\hat{S}_i)^2] - n^{1-\alpha}\tilde{E}(\Delta_n)\bigg] = 0, \tag{8}$$

where $\tilde{E}(\Delta_n)$ is the global minimizer of $\min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n)$.

Remark 2 *Some remarks are in order.*

- *Our theorem allows Δ_n to be dependent on n, where Δ_n is assumed to be fixed in (Barbier and Macris, 2017).*

1. This constraint is less strict than the one in (Barbier and Macris, 2017)

- *For $\alpha = 1$ (or $m = \delta n$), and $\Delta_n = \Delta$ for some fixed $\Delta > 0$, our results recover the classical result as in (Tanaka, 2002; Dongning Guo and Verdu, 2005; Reeves and Pfister, 2019; Barbier et al., 2019). In these classical papers, the authors assume that $\{S_n\}_{n=1}^{\infty}$ are i.i.d and $S_1 \sim \tilde{P}_0$ which is a fixed distribution.*

- *The minimization problem in (7), i.e., $\min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n)$ may have multiple minimizers at some values of Δ_n. The number of minimizers depends on the prior distribution. By limiting $E \in [0, \nu_n]$, we may avoid this phenomenon for some cases.*

- *In our model, the sparsity is in the expected sense, which is different from the models in (Reeves et al., 2019b, Eqn. (3)) or (Scarlett and Cevher, 2017, Cor. 1), where the authors assume that the sparse vector is uniformly distributed over $\binom{n}{k}$ possible sparse vectors. In addition, the equivalent SNR in our model is a fixed constant, but the required SNR for (Reeves et al., 2019b, Theorem 3) to hold is greater than an ambiguous constant. Hence, the weak (strong) recovery is expected to hold at low SNR. In our numerical simulations (cf. Fig. 1), even the classical AMP can (at least) recover the sparse vector weakly, i.e. the normalized MSE (divided by n^α) is almost less than one for many ranges of SNR [2].*

- *Our results show that under the MMSE estimator, the linear regression model can be decomposed into sub-AWGN channels, and the normalized MMSE of the model is equal to the MMSE of a (time-varying SNR) sub-AWGN channel in the large system limit.*

Proof The proof of Theorem 1 is based on (Barbier and Macris, 2017; Barbier et al., 2016) with some modifications in concentration inequalities and normalized factors to account for new settings. Given the model (1), the likelihood of the observation $\mathbf{y}$ given $\mathbf{S}$ and $\mathbf{A}$ is

$$P(\mathbf{y}|\mathbf{s},\mathbf{A}) = \frac{1}{(2\pi\Delta_n)^{m/2}} \exp\left[-\frac{1}{2\Delta_n}\|\mathbf{y}-\mathbf{A}\mathbf{s}\|^2\right]. \tag{9}$$

From Bayes formula we then get the posterior distribution for $\mathbf{x} = [x_1, x_2, \cdots, x_n] \in \mathbb{R}^n$ given the observation $\mathbf{y}$ and sensing matrix $\mathbf{A}$

$$P(\mathbf{x}|\mathbf{y},\mathbf{A}) = \frac{\prod_{i=1}^n \tilde{P}_0(x_i) P(\mathbf{y}|\mathbf{x},\mathbf{A})}{\int \prod_{i=1}^n \tilde{P}_0(x_i) dx_i P(\mathbf{y}|\mathbf{x},\mathbf{A})}. \tag{10}$$

Replacing the observation $\mathbf{y}$ by its explicit expression (1) as a function of the signal and the noise we obtain

$$\begin{aligned} &P(\mathbf{x}|\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{w}\sqrt{\Delta_n}, \mathbf{A}) \\ &= \frac{\prod_{i=1}^n \tilde{P}_0(x_i) e^{-\mathcal{H}(\mathbf{x};\mathbf{A},\mathbf{s},\mathbf{w})}}{\mathcal{Z}(\mathbf{A},\mathbf{s},\mathbf{w})}, \end{aligned} \tag{11}$$

2. We can even verify that the normalized sum of MSE by n^α mostly in $[0, 1]$ by running the classical AMP in (Bayati and Montanari, 2011, Section C), an sub-optimal algorithm for this setting.

where we call

$$\mathcal{H}(\mathbf{x};\mathbf{A},\mathbf{s},\mathbf{w}) := \frac{1}{\Delta_n}\sum_{\mu=1}^{m}\bigg(\frac{1}{2}\Big[\mathbf{A}(\mathbf{x}-\mathbf{s})\Big]_\mu^2 - \Big[\mathbf{A}(\mathbf{x}-\mathbf{s})\Big]_\mu w_\mu\sqrt{\Delta_n}\bigg) \tag{12}$$

the *Hamiltonian* of the model, and the normalization factor is by definition the *partition function*:

$$\mathcal{Z}(\mathbf{A},\mathbf{s},\mathbf{w}) := \int \bigg\{\prod_{i=1}^{n}\tilde{P}_0(x_i)dx_i\bigg\}e^{-\mathcal{H}(\mathbf{x};\mathbf{A},\mathbf{s},\mathbf{w})}. \tag{13}$$

Our principal quantity of interest is

$$f_n = -\frac{1}{n^\alpha}\mathbb{E}_{\mathbf{A},\mathbf{S},\mathbf{W}}[\log\mathcal{Z}(\mathbf{A},\mathbf{S},\mathbf{W})] \tag{14}$$

$$= -\frac{1}{n^\alpha}\mathbb{E}_{\mathbf{A},\mathbf{S},\mathbf{W}}\bigg[\log\bigg(\int\bigg\{\prod_{i=1}^{n}\tilde{P}_0(x_i)dx_i\bigg\} \times \exp\bigg(-\frac{1}{\Delta_n}\sum_{\mu=1}^{m}\bigg(\frac{1}{2}\Big[\mathbf{A}(\mathbf{x}-\mathbf{S})\Big]_\mu^2 - \Big[\mathbf{A}(\mathbf{x}-\mathbf{S})\Big]_\mu W_\mu\sqrt{\Delta_n}\bigg)\bigg)\bigg)\bigg], \tag{15}$$

where $\mathbf{W} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.

By using the Bayes' rule

$$P(\mathbf{y}|\mathbf{A}) = \frac{P(\mathbf{y}|\mathbf{x},\mathbf{A})\prod_{i=1}^{n}\tilde{P}_0(x_i)}{P(\mathbf{x}|\mathbf{y}=\mathbf{A}\mathbf{s}+\mathbf{w}\sqrt{\Delta_n},\mathbf{A})}, \tag{16}$$

we have

$$P(\mathbf{y}|\mathbf{A}) = (2\pi\Delta)^{-m/2}\mathcal{Z}(\mathbf{A},\mathbf{s},\mathbf{w})e^{-\frac{\|\mathbf{w}^2\|}{2}}. \tag{17}$$

It follows that

$$\frac{I(\mathbf{S};\mathbf{Y}|\mathbf{A})}{n^\alpha} = \frac{1}{n^\alpha}\mathbb{E}_{\mathbf{A},\mathbf{S},\mathbf{Y}}\bigg[\log\bigg(\frac{P(\mathbf{S},\mathbf{Y}|\mathbf{A})}{\tilde{P}_0(\mathbf{S})P(\mathbf{Y}|\mathbf{A})}\bigg)\bigg] \tag{18}$$

$$= f_n - \frac{h(\mathbf{Y}|\mathbf{A},\mathbf{S})}{n^\alpha} + \frac{1}{2n^\alpha}\mathbb{E}[\|\mathbf{W}\|^2] + \frac{m}{2n^\alpha}\log(2\pi\Delta_n) \tag{19}$$

$$= f_n - \frac{m}{2n^\alpha}\log(2\pi e\Delta_n) + \frac{m}{2n^\alpha} + \frac{m}{2n^\alpha}\log(2\pi\Delta_n) \tag{20}$$

$$= f_n. \tag{21}$$

Hence, in order to obtain (7), it is enough to show that

$$\lim_{n\to\infty}\Big[f_n - \min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n)\Big] = 0. \tag{22}$$

Let $\mathbf{W}^{(k)} = [W_\mu^{(k)}]_{\mu=1}^m, \tilde{\mathbf{W}}^{(k)} = [\tilde{W}_i^{(k)}]_{i=1}^n$ and $\hat{\mathbf{W}} = [\hat{W}_i]_{i=1}^n$ all with i.i.d. $\mathcal{N}(0,1)$ entries for $k = 1,2,\cdots,K_n$ where K_n is chosen later. Define $\Sigma_k := \Sigma(E_k;\Delta_n)$ where the trial parameters $\{E_k\}_{k=1}^{K_n}$ are determined later on. Given any fixed $\varepsilon \in [0,1]$, as (Barbier and Macris, 2017), the (perturbed) (k,t)-interpolating Hamiltonian for this problem is defined as

$$\mathcal{H}_{k,t;\varepsilon}(\mathbf{x};\mathbf{\Theta}) := \sum_{k'=k+1}^{K_n} h\Big(\mathbf{x},\mathbf{S},\mathbf{A},\mathbf{W}^{(k')},K_n\Delta_n\Big) + \sum_{k'=1}^{k-1} h_{\mathrm{mf}}\Big(\mathbf{x},\mathbf{S},\tilde{\mathbf{W}}^{(k')},K_n\Sigma_{k'}^2\Big)$$
$$+ h\Big(\mathbf{x},\mathbf{S},\mathbf{A},\mathbf{W}^{(k)},\frac{K_n}{\gamma_k(t)}\Big) + h_{\mathrm{mf}}\Big(\mathbf{x},\mathbf{S},\tilde{\mathbf{W}}^{(k)},\frac{K_n}{\lambda_k(t)}\Big)$$
$$+ \varepsilon \sum_{i=1}^n \bigg(\frac{x_i^2}{2} - x_i S_i - \frac{x_i \hat{W}_i}{\sqrt{\varepsilon}}\bigg). \tag{23}$$

Here, $\mathbf{\Theta} := \{\mathbf{S},\mathbf{W}^{(k)},\tilde{\mathbf{W}}^{(k)}\}_{k=1}^{K_n},\hat{\mathbf{W}},\mathbf{A},k\in[K_n],t\in[0,1]$ and

$$h(\mathbf{x},\mathbf{S},\mathbf{W},\mathbf{A},\sigma^2) := \frac{1}{\sigma^2}\sum_{\mu=1}^m \bigg(\frac{[\mathbf{A}\bar{\mathbf{x}}]_\mu^2}{2} - \sigma\big[\mathbf{A}\bar{\mathbf{x}}\big]_\mu W_\mu\bigg), \tag{24}$$

$$h_{\mathrm{mf}}(\mathbf{x},\mathbf{S},\tilde{\mathbf{W}},\sigma^2) := \frac{1}{\sigma^2}\sum_{i=1}^n \bigg(\frac{\bar{x}_i^2}{2} - \sigma\bar{x}_i\tilde{W}_i\bigg), \tag{25}$$

where $\bar{\mathbf{x}} = \mathbf{x} - \mathbf{S}$ and $\bar{x}_i = x_i - S_i$.

The (k,t)-interpolating model corresponds an inference model where one has access to the following sets of noisy observations about the signal $\mathbf{S}$

$$\Big\{\mathbf{Z}^{(k')} = \mathbf{A}\mathbf{S} + \mathbf{W}^{(k')}\sqrt{K_n\Delta_n}\Big\}_{k'=k+1}^{K_n}, \tag{26}$$

$$\Big\{\tilde{\mathbf{Z}}^{(k')} = \mathbf{S} + \tilde{\mathbf{W}}^{(k')}\Sigma_{k'}\sqrt{K_n}\Big\}_{k'=1}^{k-1}, \tag{27}$$

$$\Big\{\mathbf{Z}^{(k)} = \mathbf{A}\mathbf{S} + \mathbf{W}^{(k)}\sqrt{\frac{K_n}{\gamma_k(t)}}\Big\}, \tag{28}$$

$$\Big\{\tilde{\mathbf{Z}}^{(k)} = \mathbf{S} + \tilde{\mathbf{W}}^{(k)}\sqrt{\frac{K_n}{\lambda_k(t)}}\Big\}. \tag{29}$$

The first and third sets of observation correspond to similar inference channel as the original model (1) but with a higher noise variance proportional to K_n. These correspond to the first and third terms in (23). The second and fourth sets instead correspond to decoupled Gaussian denoising models, with associated "mean-field" second and fourth term in (23). The last term in (23) is a perturbed term which corresponds to a Gaussian "side-channel" $\mathbf{Y} = \mathbf{S}\sqrt{\varepsilon} + \hat{\mathbf{Z}}$ whose signal-to-noise ratio ε will tend to zero at the end of proof. The noise variance are proportional to K_n in order to keep the average signal-to-noise ratio not dependent on K_n. A perturbed of the original and final (decoupled) models are obtained by setting $k=1, t=0$ and $k=K_n, t=1$, respectively. The interpolation is performed on both k and t. For each fixed k, at t changes from 0 to 1, the observation in (28) is removed from the original model and added to the decoupled model. An interesting point is that the $(k, t = 1)$ and $(k + 1, t = 0)$-interpolating models are statistically equivalent. This is an adjusted model of the classical interpolation model in (Guerra and Toninelli, 2002), where an interpolating path $k \in [K_n]$ is added. This is called the adaptive interpolation method. See (Barbier and Macris, 2017) for more detailed discussion.

Consider a set of observations $[\mathbf{y}, \tilde{\mathbf{y}}]$ from the following channels

$$\begin{cases} \mathbf{y} & = \mathbf{A}\mathbf{S} + \mathbf{W}\frac{1}{\sqrt{\gamma_k(t)}} \\ \tilde{\mathbf{y}} & = \mathbf{S} + \tilde{\mathbf{W}}\frac{1}{\sqrt{\lambda_k(t)}}, \end{cases} \tag{30}$$

where $\mathbf{W} \sim \mathcal{N}(0, \mathbf{I}_m), \tilde{\mathbf{W}} \sim \mathcal{N}(0, \mathbf{I}_n)$, $t \in [0, 1]$ is the interpolating parameter and the "signal-to-noise functions" $\{\gamma_k(t), \lambda_k(t)\}_{k=1}^{K_n}$ satisfy

$$\gamma_k(0) = \Delta_n^{-1}, \qquad \gamma_k(1) = 0, \tag{31}$$

$$\lambda_k(0) = 0, \qquad \lambda_k(1) = \Sigma_k^{-2}, \tag{32}$$

as well as the following constraint

$$\frac{\delta n^{\alpha-1}}{\gamma_k(t)^{-1} + E_k} + \lambda_k(t) = \frac{\delta n^{\alpha-1}}{\Delta_n + E_k} = \Sigma_k^{-2} \tag{33}$$

and thus

$$\frac{d\lambda_k(t)}{dt} = -\frac{d\gamma_k(t)}{dt}\frac{\delta n^{\alpha-1}}{(1 + \gamma_k(t)E_k)^2}. \tag{34}$$

We also require $\gamma_k(t)$ to be strictly decreasing with t. The (k, t)-interpolating model has an associated posterior distribution, Gibbs expectation $\langle - \rangle_{k,t;\varepsilon}$ and (k, t)-interpolating free energy $f_{k,t;\varepsilon}$:

$$P_{k,t;\varepsilon}(\mathbf{x}|\mathbf{\Theta}) := \frac{\prod_{i=1}^n \tilde{P}_0(x_i)e^{-\mathcal{H}_{k,t;\varepsilon}(\mathbf{x};\boldsymbol{\theta})}}{\int \{\prod_{i=1}^n \tilde{P}_0(x_i)\}e^{-\mathcal{H}_{k,t;\varepsilon}(\mathbf{x};\boldsymbol{\theta})}}, \tag{35}$$

$$\langle V(\mathbf{X})\rangle_{k,t;\varepsilon} := \int d\mathbf{x}V(\mathbf{x})P_{k,t;\varepsilon}(\mathbf{x}|\mathbf{\Theta}), \tag{36}$$

$$f_{k,t;\varepsilon} := -\frac{1}{n}\mathbb{E}_{\mathbf{\Theta}}\left[\log \int \left\{\prod_{i=1}^n dx_i\tilde{P}_0(x_i)\right\}e^{-\mathcal{H}_{k,t;\varepsilon}(\mathbf{x};\mathbf{\Theta})}\right]. \tag{37}$$

Lemma 3 *Let P_0 have finite second moment. Then for initial and final systems*

$$|f_{1,0;\varepsilon} - f_{1,0;0}| \le O\left(\frac{\varepsilon}{2n^{1-\alpha}}\right)\mathbb{E}_{S\sim P_0}[S^2] \tag{38}$$

$$|f_{K_n,1;\varepsilon} - f_{K_n,1;0}| \le O\left(\frac{\varepsilon}{2n^{1-\alpha}}\right)\mathbb{E}_{S\sim P_0}[S^2]. \tag{39}$$

Proof Using the similar arguments as Lemma 1, Section II in (Barbier and Macris, 2017), we have

$$|f_{1,0;\varepsilon} - f_{1,0;0}| \leq \frac{\varepsilon}{2}\mathbb{E}_{S\sim\tilde{P}_0}[S^2] \tag{40}$$

$$= \frac{\varepsilon}{2}\frac{k}{n}\mathbb{E}_{S\sim P_0}[S^2] \tag{41}$$

$$= O\bigg(\frac{\varepsilon}{2n^{1-\alpha}}\bigg)\mathbb{E}_{S\sim P_0}[S^2]. \tag{42}$$

Similarly, we come to the other inequality. ■

Now, by defining

$$\Sigma_{\mathrm{mf}}^{-2}(\{E_k\}_{k=1}^{K_n};\Delta) := \frac{1}{K_n}\sum_{k=1}^{K_n}\Sigma_k^2, \tag{43}$$

from (23), we have

$$\mathcal{H}_{K_n,1;0}(\mathbf{x};\boldsymbol{\Theta}) = \sum_{k=1}^{K_n} h_{\mathrm{mf}}(\mathbf{x},\mathbf{S},\mathbf{A},\tilde{\mathbf{W}}^{(k)},K_n\Sigma_k^{-2}) \tag{44}$$

$$= \sum_{k=1}^{K_n}\frac{1}{K_n\Sigma_k^2}\sum_{i=1}^{n}\bigg(\frac{\bar{x}_i^2}{2} - \sqrt{K_n\Sigma_k^2}\bar{x}_i\tilde{W}_\mu^{(k)}\bigg) \tag{45}$$

$$= \Sigma_{\mathrm{mf}}^{-2}\bigg(\sum_{i=1}^{n}\frac{\bar{x}_i^2}{2} - \Sigma_{\mathrm{mf}}\bar{x}_i\sum_{k=1}^{K_n}\frac{\Sigma_{\mathrm{mf}}}{\sqrt{K_n\Sigma_k^2}}\tilde{W}_\mu^{(k)}\bigg). \tag{46}$$

Since

$$\tilde{W} := \sum_{k=1}^{K_n}\frac{\Sigma_{\mathrm{mf}}}{\sqrt{K_n\Sigma_k^2}}\tilde{W}_\mu^{(k)} \sim \mathcal{N}(0,1), \tag{47}$$

it holds from (46) that

$$\mathcal{H}_{K_n,1;0}(\mathbf{x};\boldsymbol{\Theta}) = \Sigma_{\mathrm{mf}}^{-2}\bigg(\sum_{i=1}^{n}\frac{\bar{x}_i^2}{2} - \Sigma_{\mathrm{mf}}\bar{x}_i\tilde{W}\bigg). \tag{48}$$

Hence, we have

$$f_{K_n,1;0} = -\frac{1}{n}\mathbb{E}\bigg[\sum_{i=1}^{n}\log\int dx_i\tilde{P}_0(x_i)e^{-\Sigma_{\mathrm{mf}}^{-2}\big(\frac{\bar{x}_i^2}{2}-\Sigma_{\mathrm{mf}}\bar{x}_i\tilde{W}\big)}\bigg] \tag{49}$$

$$= \mathbb{E}\bigg[\log\int dx\tilde{P}_0(x)e^{-\Sigma_{\mathrm{mf}}^{-2}\big(\frac{\bar{x}^2}{2}-\Sigma_{\mathrm{mf}}\bar{x}\tilde{W}\big)}\bigg] \tag{50}$$

$$= \frac{1}{n^{1-\alpha}}i_{n,\mathrm{den}}(\Sigma_{\mathrm{mf}}(\{E_k\}_{k=1}^{K_n};\Delta_n), \tag{51}$$

where (51) follows from (6).

Similarly, we can show that

$$f_{1,0;0} = -\frac{1}{n}\mathbb{E}\Big[\log\int\Big\{\prod_{i=1}^{n} dx_i\tilde{P}_0(x_i)e^{-\mathcal{H}(\mathbf{x};\mathbf{A},\mathbf{S},\mathbf{W})}\Big\}\Big] \tag{52}$$

$$= \frac{f_n}{n^{1-\alpha}}. \tag{53}$$

In addition, we can prove (with $\bar{\mathbf{X}} = \mathbf{X} - \mathbf{S}$) that

$$\frac{df_{k,t;\varepsilon}}{dt} = \frac{1}{K_n}\big(\mathcal{A}_{k,t;\varepsilon} + \mathcal{B}_{k,t;\varepsilon}\big), \tag{54}$$

$$\mathcal{A}_{k,t;\varepsilon} := \frac{d\gamma_k(t)}{dt}\frac{1}{2n}\sum_{\mu=1}^{m}\mathbb{E}\bigg[\bigg\langle \big[\mathbf{A}\bar{\mathbf{X}}\big]_\mu^2 - \sqrt{\frac{K_n}{\gamma_k(t)}}\big[\mathbf{A}\bar{\mathbf{X}}\big]_\mu W_\mu^{(k)}\bigg\rangle_{k,t;\varepsilon}\bigg], \tag{55}$$

$$\mathcal{B}_{k,t;\varepsilon} := \frac{d\lambda_k(t)}{dt}\frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}\bigg[\bigg\langle \bar{X}_i^2 - \sqrt{\frac{K_n}{\lambda_k(t)}}\bar{X}_i\tilde{W}_i^{(k)}\bigg\rangle_{k,t;\varepsilon}\bigg], \tag{56}$$

where $\mathbf{E}$ denotes the average w.r.t. $\mathbf{X}$ and all quenched random variables $\mathbf{\Theta}$, and $\langle - \rangle_{k,t;\varepsilon}$ is the Gibbs average with Hamiltonian (23).

Now, since $W_\mu^k \sim \mathcal{N}(0,1)$, by using the Gaussian integral formula $\mathbb{E}[Zf(Z)] = \mathbb{E}[f'(Z)]$, we can show that

$$n^{1-\alpha}\mathcal{A}_{k,t;\varepsilon} = \frac{d\gamma_k(t)}{dt}\frac{1}{2n^\alpha}\sum_{\mu=1}^{m}\mathbb{E}\big[\big\langle\big[\mathbf{A}\bar{\mathbf{X}}\big]_\mu\big\rangle_{k,t;\varepsilon}^2\big] \tag{57}$$

$$= \frac{d\gamma_k(t)}{dt}\frac{\delta}{2}\mathrm{ymmse}_{\mathrm{k,t;\varepsilon}}, \tag{58}$$

where

$$\mathrm{ymmse}_{\mathrm{k,t;\varepsilon}} := \frac{1}{m}\mathbb{E}\big[\big\|\mathbf{A}(\langle\mathbf{X}\rangle_{k,t;\varepsilon} - \mathbf{S})\big\|^2\big] \tag{59}$$

is called "measurement minimum mean-square error".

For $\mathcal{B}_{k,t;\varepsilon}$, we proceed similarly and find

$$n^{1-\alpha}\mathcal{B}_{k,t;\varepsilon} = \frac{d\lambda_k(t)}{dt}\frac{1}{2n^\alpha}\sum_{i=1}^{n}\mathbb{E}[\langle\bar{X}_i^2\rangle_{k,t;\varepsilon}] \tag{60}$$

$$= \frac{d\lambda_k(t)}{dt}\frac{1}{2n^\alpha}\mathbb{E}[\|\langle\mathbf{X}\rangle_{k,t;\varepsilon} - \mathbf{S}\|^2] \tag{61}$$

$$= -\frac{d\gamma_k(t)}{dt}\frac{1}{(1+\gamma_k(t)E_k)^2}\frac{\delta}{2}n^{\alpha-1}\mathrm{mmse}_{\mathrm{k,t;\varepsilon}}, \tag{62}$$

where the normalized minimum mean-square-error (MMSE) defined as

$$\mathrm{mmse}_{\mathrm{k,t;\varepsilon}} := \frac{1}{n^\alpha}\mathbb{E}[\|\langle\mathbf{X}\rangle_{k,t;\varepsilon} - \mathbf{S}\|^2]. \tag{63}$$

Here, (62) follows from (34).

By the construction, we have the following coherency property: The $(k, t = 1)$ and $(k+1, t = 0)$ models are equivalent (the Hamiltonian is invariant under this change) and thus $f_{k,1;\varepsilon} = f_{k+1,0;\varepsilon}$ for any k (Barbier and Macris, 2017). This implies that the (k,t)-interpolating free energy satisfies

$$f_{1,0;\varepsilon} = f_{K_n,1;\varepsilon} + \sum_{k=1}^{K_n} (f_{k,0;\varepsilon} - f_{k,1;\varepsilon}) \tag{64}$$

$$= f_{K_n,1;\varepsilon} - \sum_{k=1}^{K_n} \int_0^1 dt \frac{df_{k,t;\varepsilon}}{dt}. \tag{65}$$

It follows that

$$f_{1,0;\varepsilon} n^{1-\alpha} = n^{1-\alpha} f_{K_n,1;\varepsilon} - n^{1-\alpha} \sum_{k=1}^{K_n} \int_0^1 dt \frac{df_{k,t;\varepsilon}}{dt} \tag{66}$$

$$= n^{1-\alpha} f_{K_n,1;\varepsilon} - \frac{1}{K_n} \sum_{k=1}^{K_n} \int_0^1 dt \Big(n^{1-\alpha} \mathcal{A}_{k,t;\varepsilon} + n^{1-\alpha} \mathcal{B}_{k,t;\varepsilon} \Big). \tag{67}$$

On the other hand, by Lemma 3, we have

$$n^{1-\alpha} |f_{1,0,0} - f_{1,0;\varepsilon}| \leq \frac{\varepsilon}{2} \mathbb{E}_{S \sim P_0} \mathbb{E}[S^2], \tag{68}$$

or

$$|f_n - n^{1-\alpha} f_{1,0;\varepsilon}| \leq \frac{\varepsilon}{2} \mathbb{E}_{S \sim P_0} \mathbb{E}[S^2] \tag{69}$$

where (69) follows from (53).

From (58), (62), and (67), we obtain

$$\int_{a_n}^{b_n} d\varepsilon f_n = n^{1-\alpha} \int_{a_n}^{b_n} d\varepsilon f_{1,0;\varepsilon} \pm \frac{\varepsilon}{2} \mathbb{E}_{S \sim P_0} \mathbb{E}[S^2] \tag{70}$$

$$= \int_{a_n}^{b_n} n^{1-\alpha} d\varepsilon \Big\{ f_{K_n,1;\varepsilon} - f_{K_n,1;0} \Big\} + \int_{a_n}^{b_n} d\varepsilon i_{n,\mathrm{den}} \Big(\Sigma_{\mathrm{mf}} \big(\{E_k\}_{k=1}^{K_n}; \Delta \big) \Big)$$
$$- \frac{\delta}{2} \int_{a_n}^{b_n} d\varepsilon \frac{1}{K_n} \sum_{k=1}^{K_n} \int_0^1 dt \frac{d\gamma_k(t)}{dt} \bigg(\mathrm{ymmse}_{k,t;\varepsilon} - \frac{\mathrm{mmse}_{k,t;\varepsilon} n^{\alpha-1}}{(1+\gamma_k(t) E_k)^2} \bigg) \pm \frac{\varepsilon}{2} \mathbb{E}_{S \sim P_0} \mathbb{E}[S^2], \tag{71}$$

where (70) follows from (69).

The following lemma can be verified to hold for the new settings:

Lemma 4 *For any sequence $K_n \to +\infty$ and $0 < a_n < b_n < 1$ (that tend to zero slowly enough in the application), and trial parameters $\{E_k = E_k^{(n)}(\varepsilon)\}_{k=1}^{K_n}$ which are differentiable, bounded and non-increasing in ε, we have*

$$\int_{a_n}^{b_n} \frac{1}{K_n} \sum_{k=1}^{K_n} \int_0^1 dt \frac{d\gamma_k(t)}{dt} \bigg(\mathrm{ymmse}_{\mathrm{k,t};\varepsilon} - \frac{\mathrm{mmse}_{\mathrm{k,t};\varepsilon}}{n^{1-\alpha} + \gamma_k(t) \mathrm{mmse}_{\mathrm{k,t};\varepsilon}} \bigg)$$
$$= O\bigg(\max \bigg\{ o\Big(\frac{b_n - a_n}{\Delta_n} \Big), a_n^{-2} n^{-\gamma} \bigg\} \bigg) \tag{72}$$

as $n \to \infty$ *for some* $0 < \gamma < 1$.

Proof This lemma is a generalization of (93) in (Barbier and Macris, 2017). The proof of this lemma can be found in Appendix A. ■

In addition, the following fact can be proved (see Appendix B).

Lemma 5 *The following holds:*

$$\left|\mathrm{mmse}_{k,t;\varepsilon} - \mathrm{mmse}_{k,0;\varepsilon}\right| = O\bigg(\frac{n^{\frac{3}{2}(1+\alpha)}}{K_n \Delta_n}\bigg). \tag{73}$$

Based on the proof of Lemma 4, another interesting fact can also be derived.

Lemma 6

$$\mathrm{ymmse}_{1,0;0} = \frac{\mathrm{mmse}_{1,0;0}\mathrm{n}^{\alpha-1}}{1 + \mathrm{mmse}_{1,0;0}\mathrm{n}^{\alpha-1}/\Delta_\mathrm{n}} + \mathrm{o}_\mathrm{n}(1). \tag{74}$$

Proof We can obtain (74) by setting $(k = 1, t = 0, \varepsilon = 0)$ in Eq. (313) of the proof of Lemma 4 with noting that $\gamma_k(0) = \Delta_n^{-1}$. ■

Return to the proof of our main theorem. It is easy to verify the following identity:

$$\psi(E_k; \Delta_n) := \frac{\delta}{2} \int_0^1 dt \frac{d\gamma_k(t)}{dt} \bigg(\frac{E_k}{(1 + \gamma_k(t) E_k)^2} - \frac{E_k}{1 + \gamma_k(t) E_k}\bigg). \tag{75}$$

Let

$$\tilde{f}_{n,\mathrm{RS}}(\{E_k\}_{k=1}^{K_n}; \Delta_n) := i_{n,\mathrm{den}}\bigg(\Sigma_{\mathrm{mf}}\{E_k\}_{k=1}^{K_n}; \Delta_n\bigg) + \frac{1}{K_n} \sum_{k=1}^{K_n} \psi(E_k; \Delta_n). \tag{76}$$

From (71) and Lemma 4, we obtain as $a_n \to 0$ that

$$\begin{aligned}\int_{a_n}^{2a_n} d\varepsilon f_n = &\int_{a_n}^{2a_n} n^{1-\alpha} d\varepsilon \Big\{ f_{K_n,1;\varepsilon} - f_{K_n,1;0} \Big\} + \int_{a_n}^{2a_n} d\varepsilon \tilde{f}_{n,\mathrm{RS}}(\{E_k\}_{k=1}^{K_n}; \Delta_n) \\ &+ \frac{\delta}{2K_n} \int_{a_n}^{2a_n} d\varepsilon \sum_{k=1}^{K_n} \int_0^1 dt \frac{d\gamma_k(t)}{dt} \frac{\gamma_k(t)(E_k - \mathrm{mmse}_{\mathrm{k,t;\varepsilon}} n^{\alpha-1})^2}{(1 + \gamma_k(t) E_k)^2 (1 + \gamma_k(t) \mathrm{mmse}_{\mathrm{k,t;\varepsilon}} n^{\alpha-1})} \\ &+ O\bigg(\max\Big\{ o\Big(\frac{a_n}{\Delta_n}\Big), a_n^{-2} n^{-\gamma} \Big\} \bigg).\end{aligned} \tag{77}$$

Now, since $\gamma_k(t)$ is non-creasing in $t \in [0, 1]$, it holds that $\frac{d\gamma_k(t)}{dt} \leq 0$. Hence, from (77), we obtain

$$\begin{aligned}\int_{a_n}^{2a_n} d\varepsilon f_n \leq &\int_{a_n}^{2a_n} n^{1-\alpha} d\varepsilon \Big\{ f_{K_n,1;\varepsilon} - f_{K_n,1;0} \Big\} \\ &+ \int_{a_n}^{2a_n} d\varepsilon \tilde{f}_{n,\mathrm{RS}}(\{E_k\}_{k=1}^{K_n}; \Delta_n) + O\bigg(\max\Big\{ o\Big(\frac{a_n}{\Delta_n}\Big), a_n^{-2} n^{-\gamma} \Big\} \bigg).\end{aligned} \tag{78}$$

By setting $E_k = \arg\min_{E \in [0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n)$ for all $k \in [K_n]$, from (78), we have

$$f_n \leq O(1)\varepsilon + \min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n) + O\bigg(\max\bigg\{o\bigg(\frac{1}{\Delta_n}\bigg), a_n^{-3}n^{-\gamma}\bigg\}\bigg) \tag{79}$$

for some $\gamma > 0$. By taking $\varepsilon \to 0$, we can achieve the following upper bound:

$$f_n - \min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n) \leq O\bigg(\max\bigg\{o\bigg(\frac{1}{\Delta_n}\bigg), a_n^{-3}n^{-\gamma}\bigg\}\bigg). \tag{80}$$

On the other hand, from Lemma 5 and (77), by choosing $K_n = \Omega(n^b)$ for some sufficient large b such that $|\mathrm{mmse}_{k,t;\varepsilon} - \mathrm{mmse}_{k,0;\varepsilon}|$ decays sufficiently fast, we obtain

$$\begin{aligned}\int_{a_n}^{2a_n} d\varepsilon f_n &= \int_{a_n}^{2a_n} n^{1-\alpha} d\varepsilon \Big\{ f_{K_n,1;\varepsilon} - f_{K_n,1;0}\Big\} + \int_{a_n}^{2a_n} d\varepsilon \tilde{f}_{n,\mathrm{RS}}(\{E_k\}_{k=1}^{K_n};\Delta_n)\\ &\quad + \frac{\delta}{2K_n}\int_{a_n}^{2a_n} d\varepsilon \sum_{k=1}^{K_n}\int_0^1 dt \frac{d\gamma_k(t)}{dt}\frac{\gamma_k(t)(E_k - \mathrm{mmse}_{\mathrm{k},0;\varepsilon}n^{\alpha-1})^2}{(1+\gamma_k(t)E_k)^2(1+\gamma_k(t)\mathrm{mmse}_{\mathrm{k},0;\varepsilon}n^{1-\alpha})}\\ &\quad + O\bigg(\max\bigg\{o\bigg(\frac{a_n}{\Delta_n}\bigg), a_n^{-2}n^{-\gamma}\bigg\}\bigg).\end{aligned} \tag{81}$$

By choosing $E_k = \mathrm{mmse}_{\mathrm{k},0;\varepsilon}n^{\alpha-1}$ for all $k \in [K_n]$, it holds that

$$\begin{aligned} E_k &= \mathrm{mmse}_{1,0;\varepsilon}n^{\alpha-1} &(82)\\ &= \frac{1}{n^\alpha}\mathbb{E}\big[\|\mathbf{S}-\mathbb{E}[\mathbf{S}|\mathbf{Y}]\|^2\big]n^{\alpha-1} &(83)\\ &\leq \frac{1}{n^\alpha}\mathbb{E}[\|\mathbf{S}\|^2]n^{\alpha-1} &(84)\\ &= \frac{n}{n^\alpha}\mathbb{E}_{S\sim\tilde{P}_0}[S^2]n^{\alpha-1} &(85)\\ &= \frac{n}{n^\alpha}\frac{n^\alpha}{n}\mathbb{E}_{S\sim P_0}[S^2]n^{\alpha-1} &(86)\\ &= \mathbb{E}_{S\sim P_0}[S^2]n^{\alpha-1} &(87)\\ &= \nu_n, &(88)\end{aligned}$$

where ν_n is defined in Theorem 1. Here, (84) follows from the fact that MMSE estimation gives the lowest MSE.

Hence, from (81) we have

$$\begin{aligned}\int_{a_n}^{2a_n} d\varepsilon f_n &= \int_{a_n}^{2a_n} n^{1-\alpha} d\varepsilon \Big\{ f_{K_n,1;\varepsilon} - f_{K_n,1;0}\Big\}\\ &\quad + \int_{a_n}^{2a_n} d\varepsilon \tilde{f}_{n,\mathrm{RS}}(\{E_k\}_{k=1}^{K_n};\Delta_n) + O\bigg(\max\bigg\{o\bigg(\frac{a_n}{\Delta_n}\bigg), a_n^{-2}n^{-\gamma}\bigg\}\bigg)\end{aligned} \tag{89}$$

Now, let $\Sigma_k^{-2} := \delta n^{\alpha-1}/(E_k + \Delta_n)$ for all $k \in [K_n]$, then it holds that $\Sigma_k^{-2} \geq \frac{\delta n^{\alpha-1}}{\nu_n+\Delta_n}$ by (88) . For a given Δ_n, set $\psi_{\Delta_n}(\Sigma^{-2}) = \psi(\delta n^{\alpha-1}/\Sigma^{-2} - \Delta_n;\Delta_n)$. Since $\psi_{\Delta_n}(\cdot)$ is a convex function, from (76), we are easy to see that

$$\tilde{f}_{n,\mathrm{RS}}(\{E_k\}_{k=1}^{K_n};\Delta_n) = i_{n,\mathrm{den}}\bigg(\Sigma_{\mathrm{mf}}\Big(\{E_k\}_{k=1}^{K_n};\Delta_n\Big)\bigg) + \frac{1}{K_n}\sum_{k=1}^{K_n}\psi_{\Delta}(\Sigma_k^{-2}) \tag{90}$$

$$\geq i_{n,\mathrm{den}}\bigg(\Sigma_{\mathrm{mf}}\Big(\{E_k\}_{k=1}^{K_n};\Delta_n\Big)\bigg) + \psi_{\Delta_n}\Big(\Sigma_{\mathrm{mf}}^{-2}\{E_k\}_{k=1}^{K_n};\Delta_n\Big) \tag{91}$$

$$\geq \min_{\Sigma\in\Big[0,\sqrt{\frac{\nu_n+\Delta_n}{\delta n^{\alpha-1}}}\Big]}\Big(i_{n,\mathrm{den}}(\Sigma)+\psi_{\Delta_n}(\Sigma^{-2})\Big) \tag{92}$$

$$\geq \min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n) \tag{93}$$

From (89) and (93), we obtain a lower bound

$$f_n \geq \min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n) + O(1)\varepsilon + O\bigg(\max\Big\{o\Big(\frac{1}{\Delta_n}\Big), a_n^{-3}n^{-\gamma}\Big\}\bigg). \tag{94}$$

From (80) and (94), we have

$$f_n - \min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n) = O\bigg(\max\Big\{o\Big(\frac{1}{\Delta_n}\Big), a_n^{-3}n^{-\gamma}\Big\}\bigg), \tag{95}$$

or

$$\frac{I(\mathbf{S};\mathbf{Y}|\mathbf{A})}{n^{\alpha}} - \min_{E\in[0,\nu_n]} f_{n,\mathrm{RS}}(E;\Delta_n) = O\bigg(\max\Big\{o\Big(\frac{1}{\Delta_n}\Big), a_n^{-3}n^{-\gamma}\Big\}\bigg) \tag{96}$$

which leads to (22) by choosing the sequence $a_n \to 0$ and $a_n^{-3}n^{-\gamma} \to 0$.

On the other hand, by Lemma 6, we have the following relation:

$$\mathrm{ymmse}_{1,0;0} = \frac{\mathrm{mmse}_{1,0;0}\mathrm{n}^{\alpha-1}}{1+\mathrm{mmse}_{1,0;0}\mathrm{n}^{\alpha-1}/\Delta_{\mathrm{n}}} + \mathrm{o_n}(1). \tag{97}$$

Now, denote by

$$\tilde{i}_{n,\mathrm{den}}(\Sigma) := I(S;S+\tilde{W}\Sigma), \tag{98}$$

$$\tilde{\Sigma}(E(\Delta_n);\Delta_n)^{-2} = \frac{\delta}{E(\Delta_n)+\Delta_n}, \tag{99}$$

then from (3) and (4), we have

$$i_{n,\mathrm{den}}(\Sigma) = n^{1-\alpha}\tilde{i}_{n,\mathrm{den}}(\Sigma), \tag{100}$$

$$\Sigma(E(\Delta_n);\Delta_n)^{-2} := n^{\alpha-1}\tilde{\Sigma}(E(\Delta_n);\Delta_n)^{-2}. \tag{101}$$

Then, from (5), we have

$$\frac{df_{n,\mathrm{RS}}(\tilde{E}(\Delta_n);\Delta_n)}{d\Delta_n^{-1}} = \frac{d\psi(\tilde{E}(\Delta_n);\Delta_n)}{d\Delta_n^{-1}} + \frac{di_{n,\mathrm{den}}(\Sigma(\tilde{E}(\Delta_n);\Delta_n)}{d\Delta_n^{-1}} \tag{102}$$

$$= \frac{d\psi(\tilde{E}(\Delta_n);\Delta_n)}{d\Delta_n^{-1}} + \left(\frac{di_{n,\mathrm{den}}(\Sigma(\tilde{E}(\Delta_n);\Delta_n)}{\Sigma(\tilde{E}(\Delta_n);\Delta_n)^{-2}}\right)\left(\frac{d\Sigma(\tilde{E}(\Delta_n);\Delta_n)^{-2}}{d\Delta_n^{-1}}\right) \tag{103}$$

$$= \frac{d\psi(\tilde{E}(\Delta_n);\Delta_n)}{d\Delta_n^{-1}} + \left(\frac{d\tilde{i}_{n,\mathrm{den}}(\Sigma(\tilde{E}(\Delta_n);\Delta_n)}{\Sigma(\tilde{E}(\Delta_n);\Delta_n)^{-2}}\right)\left(\frac{d\tilde{\Sigma}(\tilde{E}(\Delta_n);\Delta_n)^{-2}}{d\Delta_n^{-1}}\right) \tag{104}$$

$$= \frac{\delta}{2}\left(\frac{\tilde{E}(\Delta_n)}{1+\tilde{E}(\Delta_n)/\Delta_n}\right), \tag{105}$$

where (105) follows from (Barbier et al., 2016).

Hence, for $n \to \infty$, we have

$$\mathrm{ymmse}_{1,0;0} = \frac{1}{\delta n^{\alpha}}\frac{dI(\mathbf{S};\mathbf{Y}|\mathbf{A})}{d\Delta_n^{-1}} \tag{106}$$

$$= \frac{2}{\delta}\left(\frac{df_{n,\mathrm{RS}}(\tilde{E}(\Delta_n);\Delta_n)}{d\Delta_n^{-1}}\right) \tag{107}$$

$$= \frac{\tilde{E}(\Delta_n)}{1+\tilde{E}(\Delta_n)/\Delta_n}, \tag{108}$$

where (106) follows from (Dongning Guo et al., 2005) and (59) with $m = \delta n^{\alpha}$, (107) follows from (96), and (108) follows from (105).

From (97) and (108), we obtain (8). This concludes our proof of Theorem 1. ■

4. Algorithm and performance guarantee

In this section, we propose a way to modify the Approximate Message Passing (AMP) algorithm (Bayati and Montanari, 2011) to make it work for sub-linear regimes. See our following Algorithm 1. Compared with the AMP in (Bayati and Montanari, 2011), we multiply an extra term $n^{1-\alpha}$ in the steps 1 (initialize) and 5 (state evolution).

From now on, we denote by $\hat{\mathbf{x}}^{(t)}, \mathbf{y}^{(t)}, \mathbf{z}^{(t)}, \mathbf{h}^{(t)}, \tau_t$ the value of $\hat{\mathbf{x}}, \mathbf{y}, \mathbf{z}, \mathbf{h}, \tau$ at the iteration t, respectively.

First, we prove a few important lemmas. We begin with a generalization of the general law of large numbers in (Fazekas and Klesov, 2001, Theorem 2.1). Our generalization may be of independent interest.

Lemma 7 *Let $\{b_n\}_{n=1}^{\infty}$ be a nondecreasing unbounded sequence of positive numbers, and $\{S_n\}_{n=1}^{\infty}$ be a sequence of random variables. Let $\{\nu_n\}_{n=1}^{\infty}$ be nonnegative numbers. Let $r > 0$ and $\rho \geq 0$ be two fixed numbers. Assume that for each $n \geq 1$*

$$\mathbb{E}\left[\max_{1\leq l\leq n}\left|S_l\right|^r\right] \leq d_n^{1/(1+\rho)}\sum_{i=1}^{n}\nu_i \tag{109}$$

Algorithm 1 AMP for sub-linear regimes.

Input: observation $\mathbf{y}$, matrix sizes m, n, other parameters α, δ, number of iterations itermax, $t = 1$, $U_0 \sim \tilde{P}_0, W \sim \mathcal{N}(0,1)$, $\{\eta_t\}_{t=1}^{\text{itermax}}$ are given Lipschitz continuous functions.

repeat

Initialize $\tau = \sqrt{\Delta_n + n^{1-\alpha}\mathbb{E}_{S\sim\tilde{P}_0}[S^2]/\delta}, \mathbf{z} = \underline{0}, \hat{\mathbf{x}} = \underline{0}, d = 0.$

$\mathbf{z} \leftarrow \mathbf{y} - \mathbf{A}\hat{\mathbf{x}} + \frac{1}{\delta}\mathbf{z}d$

$\mathbf{h} \leftarrow \mathbf{A}^*\mathbf{z} + \hat{\mathbf{x}}$

$\hat{\mathbf{x}} \leftarrow \eta_t(\mathbf{h}, \tau), \quad d \leftarrow \text{Mean}\big(\frac{\mathrm{d}\eta_t}{\mathrm{dx}}(\mathbf{h}, \tau)\big)$

$\tau \leftarrow \sqrt{\Delta_n + (n^{1-\alpha}/\delta)\mathbb{E}[(\eta_t(U_0 + \tau W, \tau) - U_0)^2]}$

$t \leftarrow t + 1$

until t=itermax

Output: $\hat{\mathbf{x}}$.

for some positive and non-increasing sequence $\{d_n\}_{n=1}^{\infty}$ with $d_1 = 1$. Under the condition that $d_n b_n \to \infty$ and

$$\sum_{l=1}^{\infty} \frac{\nu_l}{b_l^r d_l^{r-1/(1+\rho)}} < \infty, \tag{110}$$

then

$$\lim_{n\to\infty} \frac{S_n}{d_n b_n} = 0, \qquad a.s.. \tag{111}$$

Remark 8 *For $d_n = 1$ for all n, this lemma recovers the general strong law of large numbers in (Fazekas and Klesov, 2001, Theorem 2.1).*

Proof See Appendix C for a detailed proof. ■

Next, we recall the following lemma from (Bayati and Montanari, 2011).

Lemma 9 *Let $\phi : \mathbb{R}^q \to \mathbb{R}$ be a pseudo-Lipschitz of order k, then the following hold:*

- *There is a constant L' such that for all $\mathbf{x} \in \mathbb{R}^q : |\phi(\mathbf{x})| \leq L'(1 + \|\mathbf{x}\|^k)$.*
- *ϕ is locally Lipschitz, that is for any $Q > 0$, there exists a constant $L_{Q,q} < \infty$ such that for all $\mathbf{x}, \mathbf{y} \in [-Q, Q]^q$,*

$$|\phi(\mathbf{x}) - \phi(\mathbf{y})| \leq L_{Q,q}\|\mathbf{x} - \mathbf{y}\|. \tag{112}$$

Further, $L_{Q,q} \leq c[1 + (Q\sqrt{q})^{k-1}]$ for some constant c.

Now, we prove the following important lemma.

Lemma 10 *Let $\{f_t\}_{t\geq 0}$ and $\{g_t\}_{t\geq 0}$ be two sequences of functions, where for each $t \in \mathbb{Z}^+$, $f_t : \mathbb{R}^2 \to \mathbb{R}$ and $g_t : \mathbb{R}^2 \to \mathbb{R}$ are assumed to be Lipschitz continuous. Given $\tilde{\mathbf{w}} \in \mathbb{R}^m$ and $\mathbf{s}_0 \in \mathbb{R}^n$, define the sequence of vectors $\mathbf{h}^{(t)}, \mathbf{q}^{(t)} \in \mathbb{R}^n$ and $\mathbf{b}^{(t)}, \mathbf{m}^{(t)} \in \mathbb{R}^m$ such that*

$$\mathbf{h}^{(t+1)} = \mathbf{A}^* \mathbf{m}^{(t)} - \zeta_t \mathbf{q}^{(t)}, \qquad \mathbf{m}^{(t)} = g_t(\mathbf{b}^{(t)}, \tilde{\mathbf{w}}) \tag{113}$$

$$\mathbf{b}^{(t)} = \mathbf{A}\mathbf{q}^{(t)} - \lambda_t \mathbf{m}^{(t-1)}, \quad \mathbf{q}^{(t)} = f_t(\mathbf{h}^{(t)}, \mathbf{s}_0), \tag{114}$$

where $\zeta_t = \langle g_t'(\mathbf{b}^t, \tilde{\mathbf{w}})\rangle, \lambda_t = \frac{1}{\delta}\langle f_t'(\mathbf{h}^{(t)}, \mathbf{s}_0)\rangle$ (both derivatives are with respect to the first argument.) Assume that

$$\sigma_0^2 = \frac{n^{1-\alpha}}{\delta}\left(\frac{\mathbb{E}\big[\|\mathbf{q}^{(0)}\|^2\big]}{n}\right) \tag{115}$$

is positive and finite, for a sequence of initial conditions of increasing dimensions. State evolution defines quantities $\{\tau_t^2\}_{t\geq 0}$ and $\{\sigma_t^2\}_{t\geq 0}$ via

$$\tau_t^2 = \mathbb{E}\big[g_t(\sigma_t Z, \tilde{W})^2\big], \tag{116}$$

$$\sigma_t^2 = \frac{n^{1-\alpha}}{\delta}\mathbb{E}\big[f_t\big(\tau_{t-1} Z, U_0\big)^2\big] \tag{117}$$

where $\tilde{W} \sim \mathcal{N}(0, \Delta_n)$ and $U_0 \sim \tilde{P}_0$ which are independent of $Z \sim \mathcal{N}(0,1)$. Then, for any pseudo-Lipschitz function $\psi : \mathbb{R}^2 \to \mathbb{R}_+$ of order k and $t \geq 0$, it holds almost surely that

$$\lim_{n\to\infty} \frac{1}{n^\alpha}\sum_{i=1}^n \phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) - n^{1-\alpha}\mathbb{E}\bigg[\phi_h(\tau_0 Z_0, \tau_1 Z_1, \cdots, \tau_t Z_t, U_0)\bigg] = 0, \tag{118}$$

$$\lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m \phi_b(b_i^{(1)}, b_i^{(2)}, \cdots, b_i^{(t+1)}, \tilde{w}_i) - \mathbb{E}\bigg[\phi_b(\sigma_0 \hat{Z}_0, \sigma_1 \hat{Z}_1, \cdots, \sigma_t \hat{Z}_t, \tilde{W})\bigg] = 0, \tag{119}$$

where $(Z_0, Z_1, \cdots, Z_t)$ and $(\hat{Z}_0, \hat{Z}_1, \cdots, \hat{Z}_t)$ are two zero-mean Gaussian vectors independent of U_0 and W, with $Z_i, \hat{Z}_i \sim \mathcal{N}(0,1)$.

Remark 11 *Some remarks are in order.*

- *The proof is based on (Bayati and Montanari, 2011, Theorem 1). As in the converse proof, we need to make use of the sparsity of the signal to achieve* (118) *and* (119)*. Two equations* (118) *and* (116) *are the main differences between the proof of Lemma 10 and the proof of (Bayati and Montanari, 2011, Theorem 1). To show this fact, we need to use Lemma 7 instead of (Bayati and Montanari, 2011, Theorem 3).*
- *The states $\{\tau_t\}$ are defined in non-asymptotic sense. This means that we allow them to depend on n. In (Bayati and Montanari, 2011), all states are defined in the asymptotic sense.*
- *Compared with (Bayati and Montanari, 2011, Theorem 3), we constraint the set of pseudo-Lipschitz functions with co-domain $\mathbb{R}_+$ instead of $\mathbb{R}$. This is likely caused by our proof technique.*

Proof The proof of Lemma 10 is based on (Bayati and Montanari, 2011, Proof of Theorem 3) with some important changes to account for the new settings. In the following, we outline the proof and present all these changes.

- *Step 1:* Let $\mathcal{F}_{t_1,t_2}^{(n)} := \sigma(\mathbf{b}_0^{t_1-1}, \mathbf{m}_0^{t_1-1}, \mathbf{h}_1^{t_2}, \mathbf{q}_0^{t_2}, \tilde{\mathbf{w}})$, which is the σ-algebra generated by all random variables in the bracket. Note that this σ-algebra is slightly different from the σ-algebra in (Bayati and Montanari, 2011, Proof of Theorem 2) since it does not covers $\mathbf{s}_0$.

  Let $\mathbf{m}_{\parallel}^{(t)}$ and $\mathbf{q}_{\parallel}^{(t)}$ be orthogonal projections of $\mathbf{m}^{(t)}$ and $\mathbf{q}^{(t)}$ onto $\text{span}(\mathbf{m}^{(0)}, \cdots, \mathbf{m}^{(\text{t}-1)})$ and $\text{span}(\mathbf{q}^{(0)}, \cdots, \mathbf{q}^{(\text{t}-1)})$, respectively. Then, we can express

  $$\mathbf{m}_{\parallel}^{(t)} = \sum_{i=1}^{t-1} \alpha_i \mathbf{m}^{(i)}, \qquad \mathbf{q}_{\parallel}^{(t)} = \sum_{i=1}^{t-1} \beta_i \mathbf{q}^{(i)} \tag{120}$$

  for some tuples $(\alpha_1, \alpha_2, \cdots, \alpha_{t-1})$ and $(\beta_1, \beta_2, \cdots, \beta_{t-1})$, respectively. Define $\mathbf{m}_{\perp}^{(t)} = \mathbf{m}^{(t)} - \mathbf{m}_{\parallel}^{(t)}$ and $\mathbf{q}_{\perp}^{(t)} = \mathbf{q}^{(t)} - \mathbf{q}_{\parallel}^{(t)}$. Then, it can be shown that (Bayati and Montanari, 2011):

  $$\mathbf{h}^{(t+1)}\big|\mathcal{F}_{t+1,t}^{(n)} \cup \sigma(\mathbf{s}_0) \overset{(d)}{=} \sum_{i=0}^{t-1} \alpha_i \mathbf{h}^{(i+1)} + \tilde{\mathbf{A}}^* \mathbf{m}_{\perp}^{(t)} + \tilde{\mathbf{Q}}_{t+1} \vec{o}_{t+1}(1) \tag{121}$$

  $$\mathbf{b}^{(t)}\big|\mathcal{F}_{t,t}^{(n)} \cup \sigma(\mathbf{s}_0) \overset{(d)}{=} \sum_{i=0}^{t-1} \beta_i \mathbf{b}^{(i)} + \tilde{\mathbf{A}} \mathbf{q}_{\perp}^{(t)} + \tilde{\mathbf{M}}_t \vec{o}_t(1), \tag{122}$$

  where $\tilde{\mathbf{A}}$ is an independent copy of $\mathbf{A}$, and the matrices $\tilde{\mathbf{Q}}_t$ and $\tilde{\mathbf{M}}_t$ are such their columns form orthogonal bases for $\text{span}(\mathbf{m}^{(0)}, \cdots, \mathbf{m}^{(t-1)})$ and $\text{span}(\mathbf{q}^{(0)}, \cdots, \mathbf{q}^{(t-1)})$, respectively.

  In addition, let

  $$\mathbf{M}_t := \big[\mathbf{m}^{(0)} | \mathbf{m}^{(1)} | \cdots | \mathbf{m}^{(t-1)}\big], \tag{123}$$

  $$\mathbf{Q}_t := \big[\mathbf{q}^{(0)} | \mathbf{q}^{(1)} | \cdots | \mathbf{q}^{(t-1)}\big], \tag{124}$$

  $$\mathbf{X}_t := \mathbf{A}^* \mathbf{M}_t \tag{125}$$

  $$\mathbf{Y}_t := \mathbf{A} \mathbf{Q}_t. \tag{126}$$

  Here, we denote by $[\mathbf{a}_1 | \mathbf{a}_2 | \cdots \mathbf{a}_k]$ the matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$. Then, from Lemma (Bayati and Montanari, 2011, Lemma 10), it can show that

  $$\mathbf{h}^{(t+1)}\big|\mathcal{F}_{t+1,t}^{(n)} \cup \sigma(\mathbf{s}_0) \overset{(d)}{=} \mathbf{H}_t (\mathbf{M}_t^* \mathbf{M}_t)^{-1} \mathbf{M}_t^* \mathbf{m}_{\parallel}^{(t)} + \mathbf{P}_{\mathbf{Q}_{t+1}}^{\perp} \tilde{\mathbf{A}}^* \mathbf{P}_{\mathbf{M}_t}^{\perp} \mathbf{m}^{(t)} + \mathbf{Q}_t \vec{o}_t(1), \tag{127}$$

  $$\mathbf{b}^{(t)}\big|\mathcal{F}_{t,t}^{(n)} \cup \sigma(\mathbf{s}_0) \overset{(d)}{=} \mathbf{B}_t (\mathbf{Q}_t^* \mathbf{Q}_t)^{-1} \mathbf{Q}_t^* \mathbf{q}_{\parallel}^{(t)} + \mathbf{P}_{\mathbf{M}_t}^{\perp} \tilde{\mathbf{A}} \mathbf{P}_{\mathbf{Q}_t}^{\perp} \mathbf{q}^{(t)} + \mathbf{M}_t \vec{o}_t(1), \tag{128}$$

  where

  $$\mathbf{B}_t := \big[\mathbf{b}^{(0)} | \mathbf{b}^{(1)} | \cdots | \mathbf{b}^{(t-1)}\big], \tag{129}$$

  $$\mathbf{H}_t := \big[\mathbf{h}^{(0)} | \mathbf{h}^{(1)} | \cdots | \mathbf{h}^{(t-1)}\big], \tag{130}$$

  $$\mathbf{P}_{\mathbf{Q}_t}^{\perp} := \mathbf{I} - \mathbf{P}_{\mathbf{Q}_t}, \tag{131}$$

  $$\mathbf{P}_{\mathbf{M}_t}^{\perp} := \mathbf{I} - \mathbf{P}_{\mathbf{M}_t}, \tag{132}$$

and $\mathbf{P}_{\mathbf{M}_t}$ and $\mathbf{P}_{\mathbf{Q}_t}$ are orthogonal projectors onto column spaces of $\mathbf{Q}_t$ and $\mathbf{M}_t$, respectively.

- *Step 2:* By using Lemma 7, for all pseudo-Lipschitz $\phi_h : \mathbb{R}^{t+2} \to \mathbb{R}_+$ of order k, it holds almost surely that

$$\lim_{n\to\infty} \frac{1}{n^\alpha}\bigg(\sum_{i=1}^n \phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) - \mathbb{E}\Big[\phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i})\Big]\bigg) = 0. \tag{133}$$

To show (133), we set

$$T_n := \sum_{i=1}^n \phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}). \tag{134}$$

As (Bayati and Montanari, 2011), given $\mathcal{F}_{t+2,t+1}$, the effects of all terms containing $\vec{o}_{t+1}$ in the distribution of $\mathbf{h}^{(t+1)}$ and $\mathbf{b}^{(t+1)}$ in (121), (122) can be neglected. This fact can be easily explained as follows. Since the limits of $\mathbf{h}^{(t+1)}$ and $\mathbf{b}^{(t+1)}$ are Gaussian vectors which have bounded moments, by applying Lemma 9 and the dominated convergence theorem (Billingsley, 1995), the orders of limits and expectations are interchangeable. Hence, we only need to work with $\phi_h(\lim_{n\to\infty} \mathbf{h}^{(t+1)})$ and $\phi_b(\lim_{n\to\infty} \mathbf{b}^{(t+1)})$ instead of $\phi_h(\mathbf{h}^{(t+1)})$ or $\phi_b(\mathbf{h}^{(t+1)})$ inside all the expectations of these random variables.

In all the proofs in this paper, as Bayati and Montanari (2011), we also define $\mathbb{E}[f(\mathcal{F}, X)|\mathcal{F}]$ as the expectation of the random function $f(\mathcal{F}, X)$ given that all the random in the sigma-algebra $\mathcal{F}$ are fixed. This also means that $\mathbb{E}[f(\mathcal{F}, X)|\mathcal{F}]$ is a constant, not a random variable as in the standard definition of the conditional expectation in probability (Billingsley, 1995).

Now, let

$$\mathcal{F}_{t+2,t+1} := \mathcal{F}^{\infty}_{t+2,t+1} \tag{135}$$

$$= \bigcup_{n=1}^{\infty} \mathcal{F}^{(n)}_{t+2,t+1}. \tag{136}$$

Then, for any $\nu > 0$ and $\gamma > 0$, we have

$$\mathbb{E}\Big[\max_{1\le l\le n} |T_l - \mathbb{E}[T_l|\mathcal{F}_{t+2,t+1}]|^{2-\nu}\Big|\mathcal{F}_{t+2,t+1}\Big]$$

$$= \int_0^\infty \mathbb{P}\big[\max_{1\le l\le n} |T_l - \mathbb{E}[T_l|\mathcal{F}_{t+2,t+1}]|^{2-\nu} > t\Big|\mathcal{F}_{t+2,t+1}\big]dt \tag{137}$$

$$= \int_0^{n^\gamma} \mathbb{P}\big[\max_{1\le l\le n} |T_l - \mathbb{E}[T_l|\mathcal{F}_{t+2,t+1}]|^{2-\nu} > t\Big|\mathcal{F}_{t+2,t+1}\big]dt$$

$$+ \int_{n^\gamma}^\infty \mathbb{P}\big[\max_{1\le l\le n} |T_l - \mathbb{E}[T_l|\mathcal{F}_{t+2,t+1}]|^{2-\nu} > t\big|\mathcal{F}_{t+2,t+1}\big]dt \tag{138}$$

$$\leq n^\gamma + \int_{n^\gamma}^\infty \mathbb{P}\big[\max_{1\leq l\leq n} |T_l - \mathbb{E}[T_l|\mathcal{F}_{t+2,t+1}]|^2 > t^{\frac{2}{2-\nu}} \Big| \mathcal{F}_{t+2,t+1}\big] dt \tag{139}$$

$$= n^\gamma + \int_{n^\gamma}^\infty \mathbb{P}\big[\max_{1\leq l\leq n} |T_l - \mathbb{E}[T_l|\mathcal{F}_{t+2,t+1}]| > t^{\frac{1}{2-\nu}} \Big| \mathcal{F}_{t+2,t+1}\big] dt \tag{140}$$

$$\leq n^\gamma + \int_{n^\gamma}^\infty \frac{\mathsf{Var}(T_n|\mathcal{F}_{t+2,t+1})}{t^{\frac{2}{2-\nu}}} dt \tag{141}$$

$$= n^\gamma + \mathbb{E}(T_n|\mathcal{F}_{t+2,t+1}) \int_{n^\gamma}^\infty \frac{1}{t^{\frac{2}{2-\nu}}} dt \tag{142}$$

$$= n^\gamma + o(\mathsf{Var}(T_n|\mathcal{F}_{t+2,t+1})), \tag{143}$$

where (141) follows from Kolmogorov's maximal inequality (Billingsley, 1995) since given $\mathcal{F}_{t+2,t+1}$, $T_n - \mathbb{E}[T_n|\mathcal{F}_{t+2,t+1}]$ is a sum of independent random variables with zero means for each $n \geq 1$ by the i.i.d. generation of the sequence $\mathbf{s}_0 = (s_{0,1}, s_{0,2}, \cdots)$.

It follows that

$$\mathsf{Var}(T_n|\mathcal{F}_{t+2,t+1}) = \mathbb{E}\bigg[\big(T_n - \mathbb{E}[T_n|\mathcal{F}_{t+2,t+1}]\big)^2 \bigg| \mathcal{F}_{t+2,t+1}\bigg] \tag{144}$$

$$= \sum_{i=1}^n \mathsf{Var}\left[\phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) \middle| \mathcal{F}_{t+2,t+1}\right] \tag{145}$$

$$= \sum_{i=1}^n \mathbb{E}\bigg[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) \bigg| \mathcal{F}_{t+2,t+1}\bigg] - \bigg(\mathbb{E}\big[\phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) \Big| \mathcal{F}_{t+2,t+1}\big]\bigg)^2, \tag{146}$$

where (145) follows from the fact that given $\mathcal{F}_{t+2,t+1}$, $T_n - \mathbb{E}[T_n|\mathcal{F}_{t+2,t+1}]$ is a sum of independent random variables with zero means for each $n \geq 1$ by the i.i.d. generation of the sequence $\mathbf{s}_0 = (s_{0,1}, s_{0,2}, \cdots)$.

Now, we have

$$\begin{aligned}
&\mathbb{E}\bigg[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) \bigg| \mathcal{F}_{t+2,t+1}\bigg] \\
&= \mathbb{E}\bigg[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, 0) \bigg| \mathcal{F}_{t+2,t+1}, s_{0,i} = 0\bigg] \mathbb{P}\bigg[s_{0,i} = 0 \bigg| \mathcal{F}_{t+2,t+1}\bigg] \\
&\quad + \sum_{b=1}^B \mathbb{E}\bigg[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b) \bigg| \mathcal{F}_{t+2,t+1}, s_{0,i} = a_b\bigg] \mathbb{P}\bigg[s_{0,i} = a_b \bigg| \mathcal{F}_{t+2,t+1}\bigg].
\end{aligned} \tag{147}$$

On the other hand, we also have

$$\begin{aligned}
&\bigg(\mathbb{E}_{s_{0,i}}\big[\phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) \Big| \mathcal{F}_{t+2,t+1}\big]\bigg)^2 \\
&= \bigg(\mathbb{E}\bigg[\phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, 0) \bigg| \mathcal{F}_{t+2,t+1}, s_{0,i} = 0\bigg] \mathbb{P}\bigg[s_{0,i} = 0 \bigg| \mathcal{F}_{t+2,t+1}\bigg]
\end{aligned}$$

$$+\sum_{b=1}^{B}\mathbb{E}\Big[\phi_h(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},a_b)\Big|\mathcal{F}_{t+2,t+1},s_{0,i}=a_b\Big]\mathbb{P}\Big[s_{0,i}=a_b\Big|\mathcal{F}_{t+2,t+1}\Big]\Big)^2 \tag{148}$$

$$\geq\Big(\mathbb{E}\Big[\phi_h(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},0)\Big|\mathcal{F}_{t+2,t+1},s_{0,i}=0\Big]\mathbb{P}\Big[s_{0,i}=0\Big|\mathcal{F}_{t+2,t+1}\Big]\Big)^2$$
$$+\sum_{b=1}^{B}\Big(\mathbb{E}\Big[\phi_h(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},a_b)\Big|\mathcal{F}_{t+2,t+1},s_{0,i}=a_b\Big]\mathbb{P}\Big[s_{0,i}=a_b\Big|\mathcal{F}_{t+2,t+1}\Big]\Big)^2 \tag{149}$$

$$=\Big(\mathbb{E}\Big[\phi_h^2(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},0)\Big|\mathcal{F}_{t+2,t+1},s_{0,i}=0\Big]\Big)\Big(\mathbb{P}\Big[s_{0,i}=0\Big|\mathcal{F}_{t+2,t+1}\Big]\Big)^2$$
$$+\sum_{b=1}^{B}\Big(\mathbb{E}\Big[\phi_h^2(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},s_{i,0})\Big|\mathcal{F}_{t+2,t+1},s_{0,i}=a_b\Big]\Big)\Big(\mathbb{P}\Big[s_{0,i}=a_b\Big|\mathcal{F}_{t+2,t+1}\Big]\Big)^2 \tag{150}$$

where (149) follows from $\phi_h:\mathbb{R}^{t+2}\to\mathbb{R}_+$ and $(x_1+x_2+\cdots+x_B)^2\geq\sum_{i=1}^B x_i^2$ if $x_i\geq 0$ for all $i\in[B]$, and (150) follows from the fact that given $\mathcal{F}_{t+2,t+1}$ and $s_{i,0}$, $\phi_h(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},s_{i,0})$ are constants.

From (146), (147), and (150), we obtain

$$\mathbb{E}\Big[\big(T_n-\mathbb{E}[T_n|\mathcal{F}_{t+2,t+1}]\big)^2\Big|\mathcal{F}_{t+2,t+1}\Big]$$
$$=\sum_{i=1}^{n}\Big(\mathbb{E}\Big[\phi_h^2(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},0)\Big|\mathcal{F}_{t+2,t+1},s_{0,i}=0\Big]\mathbb{P}\Big[s_{0,i}=0\Big|\mathcal{F}_{t+2,t+1}\Big]\Big)$$
$$\times\mathbb{P}\Big[s_{0,i}\neq 0\Big|\mathcal{F}_{t+2,t+1}\Big]$$
$$+\sum_{i=1}^{n}\sum_{b=1}^{B}\Big(\mathbb{E}\Big[\phi_h^2(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},a_b)\Big|\mathcal{F}_{t+2,t+1},s_{0,i}=a_b\Big]\mathbb{P}\Big[s_{0,i}\neq a_b\Big|\mathcal{F}_{t+2,t+1}\Big]\Big)$$
$$\times\mathbb{P}\Big[s_{0,i}=a_b\Big|\mathcal{F}_{t+2,t+1}\Big] \tag{151}$$
$$=\sum_{i=1}^{n}\Big(\mathbb{E}\Big[\phi_h^2(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},0)\Big|\mathcal{F}_{t+2,t+1},s_{0,i}=0\Big]\mathbb{P}\Big[s_{0,i}=0\Big|\mathcal{F}_{t+2,t+1}\Big]\Big)$$
$$\times\mathbb{P}\Big[s_{0,i}\neq 0\Big|\mathcal{F}_{t+2,t+1}\Big]$$
$$+\sum_{i=1}^{n}\sum_{b=1}^{B}\Big(\mathbb{E}\Big[\phi_h^2(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},a_b)\Big|\mathcal{F}_{t+2,t+1}\Big]\mathbb{P}\Big[s_{0,i}\neq a_b\Big|\mathcal{F}_{t+2,t+1},s_{0,i}\neq a_b\Big]\Big)$$
$$\times\mathbb{P}\Big[s_{0,i}=a_b\Big|\mathcal{F}_{t+2,t+1}\Big] \tag{152}$$

$$\leq \sum_{i=1}^n \mathbb{E}\left[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, 0)\middle| s_{0,i} = 0\right]\mathbb{P}\left[s_{0,i} \neq 0 \middle| \mathcal{F}_{t+2,t+1}\right]$$
$$+ \sum_{i=1}^n \sum_{b=1}^B \mathbb{E}\left[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i})\middle| s_{0,i} \neq a_b\right]\mathbb{P}\left[s_{0,i} = a_b \middle| \mathcal{F}_{t+2,t+1}\right] \quad (153)$$
$$\leq \sum_{i=1}^n \mathbb{E}\left[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, 0)\middle| \mathcal{F}_{t+2,t+1}\right]\mathbb{P}\left[s_{0,i} \neq 0 \middle| \mathcal{F}_{t+2,t+1}\right]$$
$$+ \sum_{i=1}^n \sum_{b=1}^B \mathbb{E}\left[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\middle| \mathcal{F}_{t+2,t+1}\right]\mathbb{P}\left[s_{0,i} = a_b \middle| \mathcal{F}_{t+2,t+1}\right], \quad (154)$$

where (152) follows from the fact that given $\mathcal{F}_{t+2,t+1}$, $\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)$ does not depends $s_{0,i} = a_b$ or $s_{0,i} \neq a_b$, and (153) follows from $\mathbb{E}[Y|B]\mathbb{P}(B) \leq \mathbb{E}[Y|B]\mathbb{P}(B) + \mathbb{E}[Y|B^c]\mathbb{P}(B^c) = \mathbb{E}[Y]$ if $Y \geq 0$ a.s.

From (153) and (154), we obtain

$$\mathbb{E}\left[\mathbb{E}\left[\left(T_n - \mathbb{E}[T_n|\mathcal{F}_{t+2,t+1}]\right)^2 \middle| \mathcal{F}_{t+2,t+1}\right]\right]$$
$$\leq \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, 0)\middle| \mathcal{F}_{t+2,t+1}\right]\mathbb{P}\left[s_{0,i} \neq 0 \middle| \mathcal{F}_{t+2,t+1}\right]\right]$$
$$+ \sum_{i=1}^n \sum_{b=1}^B \mathbb{E}\left[\mathbb{E}\left[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\middle| \mathcal{F}_{t+2,t+1}\right]\mathbb{P}\left[s_{0,i} \neq 0 \middle| \mathcal{F}_{t+2,t+1}\right]\right]. \quad (155)$$

Now, denote by $a_0 := 0$. Then, by (Bayati and Montanari, 2011), given $b \in [B] \cup \{0\}$ and $p > 1$, it holds that

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\phi_h^{2p}(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\right] < E_{p,b,t}, \quad (156)$$

for some $E_{p,b,t} < \infty$. Since $B < \infty$, it holds that

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\phi_h^{2p}(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\right] < E_{p,t}, \quad (157)$$

where $E_{p,t} := \max_{b \in [B]} E_{p,b,t}$.

Hence, for any $p > 1$ and $q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\mathbb{E}\left[\mathbb{E}\left[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\middle| \mathcal{F}_{t+2,t+1}\right]\mathbb{P}\left[s_{0,i} \neq 0 \middle| \mathcal{F}_{t+2,t+1}\right]\right]$$
$$\leq \left(\mathbb{E}\left[\left(\mathbb{E}\left[\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\middle| \mathcal{F}_{t+2,t+1}\right]\right)^p\right]\right)^{\frac{1}{p}}$$

$$\times \left(\mathbb{E}\left[\left(\mathbb{P}\left[s_{0,i} \neq 0 \middle| \mathcal{F}_{t+2,t+1}\right]\right)^q\right]\right)^{\frac{1}{q}} \tag{158}$$

$$= \left(\mathbb{E}\left[\mathbb{E}\left[\phi_h^{2p}(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b) \middle| \mathcal{F}_{t+2,t+1}\right]\right]\right)^{\frac{1}{p}}$$

$$\times \left(\mathbb{E}\left[\left(\mathbb{P}\left[s_{0,i} \neq 0 \middle| \mathcal{F}_{t+2,t+1}\right]\right)^q\right]\right)^{\frac{1}{q}} \tag{159}$$

$$\leq \left(\mathbb{E}\left[\phi_h^{2p}(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\right]\right)^{\frac{1}{p}} \left(\mathbb{E}\left[\mathbb{P}\left[s_{0,i} \neq 0 \middle| \mathcal{F}_{t+2,t+1}\right]\right]\right)^{\frac{1}{q}} \tag{160}$$

$$= \left(\mathbb{E}\left[\phi_h^{2p}(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\right]\right)^{\frac{1}{p}} \left(\mathbb{P}\left[s_{0,i} \neq 0\right]\right)^{\frac{1}{q}} \tag{161}$$

$$= \left(\frac{k}{n}\right)^{\frac{1}{q}} \left(\mathbb{E}\left[\phi_h^{2p}(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\right]\right)^{\frac{1}{p}} \tag{162}$$

$$\leq Ln^{(\alpha-1)/q}\left(\mathbb{E}\left[\phi_h^{2p}(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)\right]\right)^{\frac{1}{p}} \tag{163}$$

$$\leq Ln^{(\alpha-1)/q} E_{p,t}^{1/p} \tag{164}$$

for some $0 < L < \infty$, where (158) follows from Hölder's inequality (Royden and Fitzpatrick, 2010), (159) follows from given $\mathcal{F}_{t+2,t+1}$, $\phi_h^2(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, a_b)$ is a constant, (160) follows from $q > 1$, (163) follows from $k = O(n^\alpha)$, and (164) follows from (157).

From (143) and (164), for any $\nu > 0$ and $\gamma > 0$, we have

$$\mathbb{E}\left[\mathbb{E}\left[\max_{1 \leq l \leq n} |T_l - \mathbb{E}[T_l|\mathcal{F}_{t+2,t+1}]|^{2-\nu} \middle| \mathcal{F}_{t+2,t+1}\right]\right]$$

$$\leq n^\gamma + (B+1)Lnn^{(\alpha-1)/q} E_{p,t}^{1/p}. \tag{165}$$

Then, by setting $b_l = l, d_l = l^{\alpha-1}, \nu = 1/2, r = 2-\nu, \rho = \frac{1}{2}\mathbf{1}\{\alpha \geq 1/2\} + \frac{\alpha}{2(2-3\alpha)}\mathbf{1}\{\alpha < \frac{1}{2}\}, q = 1+\rho$ and $\nu_l = 1 + L(B+1)E_{p,t}^{1/p}$ for all $l \in \mathbb{Z}^+$, we have

$$\gamma := 1 + \frac{\alpha - 1}{q} \tag{166}$$

$$\geq 1 + \alpha - 1 \tag{167}$$

$$= \alpha > 0. \tag{168}$$

In addition, by these settings, from Lemma 7, we also have

$$\sum_{l=1}^{\infty} \frac{\nu_l}{b_l^r d_l^{r-1/(1+\rho)}} = ((B+1)LE_{p,t}^{1/p} + 1) \sum_{l=1}^{\infty} \frac{1}{l^{2-\nu} l^{(\alpha-1)(2-\nu-1/(1+\rho))}} \tag{169}$$

$$= ((B+1)LE_{p,t}^{1/p} + 1) \sum_{l=1}^{\infty} \frac{1}{l^{(2-\nu)\alpha+(1-\alpha)/(1+\rho)}} \tag{170}$$

$$< \infty, \tag{171}$$

where (171) follows from $(2-\nu)\alpha + (1-\alpha)/(1+\rho) > 1$ with the set value of ρ (note that $0 < \alpha \leq 1$). In addition, $\{b_l = l\}$ is a non-decreasing sequence, $\{d_l = l^{\alpha-1}\}$ is a non-increasing sequence with $d_1 = 1$, and $b_l d_l = l^\alpha \to \infty$ as $l \to \infty$. Hence, by applying Lemma 7, given $\mathcal{F}_{t+2,t+1}$, we have

$$\frac{1}{n^\alpha}\big(T_n - \mathbb{E}[T_n \big| \mathcal{F}_{t+2,t+1}]\big) \to 0, \qquad a.s. \tag{172}$$

Hence, we obtain (133) from (172) and (134).

Similarly, by using Lemma 7, for all pseudo-Lipschitz $\phi_b : \mathbb{R}^{t+2} \to \mathbb{R}_+$ of order k, it holds almost surely that

$$\lim_{m\to\infty} \frac{1}{m}\left(\sum_{i=1}^m \phi_b(b_i^{(1)}, b_i^{(2)}, \cdots, b_i^{(t+1)}, \tilde{w}_i) - \mathbb{E}\left[\phi_b(b_i^{(1)}, b_i^{(2)}, \cdots, b_i^{(t+1)}, \tilde{w}_i)\right]\right) = 0. \tag{173}$$

To show (173), we set

$$\tilde{T}_m := \sum_{i=1}^m \phi_b(b_i^{(1)}, b_i^{(2)}, \cdots, b_i^{(t+1)}, \tilde{w}_i). \tag{174}$$

Then, by (Bayati and Montanari, 2011), it holds that

$$\frac{1}{m}\sum_{i=1}^m \mathbb{E}\left[\phi_b^2(b_i^{(1)}, b_i^{(2)}, \cdots, b_i^{(t+1)}, \tilde{w}_{0,i})\right] < \tilde{E}_{1,t} < \infty. \tag{175}$$

Then, by setting $b_l = l, d_l = 1, r = 2, \rho = 0$, and $\nu_l = \tilde{E}_{1,t}$ for all $l \in \mathbb{Z}^+$ in Lemma 7, it follows that

$$\sum_{l=1}^\infty \frac{\nu_l}{b_l^r d_l^{r-1/(1+\rho)}} = \tilde{E}_{1,t} \sum_{l=1}^\infty \frac{1}{l^2} \tag{176}$$

$$< \infty. \tag{177}$$

Similar to the proof of (133), by applying Lemma 7, given $\mathcal{F}_{t+2,t+1}$, we have

$$\frac{1}{m}\big(\tilde{T}_m - \mathbb{E}[\tilde{T}_m \big| \mathcal{F}_{t+2,t+1}]\big) \to 0, \quad a.s. \tag{178}$$

Hence, we obtain (173) from (174) and (178).

- *Step 3:* From (Bayati and Montanari, 2011, Lemma 2), it holds that $\big[\tilde{\mathbf{A}}^* \mathbf{m}_\perp^{(t)}\big]_i \sim \mathcal{N}(0, \frac{1}{m}\|\mathbf{m}_\perp^{(t)}\|^2)$. Hence, from (121), we have

$$\mathbb{E}\left[\phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) \middle| \mathcal{F}_{t+2,t+1} \cup \sigma(\mathbf{s}_0)\right] \tag{179}$$

$$= \mathbb{E}\left[\phi_h\left(h_i^{(1)}, h_i^{(2)}, \cdots, \sum_{r=0}^{t-1} \alpha_r h_i^{(r+1)} + \frac{\|\mathbf{m}_\perp^{(t)}\| Z}{\sqrt{m}}, s_{0,i}\right) \middle| \mathcal{F}_{t+2,t+1} \cup \sigma(\mathbf{s}_0)\right] \tag{180}$$

where $Z \sim \mathcal{N}(0,1)$. It follows that

$$\begin{aligned}
&\mathbb{E}\Big[\phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i})\Big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[\phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i})\Big|\mathcal{F}_{t+2,t+1} \cup \sigma(\mathbf{s}_0)\Big]\Big] \quad (181)\\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[\phi_h\Big(h_i^{(1)}, h_i^{(2)}, \cdots, \sum_{r=0}^{t-1} \alpha_r h_i^{(r+1)} + \frac{\|\mathbf{m}_\perp^{(t)}\| Z}{\sqrt{m}}, s_{0,i}\Big)\Big|\mathcal{F}_{t+2,t+1} \cup \sigma(\mathbf{s}_0)\Big]\Big] \quad (182)\\
&\quad = \mathbb{E}\Big[\phi_h\Big(h_i^{(1)}, h_i^{(2)}, \cdots, \sum_{r=0}^{t-1} \alpha_r h_i^{(r+1)} + \frac{\|\mathbf{m}_\perp^{(t)}\| Z}{\sqrt{m}}, s_{0,i}\Big)\Big], \quad (183)
\end{aligned}$$

where (181) and (183) follow from the tower property of the conditional expectation (Durrett, 2010).

Hence, from (133) and (183), we obtain

$$\lim_{n\to\infty} \frac{1}{n^\alpha}\Big(\sum_{i=1}^n \phi_h(h_i^{(1)}, h_i^{(2)}, \cdots, h_i^{(t+1)}, s_{0,i}) - \mathbb{E}\Big[\phi_h\Big(h_i^{(1)}, h_i^{(2)}, \cdots, \sum_{r=0}^{t-1} \alpha_r h_i^{(r+1)} + \frac{\|\mathbf{m}_\perp^{(t)}\| Z}{\sqrt{m}}, s_{0,i}\Big)\Big]\Big) = 0. \quad (184)$$

Similarly, from (122) and the tower property of the conditional expectation, we can show that

$$\mathbb{E}\Big[\phi_b(b_i^{(1)}, b_i^{(2)}, \cdots, b_i^{(t+1)}, \tilde{w}_i)\Big] = \mathbb{E}\Big[\phi_b\Big(b_i^{(1)}, b_i^{(2)}, \cdots, \sum_{r=0}^{t-1} \beta_r b_i^{(r)} + \frac{\|\mathbf{q}_\perp^{(t)}\| Z}{\sqrt{m}}, \tilde{w}_i\Big)\Big]. \quad (185)$$

From (173) and (185), we obtain

$$\lim_{m\to\infty} \frac{1}{m}\Big(\sum_{i=1}^m \phi_b(b_i^{(1)}, b_i^{(2)}, \cdots, b_i^{(t+1)}, \tilde{w}_i) - \mathbb{E}\Big[\phi_b\Big(b_i^{(1)}, b_i^{(2)}, \cdots, \sum_{r=0}^{t-1} \beta_r b_i^{(r)} + \frac{\|\mathbf{q}_\perp^{(t)}\| Z}{\sqrt{m}}, \tilde{w}_i\Big)\Big]\Big) = 0. \quad (186)$$

By using induction, from (184) and (186), we have

$$\lim_{n\to\infty}\frac{1}{n^{\alpha}}\bigg(\sum_{i=1}^{n}\phi_h(h_i^{(1)},h_i^{(2)},\cdots,h_i^{(t+1)},s_{0,i}) \\ -\mathbb{E}\Big[\phi_h(\tilde{\tau}_0Z_0,\tilde{\tau}_1Z_1,\cdots,\tilde{\tau}_tZ_t,U_0)\Big]\bigg)=0 \tag{187}$$

$$\lim_{m\to\infty}\frac{1}{m}\bigg(\sum_{i=1}^{m}\phi_b(b_i^{(1)},b_i^{(2)},\cdots,b_i^{(t+1)},\tilde{w}_i) \\ -\mathbb{E}\Big[\phi_b(\tilde{\sigma}_0\hat{Z}_0,\tilde{\sigma}_1\hat{Z}_1,\cdots,\tilde{\sigma}_t\hat{Z}_t,\tilde{W})\Big]\bigg)=0, \tag{188}$$

where

$$\tilde{\tau}_t^2:=\mathbb{E}\bigg[\bigg(\sum_{r=0}^{t-1}\alpha_r\tilde{\tau}_rZ_r+\frac{\|\mathbf{m}_{\perp}^{(t)}\|Z}{\sqrt{m}}\bigg)^2\bigg], \tag{189}$$

$$\tilde{\sigma}_t^2:=\mathbb{E}\bigg[\bigg(\sum_{r=0}^{t-1}\beta_r\tilde{\sigma}_r\hat{Z}_r+\frac{\|\mathbf{q}_{\perp}^{(t)}\|Z}{\sqrt{m}}\bigg)^2\bigg], \tag{190}$$

and $(Z_1,Z_2,\cdots,Z_{t-1},Z_t)$ and $(\hat{Z}_1,\hat{Z}_2,\cdots,\hat{Z}_{t-1},\hat{Z}_t)$ are two Gaussian vectors independent of U_0 and W with $Z_i,\hat{Z}_i\sim\mathcal{N}(0,1)$.

- *Step 4:* Finally, we show that

$$\tilde{\tau}_t^2-\tau_t^2\to 0, \tag{191}$$

$$\tilde{\sigma}_t^2-\sigma_t^2\to 0, \tag{192}$$

where τ_t and σ_t follow the state evolutions in (116) and (117), respectively.

Indeed, by setting

$$\phi_h(\nu^{(1)},\nu^{(2)},\cdots,\nu^{(t+1)},u):=\big(\nu^{(t+1)}\big)^2, \tag{193}$$

$$\phi_b(\nu^{(1)},\nu^{(2)},\cdots,\nu^{(t+1)},u):=\big(\nu^{(t+1)}\big)^2 \tag{194}$$

for all $(\nu^{(1)},\nu^{(2)},\cdots,\nu^{(t+1)},u)\in\mathbb{R}^{t+2}$ for all $(\nu^{(1)},\nu^{(2)},\cdots,\nu^{(t+1)},u)\in\mathbb{R}^{t+2}$, from (187) and (188), we have

$$\lim_{n\to\infty}\frac{1}{n^{\alpha}}\bigg(\big\|\mathbf{h}^{(t+1)}\big\|^2-n\mathbb{E}\bigg[\bigg(\sum_{r=0}^{t-1}\alpha_r\tilde{\tau}_rZ_r+\frac{\|\mathbf{m}_{\perp}^{(t)}\|Z}{\sqrt{m}}\bigg)^2\bigg]\bigg)=0,\quad\forall t\geq 1, \tag{195}$$

$$\lim_{n\to\infty}\frac{1}{m}\bigg(\big\|\mathbf{b}^{(t+1)}\big\|^2-m\mathbb{E}\bigg[\bigg(\sum_{r=0}^{t-1}\alpha_r\tilde{\sigma}_r\hat{Z}_r+\frac{\|\mathbf{q}_{\perp}^{(t)}\|Z}{\sqrt{m}}\bigg)^2\bigg]\bigg)=0,\quad\forall t\geq 1. \tag{196}$$

It follows from (189), (195) and (190), (196) that

$$\lim_{n\to\infty}\frac{1}{n^{\alpha}}\Big(\big\|\mathbf{h}^{(t+1)}\big\|^2-n\tilde{\tau}_t^2\Big)=0, \tag{197}$$

$$\lim_{m\to\infty}\frac{1}{m}\Big(\big\|\mathbf{b}^{(t+1)}\big\|^2-m\tilde{\sigma}_t^2\Big)=0. \tag{198}$$

On the other hand, from (127) and (128), we can prove that (cf. (Bayati and Montanari, 2011, Eq. (3.18) and (3.19)))

$$\lim_{n\to\infty}\left(\frac{\|\mathbf{h}^{(t+1)}\|^2}{n}-\frac{\|\mathbf{m}^{(t)}\|^2}{m}\right)=0, \tag{199}$$

$$\lim_{n\to\infty}\left(\|\mathbf{b}^{(t+1)}\|^2-\lim_{n\to\infty}\|\mathbf{q}^{(t)}\|^2\right)=0, \tag{200}$$

where $\mathbf{m}^{(t)}=g_t(\mathbf{b}^{(t)},\tilde{\mathbf{w}})$ and $\mathbf{q}^{(t)}=f_t(\mathbf{h}^{(t)},\mathbf{s}_0)$ as (113) and (114), respectively. Hence, from (197) – (200), we obtain

$$\lim_{n\to\infty}\frac{\left\|g_t(\mathbf{b}^{(t)},\tilde{\mathbf{w}})\right\|^2}{m}-\tilde{\tau}_t^2=0, \tag{201}$$

$$\lim_{m\to\infty}\frac{1}{m}\left\|f_t(\mathbf{h}^{(t)},\mathbf{s}_0)\right\|^2-\tilde{\sigma}_t^2=0. \tag{202}$$

Furthermore, by setting

$$\phi_b(\nu^{(1)},\nu^{(2)},\cdots,\nu^{(t+1)},u):=g_t\big(\nu^{(t+1)},u\big)^2, \tag{203}$$

$$\phi_h(\nu^{(1)},\nu^{(2)},\cdots,\nu^{(t+1)},u):=f_t\big(\nu^{(t+1)},u\big)^2 \tag{204}$$

for all $(\nu^{(1)},\nu^{(2)},\cdots,\nu^{(t+1)},u)\in\mathbb{R}^{t+2}$, from (188) and (187), we have

$$\lim_{m\to\infty}\frac{1}{m}\left\|g_t(\mathbf{b}^{(t)},\tilde{\mathbf{w}})\right\|^2-\mathbb{E}[g_t(\tilde{\sigma}_t Z,\tilde{W})^2]=0, \tag{205}$$

$$\lim_{n\to\infty}\frac{\left\|f_t(\mathbf{h}^{(t)},\mathbf{s}_0)\right\|^2}{n^\alpha}-n^{1-\alpha}\mathbb{E}[f_t(\tilde{\tau}_t Z,U_0)^2]=0. \tag{206}$$

From (201), (205), and (202), (206), and $m=\delta n^\alpha$, we finally obtain

$$\lim_{n\to\infty}\tilde{\tau}_t^2-\mathbb{E}[g_t(\tilde{\tau}_t Z,\tilde{W})^2]=0, \tag{207}$$

$$\lim_{n\to\infty}\tilde{\sigma}_t^2-\frac{n^{1-\alpha}}{\delta}\mathbb{E}[f_t(\tilde{\tau}_t Z,U_0)^2]=0. \tag{208}$$

Finally, we achieve (191) and (192) by combining (207), (208) and (116), (117).

■

Finally, we prove the following theorem.

Theorem 12 *Let $\hat{\mathbf{x}}^{(t)}=(\hat{x}_1^{(t)},\hat{x}_2^{(t)},\cdots,\hat{x}_n^{(t)})$ be the estimate of $\mathbf{S}$ at the step t in Algorithm 1. Then, for any pseudo-Lipschitz function $\psi:\mathbb{R}^2\to\mathbb{R}_+$ of order k and all $t\geq 0$, the following holds almost surely*

$$\lim_{n\to\infty}\left(\left[\frac{1}{n^\alpha}\sum_{i=1}^n\psi(\hat{x}_i^{(t+1)},S_i)\right]-\delta(\tau_t^2-\Delta_n)\right)=0, \tag{209}$$

where τ_t satisfies the following state evolution:

$$\tau_0^2 = \Delta_n + \frac{n^{1-\alpha}}{\delta}\mathbb{E}_{S\sim\tilde{P}_0}[S^2], \tag{210}$$

$$\tau_{t+1}^2 = \Delta_n + \frac{n^{1-\alpha}}{\delta}\mathbb{E}\big[\psi\big(\eta_t(U_0+\tau_t Z),U_0\big)\big] \quad \forall t\in\mathbb{Z}_+, \tag{211}$$

where U_0 and Z are independent, $U_0\sim\tilde{P}_0$ and $Z\sim\mathcal{N}(0,1)$.

Remark 13 *Some remarks are in order.*

- *At $\alpha=1$, Theorem 12 recovers (Bayati and Montanari, 2011, Theorem 1).*
- *τ_t depends on α though τ_0 since $\tilde{P}_0$ is a function of α.*

Proof First, in Lemma 10, let

$$\tilde{\mathbf{w}} := \mathbf{w}\sqrt{\Delta_n}, \tag{212}$$

$$\mathbf{h}^{(t+1)} := \mathbf{s} - (\mathbf{A}^*\mathbf{z}^{(t)} + \hat{\mathbf{x}}^{(t)}) \tag{213}$$

$$\mathbf{q}^{(t)} := \hat{\mathbf{x}}^{(t)} - \mathbf{s} \tag{214}$$

$$\mathbf{b}^{(t)} := \tilde{\mathbf{w}} - \mathbf{z}^{(t)} \tag{215}$$

$$\mathbf{m}^{(t)} := -\mathbf{z}^{(t)}, \tag{216}$$

$$\mathbf{s}_0 := \mathbf{s}. \tag{217}$$

In addition, the function f_t and g_t are given by

$$f_t(\mathbf{u},\mathbf{s}) := \eta_{t-1}(\mathbf{s}-\mathbf{u}) - \mathbf{s} \tag{218}$$

$$g_t(\mathbf{u},\tilde{\mathbf{w}}) := \mathbf{u} - \tilde{\mathbf{w}} \tag{219}$$

and the initial condition $\mathbf{q}^{(0)} := -\mathbf{s}$. Then, we recover Algorithm 1 as a special case. Hence, by defining: $\phi_h(v^{(1)},v^{(2)},\cdots,v^{(t+1)},s) := \psi(\eta_t(s-v^{(t+1)}),s), \quad \forall(v^{(1)},v^{(2)},\cdots,v^{(t+1)})\in\mathbb{R}^{t+1}$, which is a pseudo-Lipschitz function (Bayati and Montanari, 2011), from Lemma 10, we have

$$\lim_{n\to\infty}\frac{1}{n^\alpha}\sum_{i=1}^n \psi\big(\eta_t(s_i-h_i^{(t+1)}),s_i\big) - n^{1-\alpha}\mathbb{E}\big[\psi(\eta_t(U_0+\tau_t Z_t),U_0)\big] = 0, \qquad a.s. \tag{220}$$

Note that, by Algorithm 1, we have

$$\hat{\mathbf{x}}^{(t+1)} = \eta_t\big(\mathbf{A}^*\mathbf{z}^{(t)} + \hat{\mathbf{x}}^{(t)}\big). \tag{221}$$

Hence, from (213), (220), and (221), we obtain

$$\lim_{n\to\infty}\frac{1}{n^\alpha}\sum_{i=1}^n \psi(\hat{x}_i^{(t+1)},s_i) - n^{1-\alpha}\mathbb{E}\big[\psi(\eta_t(U_0+\tau_t Z_t),U_0)\big] = 0, \tag{222}$$

where $Z_t\sim\mathcal{N}(0,1)$.

Furthermore, by (116), (117), and (219), we have

$$\tau_{t+1}^2 = \Delta_n + \sigma_{t+1}^2 \tag{223}$$
$$= \Delta_n + \frac{n^{1-\alpha}}{\delta}\mathbb{E}\big[\psi\big(\eta_t(U_0+\tau_t Z), U_0\big)\big]\Big) \tag{224}$$
$$= \Delta_n + \frac{n^{1-\alpha}}{\delta}\mathbb{E}\big[\psi\big(\eta_t(U_0+\tau_t Z), U_0\big)\big] \tag{225}$$

with

$$\tau_0^2 = \Delta_n + \sigma_0^2 \tag{226}$$
$$= \Delta_n + \frac{n^{1-\alpha}}{\delta}\bigg(\frac{\mathbb{E}\big[\|\mathbf{q}^{(0)}\|^2\big]}{n}\bigg) \tag{227}$$
$$= \Delta_n + \frac{n^{1-\alpha}}{\delta}\bigg(\frac{\mathbb{E}\big[\|\mathbf{S}\|^2\big]}{n}\bigg) \tag{228}$$
$$= \Delta_n + \frac{n^{1-\alpha}}{\delta}\mathbb{E}_{S\sim\tilde{P}}\mathbb{E}[S^2]. \tag{229}$$

From (220) and (225), we obtain (209). This concludes our proof of Theorem 12. ■

By setting $\psi(x,y) = (x-y)^2$ and using (211), the following corollary is easily derived from Theorem 12.

Corollary 14 *With the same notations as Theorem 12, the following holds:*

$$\lim_{n\to\infty}\bigg(\bigg[\frac{1}{n^\alpha}\sum_{i=1}^n\big(\hat{x}_i^{(t)} - S_i\big)^2\bigg] - \delta\big(\tau_t^2 - \Delta_n\big)\bigg) = 0, \tag{230}$$

where τ_t satisfies the following state evolution:

$$\tau_0^2 = \Delta_n + \frac{n^{1-\alpha}}{\delta}\mathbb{E}_{S\sim\tilde{P}_0}[S^2], \tag{231}$$
$$\tau_{t+1}^2 = \Delta_n + \frac{n^{1-\alpha}}{\delta}\mathbb{E}\big[\big(\eta_t(U_0+\tau_t Z) - U_0\big)^2\big] \quad \forall t \in \mathbb{Z}_+, \tag{232}$$

where $U_0 \sim \tilde{P}_0$ and $Z \sim \mathcal{N}(0,1)$.

5. Numerical Evaluations

In this section, we compare the normalized MMSE fundamental limit in Theorem 1 and the normalized MSE of Algorithm 1 in Corollary 14 for the Bernoulli-Rademacher prior. Here, the normalization means that we divide the total mean square error by $k = n^\alpha$.

More specifically, let $\Delta_n = \Delta = s_{\max} < \infty$ and $\tilde{P}_0(s) = (1-\frac{k}{n})\delta(s) + \frac{1}{2}(\frac{k}{n})(\delta(s-\sqrt{\Delta}) + \delta(s+\sqrt{\Delta}))$, which is the Bernoulli-Rademacher distribution (cf. (Barbier et al., 2020)). With this assumption, we have

$$i_{n,\text{den}}(\Sigma) = n^{1-\alpha} I(S; S + \tilde{W}\Sigma) \tag{233}$$
$$= n^{1-\alpha}\bigg[H(Y) - \frac{1}{2}\log(2\pi e\Sigma^2)\bigg], \tag{234}$$

where

$$f_Y(y) = \left(1 - \frac{k}{n}\right)\frac{1}{\Sigma\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\Sigma^2}\right) + \frac{1}{2\Sigma\sqrt{2\pi}}\left(\frac{k}{n}\right)\left(\exp\left(-\frac{(y-\sqrt{\Delta})^2}{2\Sigma^2}\right) + \exp\left(-\frac{(y+\sqrt{\Delta})^2}{2\Sigma^2}\right)\right). \tag{235}$$

For this prior distribution, we run Algorithm 1 for itermax $= 10$ iterations with the denoiser defined as following:

$$\eta(x,\tau) = \mathbb{E}[S|S+\tau Z = x] \tag{236}$$

$$= \frac{\frac{1}{2}\left(\frac{k}{n}\sqrt{\Delta}\right)\left[\exp(\frac{x\sqrt{\Delta}}{\tau^2}) - \exp(-\frac{x\sqrt{\Delta}}{\tau^2})\right]}{\left(1-\frac{k}{n}\right)\exp\left(\frac{\Delta}{2\tau^2}\right) + \frac{1}{2}\left(\frac{k}{n}\right)\left[\exp(\frac{x\sqrt{\Delta}}{\tau^2}) + \exp(-\frac{x\sqrt{\Delta}}{\tau^2})\right]}. \tag{237}$$

This denoiser has the following derivative:

$$\frac{d\eta(x,\tau)}{dx} = \frac{\frac{\Delta}{2\tau^2}\left(1-\frac{k}{n}\right)\frac{k}{n}\exp\left(\frac{\Delta}{2\tau^2}\right)\left[\exp(\frac{x\sqrt{\Delta}}{\tau^2}) + \exp(-\frac{x\sqrt{\Delta}}{\tau^2})\right]}{\left(\left(1-\frac{k}{n}\right)\exp\left(\frac{\Delta}{2\tau^2}\right) + \frac{1}{2}\left(\frac{k}{n}\right)\left[\exp(\frac{x\sqrt{\Delta}}{\tau^2}) + \exp(-\frac{x\sqrt{\Delta}}{\tau^2})\right]\right)^2} + \frac{\left(\frac{k}{n}\right)^2\frac{\Delta}{\tau^2}}{\left(\left(1-\frac{k}{n}\right)\exp\left(\frac{\Delta}{2\tau^2}\right) + \frac{1}{2}\left(\frac{k}{n}\right)\left[\exp(\frac{x\sqrt{\Delta}}{\tau^2}) + \exp(-\frac{x\sqrt{\Delta}}{\tau^2})\right]\right)^2}. \tag{238}$$

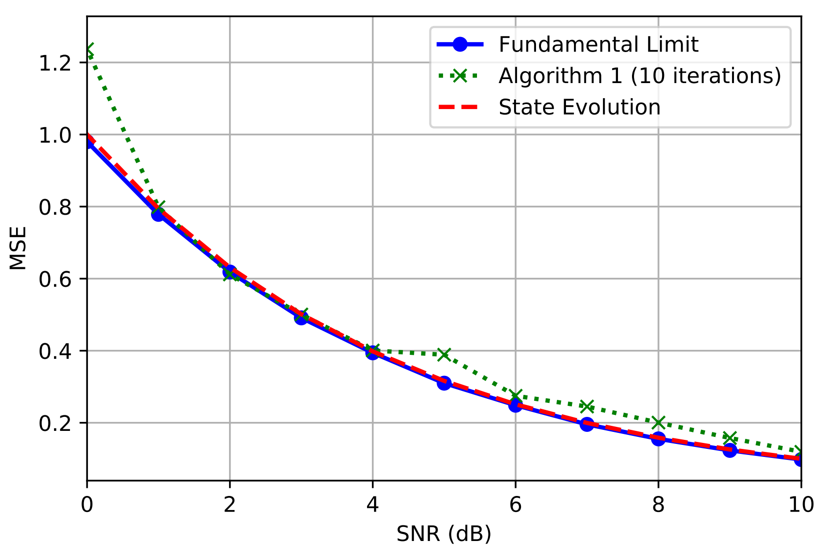

Figure 1: MMSE and the MSE of Algorithm 1 as a function of SNR at $\alpha = 0.5$ and $\delta = 0.5$ for $n = 1000$. Here, $SNR := -10\log(\Delta_n/\delta)$ (dB).

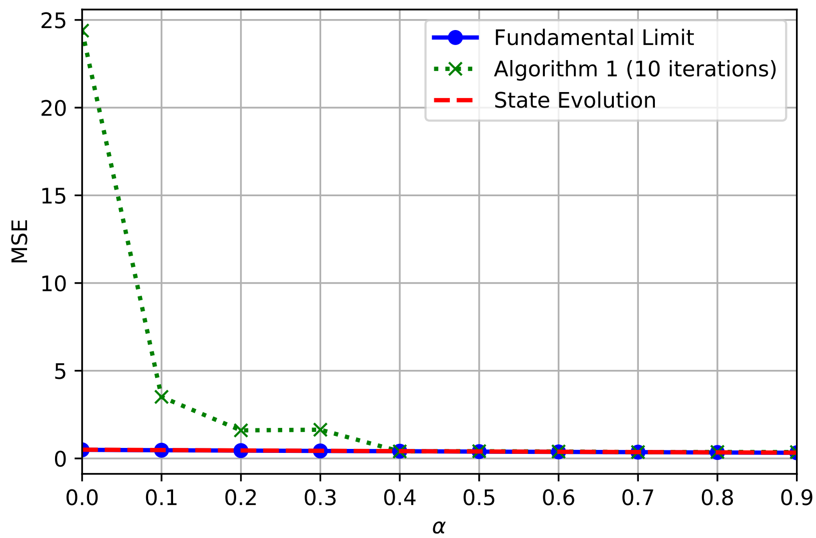

Figure 2: MSE of Algorithm 1, State Evolution, and MSE fundamental limit as functions of α at $\delta = 0.5, SNR = 10\log(2\alpha)$ dB for $n = 1000$.

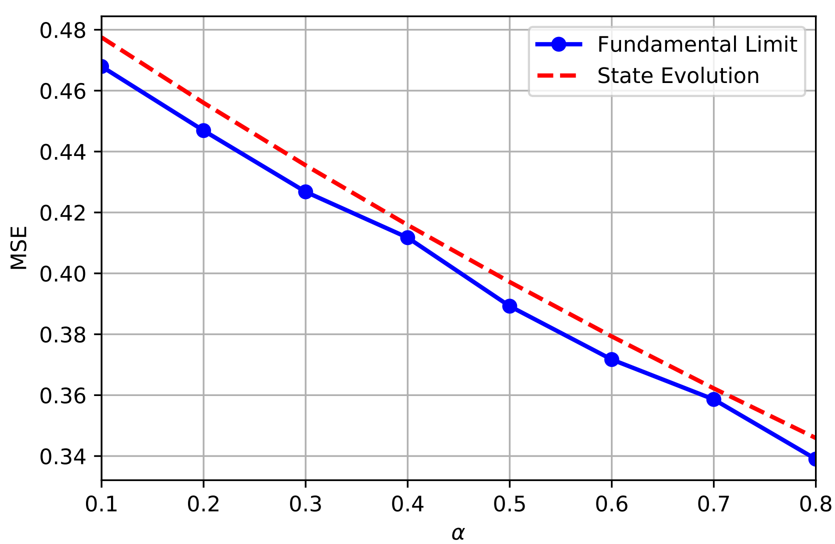

Figure 3: State Evolution of Algorithm 1 vs. Fundamental Limit as functions of α at $\delta = 0.5, SNR = 10\log(2\alpha)$ dB for $n = 1000$.

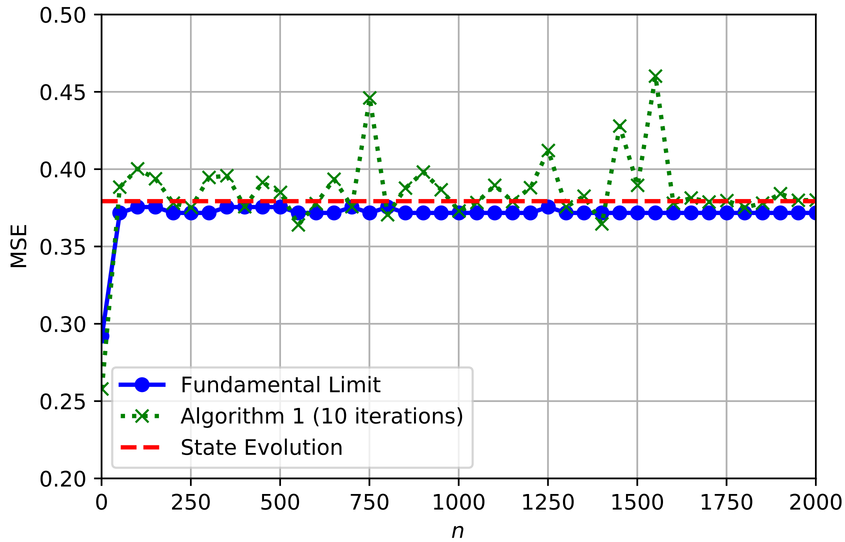

Figure 4: MSE of Algorithm 1, State Evolution, and MSE fundamental limit as functions of n at $\delta = 0.5$ and $\alpha = 0.6$.

More specifically, let $\Delta_n = \Delta$ and $\tilde{P}_0(s) = (1 - \frac{k}{n})\delta(s) + \frac{1}{2}(\frac{k}{n})(\delta(s-1) + \delta(s+1))$, which is the Bernoulli-Rademacher distribution (cf. (Barbier et al., 2020)). With this assumption, we have

$$i_{n,\text{den}}(\Sigma) = n^{1-\alpha} I(S; S + \tilde{W}\Sigma) \tag{239}$$

$$= n^{1-\alpha}\left[H(Y) - \frac{1}{2}\log(2\pi e \Sigma^2)\right], \tag{240}$$

where $Y = S + \tilde{W}\Sigma$ and

$$f_Y(y) = \left(1 - \frac{k}{n}\right)\frac{1}{\Sigma\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\Sigma^2}\right) + \frac{1}{2\Sigma\sqrt{2\pi}}\left(\frac{k}{n}\right)\left[\exp\left(-\frac{(y-1)^2}{2\Sigma^2}\right) + \exp\left(-\frac{(y+1)^2}{2\Sigma^2}\right)\right]. \tag{241}$$

In the first experiment, we set $n = 300$ and run AMP in Algorithm 1 for 10 iterations. Fig. 1 shows that the MSE achieved by Algorithm 1 via Monte-Carlo simulation is very close to the MMSE fundamental limit in Theorem 1. The state evolution in Corollary 14 tracks the MMSE fundamental limit in Theorem 1 very well. This plot also hints us that a judicious modification of the existing AMPs for linear regimes (for example, (Donoho et al., 2009)) can work well for sub-linear regimes.

Fig. 2 plots the MSE as a function of $\alpha \in (0, 1)$ for SNR $= 10\log(2\alpha)$ dB for all $\alpha \in [0, 1]$. As we can observe from the plot, the gap between the state evolution and fundamental limit is very small. However, there is big gap between the state evolution and

MSE from Algorithm 1 at low α's. This can be explained by observing that $m = \delta n^{\alpha}$ is very small at small values of α (for example, $m = 1$ at $n = 1000$ and $\alpha = 0.1$), so the LLNs in Lemma 10 do not hold. To have a better view of the relationship between the MMSE fundamental limit in Theorem 1 and the state evolution of Algorithm 1, we zoom out it in Fig. 3.

In Fig. 4, we plot MSE as a function of n for fixed $\alpha = 0.8$ and $\delta = 0.5$. The figure shows that the gap among fundamental limit in Theorem 1, the state evolution of Algorithm 1 in Corollary 14, and MSE of Algorithm 1 nearly coincide to each other at n sufficiently large.

Acknowledgments

The author is grateful to Prof. Ramji Venkataramanan, the University of Cambridge, for many suggestions to improve the manuscript. The authors also would like to thank the action editor and reviewers for many useful suggestions to improve the manuscript during the review process.

References

Rakesh Agrawal, Heikki Mannila, Ramakrishnan Srikant, Hannu Toivonen, and A. Inkeri Verkamo. *Fast Discovery of Association Rules*, page 307–328. American Association for Artificial Intelligence, USA, 1996. ISBN 0262560976.

J. Barbier, M. Dia, N. Macris, and F. Krzakala. The mutual information in random linear estimation. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 625–632, 2016.

J. Barbier, N. Macris, and C. Rush. All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation. In *Advances of Neural Information Processing Systems (NIPS)*, 2020.

Jean Barbier and N. Macris. 0-1 phase transitions in sparse spiked matrix estimation. *ArXiv*, abs/1911.05030, 2019.

Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference. *Probability Theory and Related Fields*, 174: 1133–1185, 2017.

Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. In *Proceedings of the 31st Conference On Learning Theory*, pages 728–731, 2018.

Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.

M. Bayati and A. Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. on Inform. Th.*, 57(2):764–785, Feb 2011.

P. Billingsley. *Probability and Measure.* Wiley-Interscience, 3rd edition, 1995.

E. Candès and M. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, 2008.

Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse PCA. *2014 IEEE International Symposium on Information Theory*, pages 2197–2201, 2014.

Dongning Guo and S. Verdu. Randomly spread CDMA: asymptotics via statistical physics. *IEEE Trans. on Inform. Th.*, 51(6):1983–2010, 2005.

Dongning Guo, S. Shamai, and S. Verdu. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, 2005.

D. Donoho. Compressed sensing. *IEEE Trans. on Inform. Th.*, 52(1):1289–1306, 2006.

D. Donoho, A. Maleki, and A. Montanari. Message-passing algorithms for compressed sensing. *Proc. Natl. Academy of Sciences (PNAS)*, 106(45):18914–18919, 2009.

R. Durrett. *Probability: Theory and Examples.* Cambridge Univ. Press, 4th edition, 2010.

S. F. Edwards and P. W. Anderson. Theory of spin glasses. *J. Phys. F: Metal Physics*, 5: 965–974, 1975.

M Motavali Emami, Mojtaba Sahraee-Ardakan, Parthe Pandit, Sundeep Rangan, and Alyson K. Fletcher. Generalization error of generalized linear models in high dimensions. In *ICML*, 2020.

I. Fazekas and O. Klesov. A general approach to the strong law of large numbers. *Theory of Probability and Its Applications*, 45:436–449, 2001.

Oliver Y. Feng, Ramji Venkataramanan, Cynthia Rush, and Richard J. Samworth. A unifying tutorial on approximate message passing. *Found. Trends Mach. Learn.*, 15: 335–536, 2022.

Alyson K. Fletcher and Sundeep Rangan. Scalable inference for neuronal connectivity from calcium imaging. In *NIPS*, 2014.

F. Guerra and F. L. Toninelli. The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics*, 230:71–79, 2002.

Brendan Juba. Conditional sparse linear regression, 2016.

Yoshiyuki Kabashima, Florent Krzakala, Marc Mézard, Ayaka Sakata, and Lenka Zdeborová. Phase transitions and sample complexity in bayes-optimal matrix factorization. *IEEE Transactions on Information Theory*, 62:4228–4265, 2016.

F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová. Statistical-physics-based reconstruction in compressed sensing. *Phys. Rev. X*, 2:021005, May 2012.

Clément Luneau, Jean Barbier, and N. Macris. Information theoretic limits of learning a sparse rule. *ArXiv*, abs/2006.11313, 2020.

Ryosuke Matsushita and Toshiyuki Tanaka. Low-rank matrix reconstruction and clustering via approximate message passing. In *NIPS*, 2013.

C. A. Metzler, A. Maleki, and R. G. Baraniuk. From denoising to compressed sensing. *IEEE Trans. on Inform. Th.*, 62(9):5117–514, 2016.

Christopher A. Metzler, Ali Mousavi, and Richard Baraniuk. Learned D-AMP: Principled neural network based compressive image recovery. In *NIPS*, 2017.

Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics*, 2021.

Jonathan Niles-Weed and Ilias Zadik. The all-or-nothing phenomenon in sparse tensor pca. *ArXiv*, abs/2007.11138, 2020.

Parthe Pandit, Mojtaba Sahraee, Sundeep Rangan, and Alyson K. Fletcher. Asymptotics of map inference in deep networks. *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 842–846, 2019.

Parthe Pandit, Mojtaba Sahraee-Ardakan, Sundeep Rangan, Philip Schniter, and Alyson K. Fletcher. Inference with deep generative priors in high dimensions. *IEEE Journal on Selected Areas in Information Theory*, 1:336–347, 2020.

G. Reeves and H. D. Pfister. The replica-symmetric prediction for random linear estimation with gaussian matrices is exact. *IEEE Trans. on Inform. Th.*, 65(4):2252–2283, April 2019.

Galen Reeves and Henry D. Pfister. The replica-symmetric prediction for compressed sensing with gaussian matrices is exact. *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 665–669, 2016.

Galen Reeves, Jiaming Xu, and Ilias Zadik. All-or-nothing phenomena: From single-letter to high dimensions. *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 654–658, 2019a.

Galen Reeves, Jiaming Xu, and Ilias Zadik. The all-or-nothing phenomenon in sparse linear regression. In *Proceedings of the Thirty-Second Conference on Learning Theory*, pages 2652–2663, 2019b.

H. Royden and P. Fitzpatrick. *Real Analysis*. Pearson, 4th edition, 2010.

Cynthia Rush, Adam Greig, and Ramji Venkataramanan. Capacity-achieving sparse superposition codes via approximate message passing decoding. *IEEE Transactions on Information Theory*, 63:1476–1500, 2017.

J. Scarlett and V. Cevher. Limits on support recovery with probabilistic models: An information-theoretic framework. *IEEE Trans. on Inform. Th.*, 63(1):593–620, 2017.

Philip Schniter. A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels. *IEEE Journal of Selected Topics in Signal Processing*, 5:1462–1474, 2011.

T. Tanaka. A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors. *IEEE Trans. on Inform. Th.*, 48(11):2888–2909, 2002.

L. V. Truong. On linear model with Markov signal priors. In *AISTATS*, 2022.

L. V. Truong and J. Scarlett. Support recovery in the phase retrieval model: Information-theoretic fundamental limit. *IEEE Transactions on Information Theory*, 66(12):7887–7910, 2020.

L. V. Truong, M. Aldridge, and J. Scarlett. On the all-or-nothing behavior of Bernoulli group testing. *IEEE Journal on Selected Areas in Information Theory*, 1(3):669–680, 2020.

Appendix A.

In this Appendix, we provide a proof for Lemma 4.

For $\alpha = 1$, it is known from (Barbier and Macris, 2017, Eq. (93)) that

$$\int_{a_n}^{b_n} \frac{1}{K_n} \sum_{k=1}^{K_n} \int_0^1 dt \frac{d\gamma_k(t)}{dt} \left(\text{ymmse}_{\text{k,t;}\varepsilon} - \frac{\text{mmse}_{\text{k,t;}\varepsilon}}{n^{1-\alpha} + \gamma_k(t)\text{mmse}_{\text{k,t;}\varepsilon}} \right) = O(a_n^{-2} n^{-\gamma}) \tag{242}$$

for some $0 < \gamma < 1$.

Now, assume that $0 < \alpha < 1$[3]. Observe that

$$n^{\alpha-1}\text{mmse}_{\text{k,t;}\varepsilon} = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[(S_i - \langle X_i \rangle_{k,t;\varepsilon})^2\big] \tag{243}$$

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\big(\langle S_i - X_i \rangle_{k,t;\varepsilon}\big)^2\big] \tag{244}$$

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\langle \bar{X}_i \rangle_{k,t;\varepsilon}^2\big] \tag{245}$$

since by definition $\bar{X}_i := X_i - S_i$ for all $i \in [n]$.

Now, by (Barbier et al., 2016, Section 6), for all $k \in [K_n]$ we have

$$\text{ymmse}_{\text{k,t;}\varepsilon} = \mathcal{Y}_{1,k} - \mathcal{Y}_{2,k}, \tag{246}$$

where

$$\mathcal{Y}_{1,k} := \mathbb{E}\bigg[\frac{1}{m}\sum_{\mu=1}^m \big(W_\mu^{(k)}\big)^2 \frac{1}{n}\sum_{i=1}^n \langle X_i \bar{X}_i \rangle_{k,t;\varepsilon}\bigg], \tag{247}$$

$$\mathcal{Y}_{2,k} := \sqrt{\gamma_k(t)}\mathbb{E}\bigg[\frac{1}{m}\sum_{\mu=1}^m W_\mu^{(k)} \bigg\langle [\mathbf{A}\bar{\mathbf{X}}]_\mu \frac{1}{n}\sum_{i=1}^n X_i \bar{X}_i \bigg\rangle_{k,t;\varepsilon}\bigg]. \tag{248}$$

By the law of large numbers, $\frac{1}{m}\sum_{\mu=1}^m \big(W_\mu^{(k)}\big)^2 = 1 + o_n(1)$ almost surely, so we have

$$\mathcal{Y}_{1,k} := \frac{1}{n}\mathbb{E}\bigg[\sum_{i=1}^n \langle X_i \bar{X}_i \rangle_{k,t;\varepsilon}\bigg] + o_n(1) \tag{249}$$

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\big(\langle \bar{X}_i(S_i + \bar{X}_i)\rangle_{k,t;\varepsilon}\big)\big] + o_n(1) \tag{250}$$

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\langle \bar{X}_i^2 \rangle_{k,t;\varepsilon}\big] + \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\langle \bar{X}_i \rangle_{k,t;\varepsilon} S_i] + o_n(1) \tag{251}$$

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\langle \bar{X}_i^2 \rangle_{k,t;\varepsilon}\big] + \mathbb{E}\big[\langle \bar{q}_{\mathbf{X},\mathbf{S}} \rangle_{k,t;\varepsilon}\big] + o_n(1), \tag{252}$$

3. Our proof of Lemma 4 for $\alpha < 1$ is simpler than the proof in (Barbier and Macris, 2017) for $\alpha = 1$. More specifically, the proof of concentration inequality in (285) has been simplified by making use of signal sparsity for $\alpha < 1$.

where

$$\bar{q}_{\mathbf{X},\mathbf{S}} := \frac{1}{n}\sum_{i=1}^{n} S_i \bar{X}_i. \tag{253}$$

Here, (250) follows from $X_i = \bar{X}_i + S_i$.

Now, for any a, b, by Cauchy-Schwarz inequality observe that

$$\begin{aligned}
\mathbb{E}[\langle ab\rangle_{k,t;\varepsilon}] &= \mathbb{E}[\langle a\rangle_{k,t;\varepsilon}]\mathbb{E}[\langle b\rangle_{k,t;\varepsilon}] + \mathbb{E}[\langle (a-\mathbb{E}[\langle a\rangle_{k,t;\varepsilon}])b\rangle_{k,t;\varepsilon}] && (254)\\
&= \mathbb{E}[\langle a\rangle_{k,t;\varepsilon}]\mathbb{E}[\langle b\rangle_{k,t;\varepsilon}] + O\left(\mathbb{E}\left[\sqrt{\langle (a-\mathbb{E}[\langle a\rangle_{k,t;\varepsilon}])^2\rangle_{k,t;\varepsilon}\langle b^2\rangle_{k,t;\varepsilon}}\right]\right) && (255)\\
&= \mathbb{E}[\langle a\rangle_{k,t;\varepsilon}]\mathbb{E}[\langle b\rangle_{k,t;\varepsilon}] + O\left(\sqrt{\mathbb{E}\big[\langle (a-\mathbb{E}[\langle a\rangle_{k,t;\varepsilon}])^2\rangle_{k,t;\varepsilon}\big]\mathbb{E}\big[\langle b^2\rangle_{k,t;\varepsilon}\big]}\right). && (256)
\end{aligned}$$

Let $a = \frac{1}{n}\sum_{i=1}^{n} X_i\bar{X}_i$ and $b = W_\mu^{(k)}[\mathbf{A}\bar{\mathbf{X}}]_\mu$, then we have

$$\mathbb{E}\left[W_\mu^{(k)}]\left\langle [\mathbf{A}\bar{\mathbf{X}}]_\mu \frac{1}{n}\sum_{i=1}^{n} X_i\bar{X}_i\right\rangle_{k,t;\varepsilon}\right] = \mathbb{E}[\langle ab\rangle_{k,t;\varepsilon}]. \tag{257}$$

On the other hand, by Cauchy-Schwarz inequality, we also have

$$\begin{aligned}
\mathbb{E}[\langle b^2\rangle_{k,t;\varepsilon}] &= \mathbb{E}\left[\left\langle \left(W_\mu^{(k)}[\mathbf{A}\bar{\mathbf{X}}]_\mu\right)^2\right\rangle_{k,t;\varepsilon}\right] && (258)\\
&\le \mathbb{E}\left[\sqrt{\big\langle (W_\mu^{(k)})^4\big\rangle_{k,t;\varepsilon}\big\langle ([\mathbf{A}\bar{\mathbf{X}}]_\mu)^4\big\rangle_{k,t;\varepsilon}}\right] && (259)\\
&\le \sqrt{\mathbb{E}\big[\big\langle (W_\mu^{(k)})^4\big\rangle_{k,t;\varepsilon}\big\langle ([\mathbf{A}\bar{\mathbf{X}}]_\mu)^4\big\rangle_{k,t;\varepsilon}\big]} && (260)\\
&\le \sqrt[4]{\mathbb{E}\big[\big\langle (W_\mu^{(k)})^4\big\rangle_{k,t;\varepsilon}^2\big]\mathbb{E}\big[\big\langle ([\mathbf{A}\bar{\mathbf{X}}]_\mu)^4\big\rangle_{k,t;\varepsilon}^2\big]} && (261)\\
&\le \sqrt[4]{\mathbb{E}\big[(W_\mu^{(k)})^8\big]\mathbb{E}\big[\big\langle ([\mathbf{A}\bar{\mathbf{X}}]_\mu)^8\big\rangle_{k,t;\varepsilon}\big]} && (262)\\
&= \sqrt[4]{105\mathbb{E}\big[\big\langle ([\mathbf{A}\bar{\mathbf{X}}]_\mu)^8\big\rangle_{k,t;\varepsilon}\big]} && (263)\\
&= \sqrt[4]{105\mathbb{E}\big[\big\langle ([\mathbf{A}\mathbf{X}]_\mu - [\mathbf{A}\mathbf{S}]_\mu)^8\big\rangle_{k,t;\varepsilon}\big]} && (264)\\
&= \sqrt[4]{105\mathbb{E}\left[\left\langle \sum_{i=0}^{8}\binom{8}{i}(-1)^i[\mathbf{A}\mathbf{X}]_\mu^i[\mathbf{A}\mathbf{S}]_\mu^{8-i}\right\rangle_{k,t;\varepsilon}\right]} && (265)\\
&= \sqrt[4]{105\sum_{i=0}^{8}\binom{8}{i}(-1)^i\mathbb{E}\big[\big\langle [\mathbf{A}\mathbf{X}]_\mu^i\big\rangle_{k,t;\varepsilon}[\mathbf{A}\mathbf{S}]_\mu^{8-i}\big]} && (266)\\
&\le \sqrt[4]{105\sum_{i=0}^{8}\binom{8}{i}\sqrt{\mathbb{E}\big[\big\langle [\mathbf{A}\mathbf{X}]_\mu^{2i}\big\rangle_{k,t;\varepsilon}\big]\mathbb{E}\big[[\mathbf{A}\mathbf{S}]_\mu^{2(8-i)}\big]}} && (267)
\end{aligned}$$

$$= \sqrt[4]{105 \sum_{i=0}^{8} \binom{8}{i} \sqrt{\mathbb{E}\big[[\mathbf{AS}]_\mu^{2i}\big]\mathbb{E}\big[[\mathbf{AS}]_\mu^{2(8-i)}\big]}} \tag{268}$$

$$= \sqrt[4]{\frac{105}{n^8} \sum_{i=0}^{8} \binom{8}{i} \sqrt{\mathbb{E}\big[[\sqrt{n}\mathbf{AS}]_\mu^{2i}\big]\mathbb{E}\big[[\sqrt{n}\mathbf{AS}]_\mu^{2(8-i)}\big]}}. \tag{269}$$

Now, we have

$$[\sqrt{n}\mathbf{AS}]_\mu = \sum_{i=1}^{n} \sqrt{n} A_{\mu,i} S_i. \tag{270}$$

Note that

$$\begin{aligned}
\mathsf{Var}(\sqrt{n} A_{\mu,i} S_i) &= n\mathbb{E}[A_{\mu,i}^2]\mathbb{E}_{S_i \sim \tilde{P}_0}[S_i^2] && (271)\\
&= \frac{n}{m}\mathbb{E}_{S_i \sim \tilde{P}_0}[S_i^2] && (272)\\
&= \frac{n}{\delta n^\alpha}\frac{n^\alpha}{n}\mathbb{E}_{S \sim P_0}[S^2] && (273)\\
&= \frac{1}{\delta}\mathbb{E}_{S \sim P_0}[S^2]. && (274)
\end{aligned}$$

Hence, by the central limit theorem, we have

$$[\sqrt{n}\mathbf{AS}]_\mu \to \mathcal{N}\bigg(0, \frac{1}{\delta}\mathbb{E}_{S \sim P_0}[S^2]\bigg). \tag{275}$$

It follows that $\mathbb{E}\big[[\sqrt{n}\mathbf{AS}]_\mu^{2(8-i)}\big]$ and $\mathbb{E}\big[[\sqrt{n}\mathbf{AS}]_\mu^{2i}\big]$ are bounded for each $i \in [8]$. Hence, $\mathbb{E}[\langle b^2 \rangle_{k,t;\varepsilon}]$ goes to zero uniformly in μ as $n \to \infty$ by observing (269).

Furthermore, we have

$$\begin{aligned}
\mathbb{E}\big[\langle (a - \mathbb{E}[\langle a \rangle_{k,t;\varepsilon}])^2 \rangle_{k,t;\varepsilon}\big] &= \mathbb{E}[\langle a^2 \rangle_{k,t;\varepsilon}] - \mathbb{E}\big[(\langle a \rangle_{k,t;\varepsilon})^2\big] && (276)\\
&\le \mathbb{E}[\langle a^2 \rangle_{k,t;\varepsilon}] && (277)\\
&= \mathbb{E}\bigg[\bigg\langle \bigg(\frac{1}{n}\sum_{i=1}^{n} X_i \bar{X}_i\bigg)^2 \bigg\rangle_{k,t;\varepsilon}\bigg] && (278)\\
&\le \mathbb{E}\bigg[\bigg\langle \bigg(\frac{1}{n}\sum_{i=1}^{n} |X_i||\bar{X}_i|\bigg)^2 \bigg\rangle_{k,t;\varepsilon}\bigg] && (279)\\
&\le \mathbb{E}\bigg[\bigg\langle \bigg(\frac{1}{n}\sum_{i=1}^{n} |X_i| 2s_{\max}\bigg)^2 \bigg\rangle_{k,t;\varepsilon}\bigg] && (280)\\
&= 4s_{\max}^2\mathbb{E}\bigg[\bigg(\frac{1}{n}\sum_{i=1}^{n} |X_i|\bigg)^2\bigg] && (281)\\
&\le \frac{4s_{\max}^2}{n}\mathbb{E}\bigg[\sum_{i=1}^{n} X_i^2\bigg] && (282)
\end{aligned}$$

$$= 4s_{\max}^2 \mathbb{E}_{S\sim\tilde{P}_0}[S^2] \tag{283}$$

$$= 4s_{\max}^2 \frac{n^\alpha}{n} \mathbb{E}_{S\sim P_0}[S^2] \tag{284}$$

$$= O_n\left(\frac{1}{n^{1-\alpha}}\right) \to 0 \tag{285}$$

as $0 \le \alpha < 1$, where (280) follows from the fact that $|\bar{X}_i| = |X_i - S_i| \le |X_i| + |S_i| \le 2s_{\max}$.

From (256), (257), (269), and (285), we obtain

$$\mathbb{E}\left[W_\mu^{(k)}\left\langle [\mathbf{A}\bar{\mathbf{X}}]_\mu \frac{1}{n}\sum_{i=1}^n X_i\bar{X}_i\right\rangle_{k,t;\varepsilon}\right] = \mathbb{E}[\langle a\rangle_{k,t;\varepsilon}]\mathbb{E}[\langle b\rangle_{k,t;\varepsilon}] + o_n(1) \tag{286}$$

$$= \mathbb{E}\left[\left\langle \frac{1}{n}\sum_{i=1}^n X_i\bar{X}_i\right\rangle_{k,t;\varepsilon}\right]\mathbb{E}\left[W_\mu^{(k)}\langle[\mathbf{A}\bar{\mathbf{X}}]_\mu\rangle_{k,t;\varepsilon}\right] + o_n(1), \tag{287}$$

where $o_n(1) \to 0$ uniformly in μ.

It follows that

$$\mathcal{Y}_{2,k} = \sqrt{\gamma_k(t)}\frac{1}{m}\sum_{\mu=1}^m \mathbb{E}\left[W_\mu^{(k)}\left\langle [\mathbf{A}\bar{\mathbf{X}}]_\mu \frac{1}{n}\sum_{i=1}^n X_i\bar{X}_i\right\rangle_{k,t;\varepsilon}\right] \tag{288}$$

$$= \sqrt{\gamma_k(t)}\mathbb{E}\left[\left\langle \frac{1}{n}\sum_{i=1}^n X_i\bar{X}_i\right\rangle_{k,t;\varepsilon}\right]\left(\frac{1}{m}\sum_{\mu=1}^m \mathbb{E}\left[W_\mu^{(k)}\langle[\mathbf{A}\bar{\mathbf{X}}]_\mu\rangle_{k,t;\varepsilon}\right]\right) + o_n(1). \tag{289}$$

Now, by (Barbier et al., 2016, Eq. (26)), we have

$$\mathrm{ymmse}_{\mathrm{k,t};\varepsilon} = \frac{1}{m\sqrt{\gamma_k(t)}}\sum_{\mu=1}^m \mathbb{E}\left[W_\mu^{(k)}\langle[\mathbf{A}\bar{\mathbf{X}}]_\mu\rangle_{k,t;\varepsilon}\right]. \tag{290}$$

Hence, from (289) and (290), we obtain

$$\mathcal{Y}_{2,k} = \gamma_k(t)\mathbb{E}\left[\left\langle \frac{1}{n}\sum_{i=1}^n X_i\bar{X}_i\right\rangle_{k,t;\varepsilon}\right]\mathrm{ymmse}_{\mathrm{k,t};\varepsilon} + o_n(1) \tag{291}$$

$$= \gamma_k(t)\mathrm{ymmse}_{\mathrm{k,t};\varepsilon}\tilde{\mathcal{Y}}_{1,k} + o_n(1), \tag{292}$$

where

$$\tilde{\mathcal{Y}}_{1,k} = \mathbb{E}\left[\left\langle \frac{1}{n}\sum_{i=1}^n X_i\bar{X}_i\right\rangle_{k,t;\varepsilon}\right]. \tag{293}$$

It follows from (246)–(292) and (292) that

$$\mathrm{ymmse}_{\mathrm{k,t};\varepsilon} = \mathcal{Y}_{1,k} - \mathcal{Y}_{2,k} \tag{294}$$

$$= \tilde{\mathcal{Y}}_{1,k} + o_n(1) - \tilde{\mathcal{Y}}_{1,k}\gamma_k(t)\mathrm{ymmse}_{\mathrm{k,t};\varepsilon} + o_n(1) \tag{295}$$

$$= \tilde{\mathcal{Y}}_{1,k} - \tilde{\mathcal{Y}}_{1,k}\gamma_k(t)\mathrm{ymmse}_{\mathrm{k,t};\varepsilon} + o_n(1). \tag{296}$$

This leads to

$$\text{ymmse}_{\text{k,t;}\varepsilon} = \frac{\tilde{\mathcal{Y}}_{1,k}}{1+\gamma_k(t)\tilde{\mathcal{Y}}_{1,k}} + o_n(1). \tag{297}$$

Then, it holds that

$$\text{ymmse}_{\text{k,t;}\varepsilon} - \frac{\text{mmse}_{\text{k,t;}\varepsilon} n^{\alpha-1}}{1+\gamma_k(t)\text{mmse}_{\text{k,t;}\varepsilon} n^{\alpha-1}} = \frac{\tilde{\mathcal{Y}}_{1,k}}{1+\gamma_k(t)\tilde{\mathcal{Y}}_{1,k}} - \frac{\text{mmse}_{\text{k,t;}\varepsilon} n^{\alpha-1}}{1+\gamma_k(t)\text{mmse}_{\text{k,t;}\varepsilon} n^{\alpha-1}} + o_n(1). \tag{298}$$

Now, observe that

$$\begin{aligned}
\left|\tilde{\mathcal{Y}}_{1,k} - \text{mmse}_{\text{k,t;}\varepsilon} n^{\alpha-1}\right| &= \left|\frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\langle \bar{X}_i^2\rangle_{k,t;\varepsilon}\big] - \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\langle \bar{X}_i\rangle_{k,t;\varepsilon}^2\big] + \mathbb{E}[\langle \bar{q}_{\mathbf{X},\mathbf{S}}\rangle_{k,t;\varepsilon}]\right| + o_n(1) && (299)\\
&\le \left|\frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\langle \bar{X}_i^2\rangle_{k,t;\varepsilon}\big] - \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\langle \bar{X}_i\rangle_{k,t;\varepsilon}^2\big]\right| + \left|\mathbb{E}[\langle \bar{q}_{\mathbf{X},\mathbf{S}}\rangle_{k,t;\varepsilon}]\right| + o_n(1) && (300)\\
&\le \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\langle \bar{X}_i^2\rangle_{k,t;\varepsilon}\big] + \frac{1}{n}\sum_{i=1}^n \mathbb{E}[|S_i\langle \bar{X}_i\rangle_{k,t;\varepsilon}|] + o_n(1) && (301)\\
&= \sum_{i=1}^n \mathbb{E}\big[\langle (X_i-S_i)^2\rangle_{k,t;\varepsilon}\big] + \frac{1}{n}\sum_{i=1}^n \mathbb{E}[|S_i\langle (X_i-S_i)\rangle_{k,t;\varepsilon}|] + o_n(1) && (302)\\
&\le \frac{2}{n}\sum_{i=1}^n \mathbb{E}\big[\langle X_i^2+S_i^2\rangle_{k,t;\varepsilon}\big] + \frac{1}{n}\sum_{i=1}^n \sqrt{\mathbb{E}[S_i^2]\mathbb{E}[\langle (X_i-S_i)\rangle_{k,t;\varepsilon}^2]} && (303)\\
&\le \frac{2}{n}\sum_{i=1}^n \mathbb{E}\big[\langle X_i^2+S_i^2\rangle_{k,t;\varepsilon}\big] + \frac{1}{n}\sum_{i=1}^n \sqrt{\mathbb{E}[S_i^2]\mathbb{E}[\langle 2(S_i^2+X_i^2)\rangle_{k,t;\varepsilon}]} && (304)\\
&= 6\mathbb{E}_{S\sim\tilde{P}_0}\mathbb{E}[S^2] && (305)\\
&= \frac{6n^\alpha}{n}\mathbb{E}_{S\sim P_0}\mathbb{E}[S^2] && (306)\\
&:= f(n), && (307)
\end{aligned}$$

where $f(n) = O\big(\frac{1}{n^{1-\alpha}}\big) = o(1)$ uniformly in k,t if $0 \le \alpha < 1$. Here, (303) follows from Cauchy–Schwarz inequality, (305) follows from the i.i.d. assumption of the sequence $\{S_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ under $\tilde{P}_0$, and (307) follows from the assumption that $\mathbb{E}_{S\sim P_0}[S^4] < \infty$.

Now, let

$$g_{k,t}(x) := \frac{x}{1+\gamma_k(t)x}. \tag{308}$$

It is easy to see that $g_{k,t}(x)$ is an increasing function for $x \geq 0$. More over, we have

$$0 < g'_{k,t}(x) = \frac{1}{(1+\gamma_k(t)x)^2} \leq 1 \tag{309}$$

uniformly in k, t for all $x \geq 0$ (since $\gamma_k(t) \geq 0$ uniformly in k, t). Hence, we have

$$\begin{aligned} &\left|\mathrm{ymmse}_{\mathrm{k,t;}\varepsilon} - \frac{\mathrm{mmse}_{\mathrm{k,t;}\varepsilon} n^{\alpha-1}}{1+\gamma_k(t)\mathrm{mmse}_{\mathrm{k,t;}\varepsilon} n^{\alpha-1}}\right| \\ &\quad = \left|\frac{\tilde{\mathcal{Y}}_{1,k}}{1+\gamma_k(t)\tilde{\mathcal{Y}}_{1,k}} - \frac{\mathrm{mmse}_{\mathrm{k,t;}\varepsilon} n^{\alpha-1}}{1+\gamma_k(t)\mathrm{mmse}_{\mathrm{k,t;}\varepsilon} n^{\alpha-1}}\right| + o_n(1) &(310) \\ &\quad \leq \left|g_{k,t}(\mathrm{mmse}_{\mathrm{k,t;}\varepsilon} n^{\alpha-1} \pm f(n)) - g_{k,t}(\mathrm{mmse}_{\mathrm{k,t;}\varepsilon} n^{\alpha-1})\right| + o_n(1) &(311) \\ &\quad = |g'_{k,t}(\theta) f(n)| + o_n(1) &(312) \end{aligned}$$

for some $\theta > 0$ by Taylor's expansion. This means that

$$\left|\mathrm{ymmse}_{\mathrm{k,t;}\varepsilon} - \frac{\mathrm{mmse}_{\mathrm{k,t;}\varepsilon} n^{\alpha-1}}{1+\gamma_k(t)\mathrm{mmse}_{\mathrm{k,t;}\varepsilon} n^{\alpha-1}}\right| \leq \tilde{f}(n) \tag{313}$$

uniformly in k, t where $\tilde{f}(n) := f(n) + o_n(1)$.

Then, it holds that

$$\begin{aligned} &\left|\int_{a_n}^{b_n} d\varepsilon \frac{1}{K_n}\sum_{k=1}^{K_n}\int_0^1 dt \frac{d\gamma_k(t)}{dt}\left(\mathrm{ymmse}_{\mathrm{k,t;}\varepsilon} - \frac{\mathrm{mmse}_{\mathrm{k,t;}\varepsilon}}{n^{1-\alpha}+\gamma_k(t)\mathrm{mmse}_{\mathrm{k,t;}\varepsilon}}\right)\right| \\ &\leq \int_{a_n}^{b_n} d\varepsilon \frac{1}{K_n}\sum_{k=1}^{K_n}\left|\int_0^1 dt \frac{d\gamma_k(t)}{dt}\right| \tilde{f}(n) &(314) \\ &= (b_n - a_n)\tilde{f}(n)\frac{1}{K_n}\sum_{k=1}^{K_n}\left|\gamma_k(1) - \gamma_k(0)\right| &(315) \\ &= (b_n - a_n)\tilde{f}(n)\frac{1}{K_n}\sum_{k=1}^{K_n}\frac{1}{\Delta_n} &(316) \\ &= (b_n - a_n)\tilde{f}(n)\frac{1}{\Delta_n} &(317) \\ &= o\left(\frac{b_n - a_n}{\Delta_n}\right) &(318) \end{aligned}$$

as $a_n, b_n \to 0$. Hence, we have

$$\int_{a_n}^{b_n} \frac{1}{K_n}\sum_{k=1}^{K_n}\int_0^1 dt \frac{d\gamma_k(t)}{dt}\left(\mathrm{ymmse}_{\mathrm{k,t;}\varepsilon} - \frac{\mathrm{mmse}_{\mathrm{k,t;}\varepsilon}}{n^{1-\alpha}+\gamma_k(t)\mathrm{mmse}_{\mathrm{k,t;}\varepsilon}}\right) = o\left(\frac{b_n - a_n}{\Delta_n}\right). \tag{319}$$

From (242) and (319), we obtain (72) for all $0 \leq \alpha \leq 1$.

Appendix B.

In this Appendix, we provide a proof for Lemma 5.

Observe that

$$\mathrm{mmse}_{k,t;\varepsilon} - \mathrm{mmse}_{k,0;\varepsilon} = \int_0^t \frac{d\mathrm{mmse}_{\mathrm{k},\nu;\varepsilon}}{d\nu} d\nu. \tag{320}$$

Now, we have

$$n^\alpha \frac{d\mathrm{mmse}_{\mathrm{k},\nu;\varepsilon}}{d\nu} = \frac{d}{d\nu}\mathbb{E}[\|\langle \mathbf{X}\rangle_{k,\nu;\varepsilon} - \mathbf{S}\|^2] \tag{321}$$

$$= \frac{d}{d\nu}\mathbb{E}[\|\langle \mathbf{X}\rangle_{k,\nu;\varepsilon}\|^2] - 2\mathbb{E}\bigg[\mathbf{S}^T \frac{d}{d\nu}\langle \mathbf{X}\rangle_{k,\nu;\varepsilon}\bigg] + \frac{d}{d\nu}\mathbb{E}[\|\mathbf{S}\|^2] \tag{322}$$

$$= 2\frac{d}{d\nu}\mathbb{E}[\|\mathbf{S}\|^2] - 2\mathbb{E}\bigg[\mathbf{S}^T \frac{d}{d\nu}\langle \mathbf{X}\rangle_{k,\nu;\varepsilon}\bigg] \tag{323}$$

$$= -2\mathbb{E}\bigg[\mathbf{S}^T \frac{d}{d\nu}\langle \mathbf{X}\rangle_{k,\nu;\varepsilon}\bigg]. \tag{324}$$

Now, it is easy to see that

$$\frac{d}{d\nu}\langle \mathbf{X}\rangle_{k,\nu;\varepsilon} = \langle \mathbf{X}\rangle_{k,\nu;\varepsilon}\bigg\langle \frac{d\mathcal{H}_{k,\nu;\varepsilon}(\mathbf{X},\mathbf{\Theta})}{d\nu}\bigg\rangle - \bigg\langle \mathbf{X}\frac{d\mathcal{H}_{k,\nu;\varepsilon}(\mathbf{X},\mathbf{\Theta})}{d\nu}\bigg\rangle_{k,\nu;\varepsilon}. \tag{325}$$

Define

$$q_{\mathbf{x},\mathbf{s}} := \frac{1}{n^\alpha}\sum_{i=1}^n x_i s_i, \tag{326}$$

which is a normalized overlap between $\mathbf{x}$ and $\mathbf{s}$.

Let $\mathbf{X}'$ is a replica of $\mathbf{X}$, i.e. $P_{k,\nu;\varepsilon}(\mathbf{x},\mathbf{x}'|\boldsymbol{\theta}) = P_{k,\nu;\varepsilon}(\mathbf{x}|\boldsymbol{\theta})P_{k,\nu;\varepsilon}(\mathbf{x}'|\boldsymbol{\theta})$, then it holds that

$$\mathbb{E}\bigg[\mathbf{S}^T \frac{d}{d\nu}\langle \mathbf{X}\rangle_{k,\nu;\varepsilon}\bigg] = n^\alpha \mathbb{E}\bigg[\langle q_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}\bigg\langle \frac{d\mathcal{H}_{k,\nu;\varepsilon}(\mathbf{X},\mathbf{\Theta})}{d\nu}\bigg\rangle - \bigg\langle q_{\mathbf{X},\mathbf{S}}\frac{d\mathcal{H}_{k,\nu;\varepsilon}(\mathbf{X},\mathbf{\Theta})}{d\nu}\bigg\rangle_{k,\nu;\varepsilon}\bigg] \tag{327}$$

$$= n^\alpha \mathbb{E}\bigg[\bigg\langle q_{\mathbf{X},\mathbf{S}}\bigg(\frac{d\mathcal{H}_{k,\nu;\varepsilon}(\mathbf{X}',\mathbf{\Theta})}{d\nu} - \frac{d\mathcal{H}_{k,\nu;\varepsilon}(\mathbf{X},\mathbf{\Theta})}{d\nu}\bigg)\bigg\rangle_{k,\nu;\varepsilon}\bigg]. \tag{328}$$

Now, observe that

$$\frac{d\mathcal{H}_{k,\nu;\varepsilon}(\mathbf{X},\mathbf{\Theta})}{d\nu} = \frac{d}{d\nu}h\bigg(\mathbf{X},\mathbf{S},\mathbf{A},\mathbf{W}^{(k)},\frac{K_n}{\gamma_k(\nu)}\bigg) + \frac{d}{d\nu}h_{\mathrm{mf}}\bigg(\mathbf{X},\mathbf{S},\tilde{\mathbf{W}}^{(k)},\frac{K_n}{\lambda_k(\nu)}\bigg), \tag{329}$$

where

$$\frac{d}{d\nu}h\bigg(\mathbf{X},\mathbf{S},\mathbf{A},\mathbf{W}^{(k)},\frac{K_n}{\gamma_k(\nu)}\bigg) = \frac{1}{2K_n}\bigg(\frac{d\gamma_k(\nu)}{d\nu}\bigg)\bigg(\sum_{\mu=1}^m [\mathbf{A}\overline{\mathbf{X}}]_\mu^2 - \sqrt{\frac{K_n}{\gamma_k(\nu)}}\sum_{\mu=1}^m [\mathbf{A}\overline{\mathbf{X}}]_\mu W_\mu^{(k)}\bigg), \tag{330}$$

and

$$\frac{d}{d\nu}h_{\mathrm{mf}}\bigg(\mathbf{X},\mathbf{S},\tilde{\mathbf{W}}^{(k)},\frac{K_n}{\lambda_k(\nu)}\bigg)=\frac{1}{2K_n}\bigg(\frac{d\lambda_k(\nu)}{d\nu}\bigg)\bigg(\sum_{i=1}^n\bar{X}_i^2-\sqrt{\frac{K_n}{\lambda_k(\nu)}}\sum_{i=1}^n\bar{X}_i\tilde{W}_i^{(k)}\bigg). \tag{331}$$

Using the fact that $\mathbb{E}[W_\mu^{(k)}]\sim\mathcal{N}(0,1)$ and $\mathbb{E}[\tilde{W}_\mu^{(k)}]\sim\mathcal{N}(0,1)$ and the fact that $\mathbb{E}[Zf(Z)]=\mathbb{E}[f'(Z)]$ for $Z\sim\mathcal{N}(0,1)$, we finally have

$$\mathbb{E}\bigg[\mathbf{S}^T\frac{d}{d\nu}\langle\mathbf{X}\rangle_{k,\nu;\varepsilon}\bigg]=\frac{n^\alpha}{2K_n}\bigg(\frac{d\gamma_k(\nu)}{d\nu}\bigg)\mathbb{E}\big[g(\mathbf{X}',\mathbf{S})-g(\mathbf{X},\mathbf{S})\big], \tag{332}$$

where

$$g(\mathbf{x},\mathbf{s}):=\sum_{\mu=1}^m\langle[\mathbf{A}\bar{\mathbf{x}}]_\mu q_{\mathbf{x},\mathbf{s}}\rangle\langle[\mathbf{A}\bar{\mathbf{x}}]_\mu\rangle-\frac{\delta n^{\alpha-1}}{(1+\gamma_k(\nu)E_k)^2}\sum_{i=1}^n\langle\bar{x}_iq_{\mathbf{x},\mathbf{s}}\rangle_{k,\nu;\varepsilon}\langle\bar{x}_i\rangle_{k,\nu;\varepsilon}. \tag{333}$$

Hence, we have

$$\begin{aligned}\big|\mathbb{E}\big[g(\mathbf{X},\mathbf{S})\big]\big|&\le\mathbb{E}\bigg[\bigg|\sum_{\mu=1}^m\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu q_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu\rangle_{k,\nu;\varepsilon}\bigg|\bigg]\\&\quad+\frac{\delta n^{\alpha-1}}{(1+\gamma_k(\nu)E_k)^2}\mathbb{E}\bigg[\bigg|\sum_{i=1}^n\langle\bar{X}_iq_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}\langle\bar{X}_i\rangle_{k,\nu;\varepsilon}\bigg|\bigg] &(334)\\&\le\sum_{\mu=1}^m\mathbb{E}\bigg[\big|\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu q_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu\rangle_{k,\nu;\varepsilon}\big|\bigg]\\&\quad+\delta n^{\alpha-1}\sum_{i=1}^n\mathbb{E}\bigg[\big|\langle\bar{X}_iq_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}\langle\bar{X}_i\rangle_{k,\nu;\varepsilon}\big|\bigg]. &(335)\end{aligned}$$

Now, by using Cauchy's and Cauchy-Schwarz's inequalities, we have

$$\begin{aligned}&\mathbb{E}\bigg[\big|\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu q_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu\rangle_{k,\nu;\varepsilon}\big|\bigg]\\&\quad\le\frac{1}{2}\mathbb{E}\big[\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu q_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}^2\big]+\frac{1}{2}\mathbb{E}\big[\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu\rangle_{k,\nu;\varepsilon}^2\big] &(336)\\&\quad\le\frac{1}{2}\sqrt{\mathbb{E}[\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu\rangle_{k,\nu;\varepsilon}^4]\mathbb{E}[\langle q_{\mathbf{X},\mathbf{S}}^4\rangle_{k,\nu;\varepsilon}]}+\frac{1}{2}\sqrt{\mathbb{E}[\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu^4\rangle_{k,\nu;\varepsilon}]}. &(337)\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\mathbb{E}[\langle[\mathbf{A}\overline{\mathbf{X}}]_\mu\rangle_{k,\nu;\varepsilon}^4]&=\mathbb{E}[\langle[\mathbf{A}(\mathbf{X}-\mathbf{S})]_\mu\rangle_{k,\nu;\varepsilon}^4] &(338)\\&\le8\Big(\mathbb{E}\big[\langle[\mathbf{A}\mathbf{X}]_\mu\rangle_{k,\nu;\varepsilon}^4\big]+\mathbb{E}\big[[\mathbf{A}\mathbf{S}]_\mu^4\big]\Big) &(339)\\&=16\mathbb{E}\big[[\mathbf{A}\mathbf{S}]_\mu^4\big] &(340)\\&=16\bigg(\sum_{i=1}^n\mathbb{E}[A_{\mu,i}^4]\mathbb{E}[S_i^4]+6n(n-1)\sum_{i=1}^n\mathbb{E}[A_{\mu,i}^2]\mathbb{E}[S_i^2]\bigg) &(341)\\&=16\bigg(\frac{n}{m^2}\frac{k}{n}\mathbb{E}_{S\sim P_0}[S^4]+6n(n-1)\frac{n}{m}\frac{k}{n}\mathbb{E}_{S\sim P_0}[S^2]\bigg) &(342)\\&=O(n^2), &(343)\end{aligned}$$

where (339) follows from $(a+b)^4 \le 8(a^4+b^4)$.

On the other hand, since

$$\big|q_{\mathbf{x},\mathbf{s}}\big| = \frac{1}{n^{\alpha}}\bigg|\sum_{i=1}^{n} x_i s_i\bigg| \tag{344}$$

$$\le \frac{1}{n^{\alpha}}\sum_{i=1}^{n}\big|x_i s_i\big| \tag{345}$$

$$\le \frac{1}{n^{\alpha}} n s_{\max}^2 \tag{346}$$

$$= s_{\max}^2 n^{1-\alpha}. \tag{347}$$

From (337), (343), and (347), we obtain

$$\mathbb{E}\bigg[\big|\langle[\mathbf{A}\overline{\mathbf{X}}]_{\mu} q_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}\langle[\mathbf{A}\overline{\mathbf{X}}]_{\mu}\rangle_{k,\nu;\varepsilon}\big|\bigg] = O\big(n^{(3-\alpha)/2}\big), \tag{348}$$

where the constant does not depend on ν.

Similarly, we have

$$\mathbb{E}\bigg[\big|\langle \bar{X}_i q_{\mathbf{X},\mathbf{S}}\rangle_{k,\nu;\varepsilon}\langle \bar{X}_i\rangle_{k,\nu;\varepsilon}\big|\bigg] = O(1), \tag{349}$$

where the constant does not depend on i.

From (335), (348), and (349), we obtain

$$\big|\mathbb{E}\big[g(\mathbf{X},\mathbf{S})\big]\big| = O(n^{(3+\alpha)/2}). \tag{350}$$

From (332) and (350), we obtain

$$\bigg|\mathbb{E}\bigg[\mathbf{S}^T\frac{d}{d\nu}\langle\mathbf{X}\rangle_{k,\nu;\varepsilon}\bigg]\bigg| \le O\bigg(\frac{n^{\alpha}}{K_n}\bigg|\frac{d\gamma_k(\nu)}{d\nu}\bigg| n^{(3+\alpha)/2}\bigg), \tag{351}$$

where the constant does not depend on ν.

From (320), (324), and (351), for some constant C, we have

$$\bigg|\text{mmse}_{k,t;\varepsilon} - \text{mmse}_{k,0;\varepsilon}\bigg| \le \int_0^t \bigg|\frac{d\text{mmse}_{\text{k},\nu;\varepsilon}}{d\nu}\bigg| d\nu \tag{352}$$

$$\le \int_0^1 \bigg|\frac{d\text{mmse}_{\text{k},\nu;\varepsilon}}{d\nu}\bigg| d\nu \tag{353}$$

$$\le C\int_0^1 \frac{n^{\alpha}}{K_n}\bigg|\frac{d\gamma_k(\nu)}{d\nu}\bigg| n^{(3+\alpha)/2} d\nu \tag{354}$$

$$= -C\int_0^1 \frac{n^{\alpha}}{K_n}\bigg(\frac{d\gamma_k(\nu)}{d\nu}\bigg) n^{(3+\alpha)/2} d\nu \tag{355}$$

$$= O\bigg(\frac{n^{(3/2)(1+\alpha)}}{K_n\Delta_n}\bigg). \tag{356}$$

This concludes our proof of Lemma 5.

Appendix C.

In this Appendix, we provide a proof for Lemma 7.

First, we recall the following result.

Lemma 15 (Abel-Dini Theorem) *If $\sum_{n=1}^{\infty} a_n$ converges, then with T_n as the n-th tail, $\sum_{n=1}^{\infty} \frac{a_n}{T_n^{1+\alpha}}$ converges if and only if $\alpha < 1/2$.*

The proof is based on the proofs of (Fazekas and Klesov, 2001, Theorem 1.1) and (Fazekas and Klesov, 2001, Theorem 2.1). To begin with, we generalize Hájek-Rényi type maximal inequality (Fazekas and Klesov, 2001, Theorem 1.1) that under the condition (109), for any non-decreasing sequence of positive numbers $\{\beta_n\}_{n=1}^{\infty}$, it holds that

$$\mathbb{E}\left[\max_{1\le l\le n}\left|\frac{S_l}{\beta_l d_l}\right|^r\right] \le 4\sum_{l=1}^{n} \frac{\nu_l}{\beta_l^r d_l^{r-1/(1+\rho)}}. \tag{357}$$

To show (357), we can suppose that $\beta_1 = 1$. Let $c = 2^{1/r}$. Consider the sets

$$A_i = \left\{k : c^i \le \beta_k d_k < c^{i+1}\right\}, \qquad \forall i = 0, 1, 2, \cdots, \tag{358}$$

Denote by $i(n)$ the index of the last non-empty A_j such that $A_j \subset [n]$. It is clear that $i(n) \ge 0$ since $A_0 \ne \emptyset$ by $\beta_1 = \nu_1 = 1$. Let $k(i) = \max\{k : k \in A_i\}, i = 0, 1, 2, \cdots$ if A_i is non-empty, while $k(i) = k(i-1)$ if A_i is empty, and let $k(-1) = 0$. Let

$$\delta_l := d_{k(l)}^{1/(1+\rho)} \sum_{j=k(l-1)+1}^{k(l)} \nu_j, \qquad l = 0, 1, 2, \cdots, \tag{359}$$

where δ_l is considered to be zero if A_l is empty. We have

$$\mathbb{E}\left[\max_{1\le l\le n}\left|\frac{S_l}{\beta_l d_l}\right|^r\right] \le \sum_{i=0}^{i(n)} \mathbb{E}\left[\max_{l\in A_i}\left|\frac{S_l}{\beta_l d_l}\right|^r\right] \tag{360}$$

$$\le \sum_{i=0}^{i(n)} c^{-ir}\mathbb{E}\left[\max_{l\in A_i}\left|S_l\right|^r\right] \tag{361}$$

$$\le \sum_{i=0}^{i(n)} c^{-ir}\mathbb{E}\left[\max_{k\le k(i)}\left|S_k\right|^r\right] \tag{362}$$

$$= \sum_{i=0}^{i(n)} c^{-ir} d_{k(i)}^{1/(1+\rho)} \sum_{k=1}^{k(i)} \nu_k \tag{363}$$

$$= \sum_{i=0}^{i(n)} c^{-ir} \sum_{l=0}^{i} \delta_l \tag{364}$$

$$= \sum_{l=0}^{i(n)} \delta_l \sum_{i=l}^{i(n)} c^{-ir} \tag{365}$$

$$= \sum_{l=0}^{i(n)} \delta_l \sum_{i=l}^{\infty} c^{-ir} \tag{366}$$

$$= \frac{1}{1-c^{-r}} \sum_{l=0}^{i(n)} \delta_l c^{-lr} \tag{367}$$

$$= \frac{1}{1-c^{-r}} \sum_{l=0}^{i(n)} c^{-lr} d_{k(l)}^{1/(1+\rho)} \sum_{j=k(l-1)+1}^{k(l)} \nu_j \tag{368}$$

$$\leq \frac{c^r}{1-c^{-r}} \sum_{l=0}^{i(n)} d_{k(l)}^{1/(1+\rho)} \sum_{j=k(l-1)+1}^{k(l)} \frac{\nu_j}{\big(\beta_j d_j\big)^r} \tag{369}$$

$$\leq \frac{c^r}{1-c^{-r}} \sum_{l=0}^{i(n)} \sum_{j=k(l-1)+1}^{k(l)} \frac{\nu_j d_j^{1/(1+\rho)}}{\big(\beta_j d_j\big)^r} \tag{370}$$

$$= \frac{c^r}{1-c^{-r}} \sum_{l=0}^{i(n)} \sum_{j=k(l-1)+1}^{k(l)} \frac{\nu_j}{\beta_j^r d_j^{r-1/(1+\rho)}} \tag{371}$$

$$= \frac{c^r}{1-c^{-r}} \sum_{j=0}^{n} \frac{\nu_j}{\beta_j^r d_j^{r-1/(1+\rho)}} \tag{372}$$

$$= 4 \sum_{j=0}^{n} \frac{\nu_j}{\beta_j^r d_j^{r-1/(1+\rho)}}, \tag{373}$$

where (369) follows from $A_l = \{j : k(l-1)+1 \leq j \leq k(l)\}$ and $c^{-lr} \leq \frac{c^r}{(\beta_j d_j)^r}$ for all $j \in A_l$ which is achieved from (358), (370) follows from $d_{k(l)} \leq d_j$ for all $j \in A_l$, and (373) follows from $c = 2^{1/r}$.

Now, we return prove (111). As the proof of (Fazekas and Klesov, 2001, Theorem 2.1), we can assume that $\nu_n > 0$ for an infinite number of indices n. Otherwise, there exists an integer n_0 such that $\nu_n = 0$ for all $n \geq n_0$, so from the condition (109), we have $\mathbb{E}\big[\sup_{n\geq 1} |S_n|^r\big] < \infty$, so $\sup_{n\neq 1} |S_n| < \infty$ a.s., which easily gives (111). Now, set

$$t_n = \sum_{k=n}^{\infty} \frac{\nu_k}{b_k^r d_k^{r-1/(1+\rho)}} \tag{374}$$

$$\beta_n = \max_{1\leq k\leq n} b_k t_k^{1/(2r)}. \tag{375}$$

Then, it holds that

$$\sum_{k=1}^{\infty} \frac{\nu_k}{d_k^{r-1/(1+\rho)} \beta_k^r} \leq \sum_{k=1}^{\infty} \frac{\nu_k}{d_k^{r-1/(1+\rho)} b_k^r t_k^{1/2}} \tag{376}$$

$$< \infty, \tag{377}$$

where (376) follows from (375), and (377) follows from Lemma 15 for the convergence of series (Fazekas and Klesov, 2001) with $\alpha = -1/2$. On the other hand, from (375), β_n is non-decreasing. In addition, from (110) and (375), we have $t_n \to 0$ as $n \to \infty$. Hence, for any $\varepsilon > 0$, there exists an $n_0(\varepsilon) \in \mathbb{Z}^+$ such that

$$t_n^{1/(2r)} < \varepsilon \tag{378}$$

for all $n > n_0(\varepsilon)$. It follows that

$$\begin{aligned}
\beta_n &= \max_{1\le k\le n} b_k t_k^{1/(2r)} && (379)\\
&\le \max_{1\le k\le n_0(\varepsilon)} b_k t_k^{1/(2r)} + \max_{n_0(\varepsilon)<k\le n} b_k t_k^{1/(2r)} && (380)\\
&\le \max_{1\le k\le n_0(\varepsilon)} b_k t_k^{1/(2r)} + \varepsilon \max_{n_0(\varepsilon)<k\le n} b_k && (381)\\
&= \max_{1\le k\le n_0(\varepsilon)} b_k t_k^{1/(2r)} + \varepsilon b_n && (382)\\
& && (383)
\end{aligned}$$

where (381) follows from (378), and (382) follows from the non-decreasing property of the sequence $\{b_n\}$. From (382), we obtain

$$\begin{aligned}
0 &\le \frac{\beta_n}{b_n} && (384)\\
&\le \frac{1}{b_n} \max_{1\le k\le n_0(\varepsilon)} b_k t_k^{1/(2r)} + \varepsilon && (385)\\
&\le 2\varepsilon, && (386)
\end{aligned}$$

where (386) follows from $b_n \to \infty$. Hence, $\lim_{n\to\infty} \beta_n/b_n = 0$.

These facts imply that (357) holds, which leads to

$$\begin{aligned}
\mathbb{E}\left[\max_{l\ge 1}\left|\frac{S_l}{\beta_l d_l}\right|^r\right] &\le 4\sum_{l=1}^{\infty} \frac{\nu_l}{\beta_l^r d_l^{r-1/(1+\rho)}} && (387)\\
&< \infty, && (388)
\end{aligned}$$

where (387) follows from the monotone convergence theorem (Billingsley, 1995), and (388) follows from (110). This implies that

$$\max_{l\ge 1}\left|\frac{S_l}{\beta_l d_l}\right| < \infty, \qquad a.s. \tag{389}$$

Finally, we have

$$\begin{aligned}
0 &\le \left|\frac{S_l}{d_l b_l}\right| && (390)\\
&= \left|\frac{S_l}{d_l \beta_l}\right|\left|\frac{\beta_l}{b_l}\right| && (391)\\
&\le \Big(\sup_{l\ge 1}\left|\frac{S_l}{d_l\beta_l}\right|\Big)\left|\frac{\beta_l}{b_l}\right| && (392)\\
&\to 0 \qquad a.s. && (393)
\end{aligned}$$

as $l \to \infty$. This concludes the proof of Lemma 7.