# Universality of the Elastic Net Error*

Andrea Montanari[†] and Phan-Minh Nguyen[‡]

July 3, 2017

**Abstract**

We consider the reconstruction problem of a vector $\boldsymbol{x}_0 \in \mathbb{R}^n$ from a noisy linear observation $\bar{\boldsymbol{y}} = \bar{\boldsymbol{A}}\boldsymbol{x}_0 + \boldsymbol{w}$, where $\bar{\boldsymbol{A}} \in \mathbb{R}^{m \times n}$ is random, using the elastic net method. Assuming that the entries of $\bar{\boldsymbol{A}}$ are drawn independently and identically from any distribution that belongs to a large class, we prove the following universality result. In the high-dimensional asymptotic setting, where $n \to \infty$ and $\frac{m}{n} \to \delta > 0$, the normalized error of the elastic net minimizer converges in probability to a limit, insensitive to the exact distribution that the entries are drawn from. We also provide an explicit formula for the limit.

## 1 Background and Main Result

Consider the noisy linear observation model

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{A}}\boldsymbol{x}_0 + \boldsymbol{w}\,,$$

where $\boldsymbol{x}_0 \in \mathbb{R}^n$ is an unknown signal, $\bar{\boldsymbol{A}} \in \mathbb{R}^{m \times n}$ is a known matrix that models the observation mechanism, $\bar{\boldsymbol{y}} \in \mathbb{R}^m$ are observations and $\boldsymbol{w} \in \mathbb{R}^m$ is noise. The elastic net [ZH05] attempts to reconstruct $\boldsymbol{x}_0$ by solving the following convex program, with regularization parameters $\lambda, \rho > 0$:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \left[ \frac{1}{2} \left\| \bar{\boldsymbol{A}}\boldsymbol{x} - \bar{\boldsymbol{y}} \right\|_2^2 + \lambda \left\| \boldsymbol{x} \right\|_1 + \frac{\rho}{2} \left\| \boldsymbol{x} \right\|_2^2 \right]. \tag{1}$$

The elastic net reduces to the LASSO[1] [Tib13] for $\rho \to 0$, and to ridge regression as $\lambda \to 0$. It shares the good features of both of these approaches. In particular, the resulting estimate $\hat{\boldsymbol{x}}$ is sparse because of the $\ell_1$ term, but always uniquely defined due to the strong convexity of the regularizer. The elastic net has been applied successfully in a number of domains [HTW15, WZZ06, SWM14].

Random sensing matrices $\bar{\boldsymbol{A}}$ have attracted considerable amount of work because they offer good sparsity-undersampling tradeoffs in compressed sensing applications, and they provide a useful benchmark for deterministic constructions. In compressed sensing, it is popular to perform reconstruction using the LASSO ($\rho = 0$) or the basis pursuit if the observations are noiseless ($\rho = 0$, $\lambda \to 0$). Among other random matrix models, Gaussian matrices with i.i.d. entries proved to be amenable to several analytic approaches starting with the seminal work of Donoho which used

*This is the complete version of the conference paper [MN17].

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

[‡]Department of Electrical Engineering, Stanford University

[1]In fact, the elastic net solution converges to one of the LASSO solutions as $\rho \to 0$.

high-dimensional polytope geometry [Don05, BM12, Sto13, OTH13, TAH16]. While the Gaussian case is highly idealized, it was observed several times that predictions derived for this case are excellent approximations for a large variety of matrix models. For instance, [DT09] accumulated numerical evidence in this direction. It is expected that, for matrices $\boldsymbol{A}$ with i.i.d. entries, the asymptotic error of the LASSO (as $m, n \to \infty$) is independent of the entries distribution (under suitable tail conditions). This surprising phenomenon is known as *universality* and has classic analogues in probability and random matrix theory [Tao12].

In the last few years, various forms of universality have been proved for the LASSO and related linear inverse problems [KM11, BLM$^+$15, OT15] (see Section 1.4 for a discussion of these earlier works). However, none of these papers establishes the above conjecture, namely universality of the asymptotic estimation error of the LASSO.

In this paper we establish the analogous conjecture for the elastic net, for any $\rho > 0$. Specifically we show that the normalized error of the elastic net minimizer converges in probability to a universal limit, for sensing matrices $\bar{\boldsymbol{A}}$ with i.i.d. entries, under the assumption that $\sqrt{m}\bar{A}_{ij}$ has bounded $(4+\epsilon)$-th moment. We provide an explicit formula for this limit. As $\rho \to 0$, this formula reproduces the known asymptotics for the LASSO [BM12].

### 1.1 Mathematical conventions

We use boldfaced lower-case letters (e.g. $\boldsymbol{x}$) for vectors and boldface upper-case letters (e.g. $\boldsymbol{M}$) for matrices. As usual, $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{R}_+$ denote the set of natural numbers, real numbers, and non-negative real numbers respectively. For $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, 2, ..., n\}$. For a set $S \subseteq [n]$, $\bar{S}$ denotes its complement $[n] \backslash S$. For an event $\mathcal{E}$, its complement is denoted by $\mathcal{E}^{\complement}$.

For $\boldsymbol{x} \in \mathbb{R}^n$ and a set $S \subseteq [n]$, $\boldsymbol{x}_S$ denotes a vector in $\mathbb{R}^n$ in which its $i$-th entry is equal to $x_i$ if $i \in S$ and 0 otherwise. Likewise for $\boldsymbol{M} \in \mathbb{R}^{m \times n}$, $\boldsymbol{M}_S$ denotes a matrix in $\mathbb{R}^{m\times n}$ in which its $i$-th column is equal to the $i$-th column of $\boldsymbol{M}$ if $i \in S$ and the all-zero column vector otherwise.

For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{i=1}^n u_i v_i = \boldsymbol{u}^T \boldsymbol{v}$. We use $\mathbb{I}$ for the indicator function, i.e. $\mathbb{I}\,(\text{Clause})$ is equal to 1 if Clause is true and 0 otherwise. As usual, $\|\boldsymbol{x}\|_p$ and $\|\boldsymbol{M}\|_2$ denote the $p$-norm of $\boldsymbol{x}$ and the maximum singular value of $\boldsymbol{M}$ respectively.

We reserve the notations $\phi$ and $\Phi$ for the standard Gaussian probability density function and cumulative distribution function respectively, i.e.

$$\phi\,(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \qquad \Phi\,(x) = \int_{-\infty}^{x} \phi\,(t)\, dt.$$

We also use the phrase "with high probability" to indicate that the event occurs with probability converging to 1 as $n \to \infty$.

We recall the definition of Wasserstein distance. For $q \geq 1$, and $\mu$, $\nu$ two probability measures on $\mathbb{R}$ with finite $q$-th moment, let $\mathcal{W}_q\,(\mu, \nu)$ denote their order-$q$ Wasserstein distance on the metric space $(\mathbb{R}, \text{dist})$, where $\text{dist}\,(x, y) = |x - y|$. That is,

$$\mathcal{W}_q\,(\mu, \nu) = \left( \inf_{\lambda \in \Gamma(\mu,\nu)} \int |x - y|^q \, d\lambda\,(x, y) \right)^{1/q},$$

where $\Gamma\,(\mu, \nu)$ denotes the set of couplings of $\mu$ and $\nu$.

Following [BM11, BM12], we say that the function $\psi: \mathbb{R}^2 \to \mathbb{R}$ is pseudo-Lipschitz (of order 2) if there exists a constant $L > 0$ such that for any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$,

$$|\psi(\boldsymbol{u}) - \psi(\boldsymbol{v})| \le L\left(1 + \|\boldsymbol{u}\|_2 + \|\boldsymbol{v}\|_2\right)\|\boldsymbol{u} - \boldsymbol{v}\|_2 .$$

For pseudo-Lipschitz $\psi$ and $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, we denote $\frac{1}{n}\sum_{i=1}^n \psi(u_i, v_i)$ by $\psi_{\text{av}}(\boldsymbol{u}, \boldsymbol{v})$.

## 1.2 Main result

We assume the following setting:

- $\boldsymbol{x}_0 = \boldsymbol{x}_0(n)$ is from the sequence $\{\boldsymbol{x}_0(n)\}_{n\in\mathbb{N}}$ such that $\mathcal{W}_{10}(\mu_n^x, p_{X_0}) \to 0$ as $n \to \infty$, where $\mu_n^x$ denotes the empirical measure of $\boldsymbol{x}_0(n)$ and $p_{X_0}$ denotes the law of a random variable $X_0$ with $\mathbb{E}\left[X_0^2\right] \equiv \mathsf{M}_2$ and $\mathbb{E}\left[X_0^{10}\right] \equiv \mathsf{M}_{10} < \infty$.
- $\boldsymbol{w} = \boldsymbol{w}(n)$ is from the sequence $\{\boldsymbol{w}(n)\}_{n\in\mathbb{N}}$ such that $\mathcal{W}_4(\mu_n^w, p_W) \to 0$ as $n \to \infty$, where $\mu_n^w$ denotes the empirical measure of $\boldsymbol{w}(n)$ and $p_W$ denotes the law of a random variable $W$ with $\mathbb{E}\left[W^2\right] \equiv \sigma^2 > 0$ and $\mathbb{E}\left[W^4\right] < \infty$.
- $\bar{\boldsymbol{A}} = \bar{\boldsymbol{A}}(n) \in \mathbb{R}^{m\times n}$ for $m = m(n)$ that satisfies $m/n \to \delta > 0$, and entries of $\boldsymbol{A}(n)$ are drawn i.i.d.

Further, we call $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ a matrix of standard type if $M_{ij}$'s are i.i.d., $\mathbb{E}[M_{ij}] = 0$, $\mathbb{E}\left[M_{ij}^2\right] = 1$, and $\mathbb{E}\left[|M_{ij}|^{\mathfrak{p}}\right] < \infty$ for some $\mathfrak{p} > 4$. Note that we allow the distribution of $M_{ij}$ to be dependent on $n$. We shall consider $\bar{\boldsymbol{A}}$ such that $\sqrt{m}\bar{\boldsymbol{A}}$ is of standard type. We reserve the notation $\bar{\boldsymbol{G}}$ for matrix $\bar{\boldsymbol{A}}$ with Gaussian entries such that $\sqrt{m}\bar{\boldsymbol{G}}$ is of standard type.

We also note that by [Vil03, Theorem 7.12], $\mu_n^x$ and $\mu_n^w$ converge weakly to $X_0$ and $W$ respectively, with $\frac{1}{n}\|\boldsymbol{x}_0(n)\|_2^2 \to \mathsf{M}_2$, $\frac{1}{n}\|\boldsymbol{x}_0(n)\|_{10}^{10} \to \mathsf{M}_{10}$, $\frac{1}{m}\|\boldsymbol{w}(n)\|_2^2 \to \sigma^2$ and $\frac{1}{m}\|\boldsymbol{w}(n)\|_4^4$ converges to a finite constant as $n \to \infty$.

Let $\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)$ be the minimizer of the elastic net (1). Uniqueness of $\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)$ follows from that the objective function of (1) is strictly convex, for $\rho > 0$. We measure the estimation error of $\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)$ by $\psi_{\text{av}}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right), \boldsymbol{x}_0\right)$, where $\psi$ is an arbitrary pseudo-Lipshitz function. Describing the limit requires some definitions. We denote by $\eta: \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$ the soft-thresholding function, $\eta(x; a) = \text{sign}(x)\max(|x| - a, 0)$. For some $\lambda, \rho > 0$, let $\tau_*, \gamma, \alpha$ be the solutions to the system of equations

$$\tau_*^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left[\left(\frac{1}{\gamma+1}\eta\left(X_0 + \tau_* Z; \alpha\tau_*\right) - X_0\right)^2\right], \tag{2}$$

$$\lambda = \alpha\tau_*\left(1 - \frac{1}{\delta(\gamma+1)}\mathbb{P}\left(|X_0 + \tau_* Z| \ge \alpha\tau_*\right)\right), \tag{3}$$

$$\rho = \frac{\lambda\gamma}{\alpha\tau_*}, \tag{4}$$

in which $Z \sim \mathcal{N}(0,1)$ independent of $X_0$. The solutions can be proven to exist and be unique (see Section 5.11). Define

$$\psi^* = \mathbb{E}\left[\psi\left(\frac{1}{\gamma+1}\eta\left(X_0 + \tau_* Z; \alpha\tau_*\right), X_0\right)\right]. \tag{5}$$

We are now ready to state our main result.

**Theorem 1.** *In the above setting, let $\bar{\boldsymbol{A}}$ denote a sequence of matrices of standard type. Then, for any $\psi$ pseudo-Lipschitz, $\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right),\boldsymbol{x}_0\right)$ converges to $\psi^*$ in probability as $n\to\infty$, where $\psi^*=\psi^*\left(\lambda,\rho,\sigma,\delta,p_{X_0}\right)$ is defined as per Eq. (5).*

As a prototypical example, by specializing $\psi\left(x,y\right)$ to $\left(x-y\right)^2$ and $|x-y|$, we obtain from Theorem 1 the asymptotic formulas for the squared error $\frac{1}{n}\left\|\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)-\boldsymbol{x}_0\right\|_2^2$ and the $\ell_1$ error $\frac{1}{n}\left\|\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)-\boldsymbol{x}_0\right\|_1$. In particular, as $n\to\infty$,

$$\frac{1}{n}\left\|\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)-\boldsymbol{x}_0\right\|_2^2\to\delta\left(\tau_*^2-\sigma^2\right) \tag{6}$$

in probability, by Eq. (2) and (5).

Our proof analyzes a perturbation of (1), called the $s$-elastic net:

$$\mathsf{OPT}\left(s,\bar{\boldsymbol{A}}\right)\equiv\min_{\boldsymbol{x}\in\mathcal{B}}\frac{1}{n}\mathcal{C}\left(\boldsymbol{x};s,\bar{\boldsymbol{A}}\right), \tag{7}$$

in which

$$\mathcal{C}\left(\boldsymbol{x};s,\bar{\boldsymbol{A}}\right)=\frac{1}{2}\left\|\bar{\boldsymbol{A}}\left(\boldsymbol{x}-\boldsymbol{x}_0\right)-\boldsymbol{w}\right\|_2^2+\lambda\left\|\boldsymbol{x}\right\|_1+\frac{\rho}{2}\left\|\boldsymbol{x}\right\|_2^2+sn\psi_{\mathrm{av}}\left(\boldsymbol{x},\boldsymbol{x}_0\right),$$

$\mathcal{B}$ is an appropriately chosen compact set (which, in particular, follows Eq. (19)), and $s\in\mathbb{R}$. Observe that $\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)=\mathrm{argmin}_{\boldsymbol{x}\in\mathbb{R}^n}\,\mathcal{C}\left(\boldsymbol{x};0,\bar{\boldsymbol{A}}\right)$. We also let $\hat{\boldsymbol{x}}_s\left(\bar{\boldsymbol{A}}\right)$ be a minimizer of the $s$-elastic net. The definition (7) is motivated by the following identity:

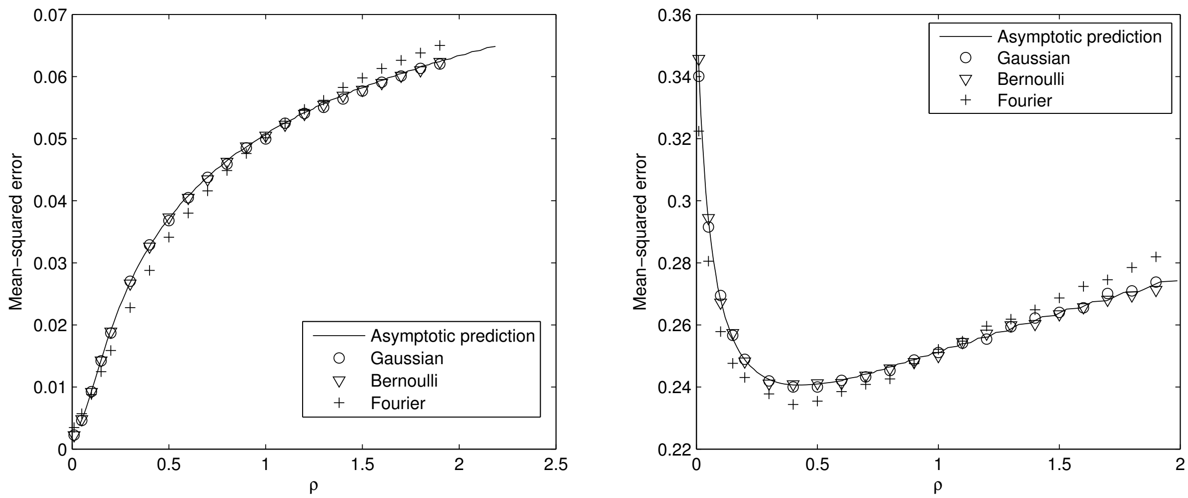
$$\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right),\boldsymbol{x}_0\right)=\frac{\mathrm{d}}{\mathrm{d}s}\mathsf{OPT}\left(0,\bar{\boldsymbol{A}}\right). \tag{8}$$

This relation suggests that one can study universality of $\psi\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right),\boldsymbol{x}_0\right)$ via universality of the cost $\mathsf{OPT}\left(s,\bar{\boldsymbol{A}}\right)$ for $s$ in a neighborhood of 0. Here we appeal to the Lindeberg's principle, following [KM11]. One main technical challenge is that universality of $\mathsf{OPT}\left(s,\bar{\boldsymbol{A}}\right)$ holds in the limit $n\to\infty$ for each fixed $s$, whereas the identity involves taking derivative w.r.t. $s$, i.e. $s\to0$. We overcome this problem by establishing a quantitative form of Eq. (8), cf. Lemma 11. Another challenge is that, in order to apply the Lindeberg method to $\mathsf{OPT}\left(s,\bar{\boldsymbol{A}}\right)$, we need $\mathcal{B}$ to be carefully chosen, and show that this restriction does not affect the optimizer. In particular, we establish *a priori* a bound on $\left\|\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)\right\|_p$ for some large $p>2$, cf. Lemma 9 and Eq. (19).

We note that the cases $\lambda=0$, $\sigma=0$ or $\delta=0$ can be treated by a continuity argument. Also, the assumptions on $\mathbb{E}\left[X_0^{10}\right]$ and $\mathbb{E}\left[W^4\right]$ are mainly introduced to simplify proofs. They can be relaxed to boundedness of $\mathbb{E}\left[X_0^p\right]$ and $\mathbb{E}\left[W^q\right]$ for some $p,q>2$, by applying the truncation technique, which we use on $\bar{\boldsymbol{A}}$ in the proof (cf. Section 2), on $\boldsymbol{x}_0$ and $\boldsymbol{w}$. In doing so, we obtain a better convergence rate of roughly $O\left(n^{-0.5}\right)$, instead of $O\left(n^{-0.188}\right)$ as to be seen in the proof of Proposition 13. We omit the details for clarity of presentation. On the other hand, $\rho>0$ is required for our proof technique to hold, and $\rho=0$ (i.e. the LASSO) requires non-trivial extension.

### 1.3 Numerical illustration

In Fig. 1, we compare the mean-squared error (MSE) results of numerical simulations for various distributions of the entries of $\bar{\boldsymbol{A}}$, as well as the asymptotic prediction from Eq. (6), for different parameter configurations. In the figures, "Gaussian" refers to $\bar{A}_{ij}\sim\mathcal{N}\left(0,\frac{1}{m}\right)$, "Bernoulli" refers




(a) (b)

Figure 1: Simulated MSE for different distributions. For both figures, we use $n = 1000$, $m = 640$, and $\lambda = 0.1$, with the simulations being averaged over 100 instances. In Figure (a), $\mathbb{P}(X_0 = 1) = \mathbb{P}(X_0 = -1) = 0.05$, $\mathbb{P}(X_0 = 0) = 0.9$, and $\sigma^2 = 10^{-4}$. In Figure (b), $\mathbb{P}(X_0 = 1) = \mathbb{P}(X_0 = -1) = 0.2$, $\mathbb{P}(X_0 = 0) = 0.6$, and $\sigma^2 = 0.2$.

to $\bar{A}_{ij} \sim \mathsf{Unif}\left(\pm\frac{1}{\sqrt{m}}\right)$, and "Fourier" refers to $\bar{\boldsymbol{A}}$ whose $m$ (distinct) rows are drawn at random from the $n$ rows of the $n \times n$ discrete cosine transform matrix.

Note that the "Fourier" case does not follow the i.i.d. entry assumption. Interestingly its elastic net MSE does not conform with the asymptotic prediction, unlike its $\ell_1 - \ell_0$ phase transition reported in [DT09].

Also, in Fig. 1(b), as the elastic net parameter $\rho$ is varied, we notice a pronounced minimum for some $\rho > 0$. In other words, it is possible that the elastic net substantially outperforms the LASSO. This is not unexpected when the non-zero coefficients are all roughly of the same size.

## 1.4 Related literature

This paper continues the line of work in [KM11, BM11, BM12]. In particular, [BM12] characterizes the asymptotic estimation error of the LASSO for Gaussian sensing matrices. The proof is based on the analysis of the approximate message passing (AMP) algorithm [BM11], which is proved to converge to the LASSO minimizer. An alternative approach based on Gordon's Gaussian min-max theorem was developed in [Sto13, OTH13, TOH15, TPH15, TAH16] and applied to several generalizations of the LASSO to structured inverse problems.

All these works are specific to the Gaussian case, and it is not immediate to export these proof techniques to non-Gaussian sensing matrices. Against this background, [KM11] used the Lindeberg method to prove universality of the value of the LASSO optimization problem, $\mathsf{OPT}\left(s = 0, \bar{\boldsymbol{A}}\right)$ for $\mathcal{B} = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_\infty \leq C\}$ a box and $\rho = 0$, for some constant $C$, cf. Eq. (7). However, the value of $\mathsf{OPT}\left(s = 0, \bar{\boldsymbol{A}}\right)$ does not have direct implications on other performance metrics. In this paper we show that a connection can instead be established by considering $s \neq 0$ in a neighborhood of 0. A different approach was developed in [BLM+15] which proved universality for the AMP algorithm, and showed that this implies universality of the $\ell_1 - \ell_0$ phase transition (for noiseless measurements), under certain conditions on the entries distributions (which –in particular– were required to have a density). This confirmed a conjecture put forward in [DT09] on the basis of extensive numerical simulations. The recent paper by Oymak and Tropp [OT15] applied again the Lindeberg method, supplemented by several geometric insights, and proved universality of the $\ell_1 - \ell_0$ phase transition under substantially weaker conditions, as well as for a broader set of linear inverse problems. The same paper also considers the case of non-vanishing noise but only proves universality of the noise stability coefficient.

From a technical perspective, our work builds directly on [KM11], and the scope of the ideas first introduced there.

We note that a large body of literature on the LASSO and elastic net in statistics and compressed sensing places an emphasis on sparsity or near-sparsity structure of $\boldsymbol{x}_0$, with the aim of its near-exact recovery. See, e.g. [HTW15]. In contrast, our result disregards such assumption and aids understanding of the elastic net method in a broader setting.

As a remark, we note that the elastic net penalization (i.e. the term $\lambda \|\boldsymbol{x}\|_1 + \frac{\rho}{2} \|\boldsymbol{x}\|_2^2$ in the objective function), with $\rho > 0$, is exploited in one crucial aspect: it is strongly convex. While our work concerns explicitly with the elastic net problem, it is foreseeable that universality can be proven to hold for other variants with strongly convex objective functions, using the outlined proof strategy. This includes the LASSO with $\delta > 1$[2].

### 1.5 Outline

The rest of the paper is dedicated to proving Theorem 1. The proof requires truncation of entries of $\bar{\boldsymbol{A}}$. In Section 2, we describe the truncation, state an analogue of Theorem 1 for the truncation (in particular, Proposition 2), and prove Proposition 2 as well as Theorem 1. The proof of Proposition 2 comprises of several key intermediate results. Some of those results, which concern with the technicalities required to establish universality as discussed in Section 1.2, are proven in Section 3. The others concern with convergence results in the Gaussian case, which is treated in Section 4, where we appeal to the mechanism of the AMP. Proofs of auxiliary lemmas are deferred to Section 5.

## 2 Proof of Theorem 1

### 2.1 Truncation of Entries of $\bar{\boldsymbol{A}}$

We define the (centered and normalized) truncation of the matrices. For some sufficiently large $\mathsf{R} > 0$, let $\boldsymbol{A}, \boldsymbol{G} \in \mathbb{R}^{m \times n}$ be such that

$$A_{ij} = \frac{\tilde{A}_{ij}}{\sqrt{m\mathbb{E}\left[\tilde{A}_{ij}^2\right]}}, \quad G_{ij} = \frac{\tilde{G}_{ij}}{\sqrt{m\mathbb{E}\left[\tilde{G}_{ij}^2\right]}} \tag{9}$$

[2] In the case of the LASSO with $\delta > 1$, strong convexity holds with high probability thanks to Theorem 26. In fact, the proof that universality holds for this case is an easy modification of the presented proof for the elastic net.

in which

$$\tilde{A}_{ij} = \bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right| \le \mathsf{R}\right) - \mathbb{E}\left[\bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right| \le \mathsf{R}\right)\right] \tag{10}$$

$$\tilde{G}_{ij} = \bar{G}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{G}_{ij}\right| \le \mathsf{R}\right) - \mathbb{E}\left[\bar{G}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{G}_{ij}\right| \le \mathsf{R}\right)\right] \tag{11}$$

It is immediate that $\mathbb{E}\left[A_{ij}\right] = \mathbb{E}\left[G_{ij}\right] = 0$ and $\mathbb{E}\left[A_{ij}^2\right] = \mathbb{E}\left[G_{ij}^2\right] = \frac{1}{\sqrt{m}}$. We have a universality result for the truncation.

**Proposition 2.** *Assume the setting in Theorem 1. For $\psi$ pseudo-Lipschitz, $\psi_{\text{av}}\left(\hat{\boldsymbol{x}}\left(\boldsymbol{A}\right), \boldsymbol{x}_0\right)$ converges to $\psi^*$ in probability as $n \to \infty$ then $\mathsf{R} \to \infty$, where $\psi^*$ is the constant given in Eq. (5).*

In the following, we show that Theorem 1 follows from this result. The next three auxiliary lemmas, concerning with properties of pseudo-Lipschitz functions, of the truncation on $\bar{\boldsymbol{A}}$, and geometry of the elastic net minimizer, shall be used repeatedly throughout our proofs. Their proofs are deferred to Section 5.

**Lemma 3.** *For $\psi$ pseudo-Lipschitz and $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{r}, \boldsymbol{t} \in \mathbb{R}^n$,*

$$\left|\psi_{\text{av}}\left(\boldsymbol{u}, \boldsymbol{v}\right) - \psi_{\text{av}}\left(\boldsymbol{r}, \boldsymbol{t}\right)\right| \le L\sqrt{5}\frac{\left\|\boldsymbol{u}-\boldsymbol{r}\right\|_2 + \left\|\boldsymbol{v}-\boldsymbol{t}\right\|_2}{\sqrt{n}}\sqrt{1 + \frac{\left\|\boldsymbol{u}\right\|_2^2}{n} + \frac{\left\|\boldsymbol{v}\right\|_2^2}{n} + \frac{\left\|\boldsymbol{r}\right\|_2^2}{n} + \frac{\left\|\boldsymbol{t}\right\|_2^2}{n}} \tag{12}$$

**Lemma 4.** *Some properties of the truncation $\boldsymbol{A}$:*

- *There exists $\mathsf{R}_0 > 0$ such that for all $\mathsf{R} \ge \mathsf{R}_0$, $\sqrt{m}\left|A_{ij}\right| \le 3\mathsf{R}$ with probability 1.*
- *$\left\|\boldsymbol{A} - \bar{\boldsymbol{A}}\right\|_2 \to 0$ in probability as $n \to \infty$ then $\mathsf{R} \to \infty$.*
- *$\left\|\boldsymbol{A}\right\|_2$ converges in probability to $1 + \frac{1}{\sqrt{\delta}}$ as $n \to \infty$.*

**Lemma 5.** *With high probability, $\left\|\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)\right\|_2^2 < n\mathsf{T} < \infty$ for some $\mathsf{T} = \mathsf{T}\left(\lambda, \rho, \sigma, \delta, p_{X_0}\right)$. Similarly, $\left\|\hat{\boldsymbol{x}}\left(\boldsymbol{A}\right)\right\|_2^2 < n\mathsf{T}$ with high probability. In particular, we choose $\mathsf{T} \ge 100\mathsf{M}_2$ so that $\left\|\boldsymbol{x}_0\right\|_2^2 \le n\mathsf{T}$ for sufficiently large $n$.*

These lemmas imply the following.

**Lemma 6.** $\left|\psi_{\text{av}}\left(\hat{\boldsymbol{x}}\left(\boldsymbol{A}\right), \boldsymbol{x}_0\right) - \psi_{\text{av}}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right), \boldsymbol{x}_0\right)\right| \to 0$ *in probability as $n \to \infty$ then $\mathsf{R} \to \infty$.*

*Proof.* Let $\bar{\boldsymbol{x}} = \hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)$ and $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}\left(\boldsymbol{A}\right)$ for brevity. We first claim that $\frac{1}{n}\left\|\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}}\right\|_2^2 \to 0$ in probability as $n \to \infty$ then $\mathsf{R} \to \infty$. Then the thesis follows easily in light of Lemma 3, the facts that $\frac{1}{n}\left\|\hat{\boldsymbol{x}}\right\|_2^2$ and $\frac{1}{n}\left\|\bar{\boldsymbol{x}}\right\|_2^2$ are bounded with high probability by Lemma 5, and that $\frac{1}{n}\left\|\boldsymbol{x}_0\right\|_2^2$ is bounded eventually.

To show the claim, by the KKT condition, we have

$$\bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}}\left(\bar{\boldsymbol{x}} - \boldsymbol{x}_0\right) - \bar{\boldsymbol{A}}^T\boldsymbol{w} + \lambda\boldsymbol{u} + \rho\bar{\boldsymbol{x}} = 0 \tag{13}$$

for some $\boldsymbol{u} \in \partial \|\bar{\boldsymbol{x}}\|_1$. As such, we get

$$
\begin{aligned}
0 &\le \mathcal{C}(\bar{\boldsymbol{x}}; 0, \boldsymbol{A}) - \mathcal{C}(\hat{\boldsymbol{x}}; 0, \boldsymbol{A}) && (14)\\
&= \frac{1}{2}\left[\boldsymbol{A}(\bar{\boldsymbol{x}} + \hat{\boldsymbol{x}} - 2\boldsymbol{x}_0) - 2\boldsymbol{w}\right]^T \boldsymbol{A}(\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}}) + \lambda\left(\|\bar{\boldsymbol{x}}\|_1 - \|\hat{\boldsymbol{x}}\|_1\right) + \frac{\rho}{2}\left(\|\bar{\boldsymbol{x}}\|_2^2 - \|\hat{\boldsymbol{x}}\|_2^2\right) && (15)\\
&= -\frac{1}{2}\|\boldsymbol{A}(\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}})\|_2^2 + \left[\left(\boldsymbol{A}^T\boldsymbol{A} - \bar{\boldsymbol{A}}^T\bar{\boldsymbol{A}}\right)(\bar{\boldsymbol{x}} - \boldsymbol{x}_0) - \left(\boldsymbol{A} - \bar{\boldsymbol{A}}\right)^T \boldsymbol{w}\right]^T (\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}})\\
&\quad + \lambda\left[\|\bar{\boldsymbol{x}}\|_1 - \|\hat{\boldsymbol{x}}\|_1 - \boldsymbol{u}^T(\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}})\right] - \frac{\rho}{2}\|\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2 && (16)\\
&\overset{(a)}{\le} \left(\left\|\boldsymbol{A} + \bar{\boldsymbol{A}}\right\|_2 \left\|\boldsymbol{A} - \bar{\boldsymbol{A}}\right\|_2 \|\bar{\boldsymbol{x}} - \boldsymbol{x}_0\|_2 + \left\|\boldsymbol{A} - \bar{\boldsymbol{A}}\right\|_2 \|\boldsymbol{w}\|_2\right) \|\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2 - \frac{\rho}{2}\|\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2 && (17)
\end{aligned}
$$

which yields, for $\rho > 0$,

$$
\|\bar{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2 \overset{(b)}{\le} \frac{2}{\rho}\left[\left(\|\boldsymbol{A}\|_2 + \left\|\bar{\boldsymbol{A}}\right\|_2\right)\left(\|\bar{\boldsymbol{x}}\|_2 + \|\boldsymbol{x}_0\|_2\right) + \|\boldsymbol{w}\|_2\right] \left\|\boldsymbol{A} - \bar{\boldsymbol{A}}\right\|_2 \tag{18}
$$

where $(a)$ follows from convexity of $\|\cdot\|_1$ and the Cauchy-Schwarz inequality, and $(b)$ is by the triangular inequality. The proof is complete with the second property in Lemma 4, along with the facts that $\frac{1}{n}\|\bar{\boldsymbol{x}}\|_2^2$, $\frac{1}{n}\|\hat{\boldsymbol{x}}\|_2^2$, $\|\boldsymbol{A}\|_2$ and $\left\|\bar{\boldsymbol{A}}\right\|_2$ are bounded with high probability by Lemma 5, the third property in Lemma 4 and Theorem 26, and that $\frac{1}{n}\|\boldsymbol{x}_0\|_2^2$ and $\frac{1}{n}\|\boldsymbol{w}\|_2^2$ are bounded eventually. □

*Proof of Theorem 1.* This is immediate from Lemma 6 and Proposition 2. □

The rest of the section focuses on the proof of Proposition 2.

## 2.2 Proof of Proposition 2

Recall the definition of $\boldsymbol{G}$, the truncation of $\bar{\boldsymbol{G}}$. We state some intermediate results. The next two lemmas concern with convergence results of $\hat{\boldsymbol{x}}(\boldsymbol{G})$ and $\mathsf{OPT}(0, \boldsymbol{G})$. Their proofs are deferred to Section 4.

**Lemma 7.** *For $\psi$ pseudo-Lipschitz, $\psi_{\mathrm{av}}(\hat{\boldsymbol{x}}(\boldsymbol{G}), \boldsymbol{x}_0)$ converges to $\psi^*$ in probability as $n \to \infty$ then $\mathsf{R} \to \infty$.*

**Lemma 8.** $\mathsf{OPT}(0, \boldsymbol{G})$*, defined with $\mathcal{B}$ given in Eq. (19) below, converges to $\mathcal{L}^*$ in probability as $n \to \infty$ then $\mathsf{R} \to \infty$, where $\mathcal{L}^* = \mathcal{L}^*(\lambda, \rho, \sigma, \delta, p_{X_0})$ is a (non-random) constant.*

The next lemma describes another geometric property of $\hat{\boldsymbol{x}}(\boldsymbol{A})$, in addition to Lemma 5. This property is crucial to establishing universality of $\mathsf{OPT}(s, \boldsymbol{A})$.

**Lemma 9.** *With high probability, $\|\hat{\boldsymbol{x}}(\boldsymbol{A})\|_\infty = O\left(n^{0.104}\right)$.*

By the above lemma and Lemma 5, with high probability, $\hat{\boldsymbol{x}}(\boldsymbol{A}) = \mathrm{argmin}_{\boldsymbol{x} \in \mathcal{B}} \mathcal{C}(\boldsymbol{x}; 0, \boldsymbol{A})$ for

$$
\mathcal{B} = \left\{\boldsymbol{x} \in \mathbb{R}^n :\ \|\boldsymbol{x}\|_2^2 \le n\mathsf{T},\ \|\boldsymbol{x}\|_\infty \le g(n)\right\}, \tag{19}
$$

for some $g(n) = O\left(n^{0.104}\right)$, and a constant $\mathsf{T}$ defined as per Lemma 5. We then define the $s$-elastic net (7) using this $\mathcal{B}$. As a companion note to the remark at the end of Section 1.2, in the situation that $\|\boldsymbol{x}_0\|_\infty \le c$ for some constant $c$, one can obtain that $\|\hat{\boldsymbol{x}}(\boldsymbol{A})\|_\infty = O(\sqrt{\log n})$ with high probability, a bound that is better than the one presented here. This situation arises as a consequence of applying the truncation technique to $\boldsymbol{x}_0$.

For $k > 0$, one can easily find functions $h_k^-$ and $h_k^+$ mapping from $\mathbb{R}$ to $[0,1]$, thrice continuously differentiable and non-increasing, such that $h_k^-(x) \le \mathbb{I}(x \le 0) \le h_k^+(x)$ and $h_k^+(x) \to \mathbb{I}(x \le 0)$, $h_k^-(x) \to \mathbb{I}(x < 0)$ as $k \to \infty$, for any $x \in \mathbb{R}$. In particular, we consider $h_k^-$ such that $h_k^-(x) = 1$ for $x \le -\frac{1}{k}$ and $h_k^-(x) = 0$ for $x \ge 0$, and $h_k^+(x) = h_k^-\left(x - \frac{|s|}{k}\right)$. Note that $h_k^-$ and $h_k^+$ depend on $s$, but we do not make this explicit for economy of notations. This dependency is immaterial for most parts, except for the proof of Proposition 2 below.

The following proposition establishes universality of $\mathsf{OPT}(s, \boldsymbol{A})$, where Lemma 9 is crucially made use of.

**Proposition 10.** *For any $s$ and any $\ell \in \mathbb{R}$, as $n \to \infty$,*

$$\left|\mathbb{E}\left[h_k^-\left(\mathsf{OPT}(s, \boldsymbol{A}) - \ell\right)\right] - \mathbb{E}\left[h_k^-\left(\mathsf{OPT}(s, \boldsymbol{G}) - \ell\right)\right]\right| \to 0 \,. \tag{20}$$

The following lemma, establishing an analogue of the identity (8), bridges universality of $\mathsf{OPT}(s, \boldsymbol{A})$ to that of $\psi_{\mathrm{av}}(\hat{\boldsymbol{x}}(\boldsymbol{A}), \boldsymbol{x}_0)$.

**Lemma 11.** *Let $\Delta(s, \boldsymbol{A}) = \mathsf{OPT}(s, \boldsymbol{A}) - \mathsf{OPT}(0, \boldsymbol{A})$. There exists $\epsilon(s) > 0$ such that $\epsilon(s)$ is independent of* $\mathsf{R}$*, $\epsilon(s) \downarrow 0$ as $s \to 0$ and*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\psi_{\mathrm{av}}(\hat{\boldsymbol{x}}(\boldsymbol{A}), \boldsymbol{x}_0) - \frac{\Delta(s, \boldsymbol{A})}{s}\right| \le \epsilon(s)\right) = 1 \,. \tag{21}$$

The proofs of Lemma 9, Proposition 10 and Lemma 11 can be found in Section 3. We are now ready for the proof of Proposition 2.

*Proof of Proposition 2.* By Lemma 11 and Proposition 8, there exists $\epsilon(s) > 0$ independent of $\mathsf{R}$ such that $\epsilon(s) \downarrow 0$ as $s \to 0$ and

$$\lim_{\mathsf{R}\to\infty} \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\mathsf{OPT}(s, \boldsymbol{G})}{s} - \frac{\mathcal{L}^* + s\psi^*}{s}\right| \le \epsilon(s)\right) = 1 \,. \tag{22}$$

Next, consider $s > 0$. By Proposition 10, for any $\ell \in \mathbb{R}$,

$$\mathbb{P}\left(\mathsf{OPT}(s, \boldsymbol{A}) \le \ell\right) \le \mathbb{E}\left[h_k^+\left(\mathsf{OPT}(s, \boldsymbol{A}) - \ell\right)\right] \tag{23}$$

$$= \mathbb{E}\left[h_k^-\left(\mathsf{OPT}(s, \boldsymbol{A}) - \ell - \frac{s}{k}\right)\right] \tag{24}$$

$$\le \mathbb{E}\left[h_k^-\left(\mathsf{OPT}(s, \boldsymbol{G}) - \ell - \frac{s}{k}\right)\right] + o_n(1) \tag{25}$$

$$\le \mathbb{P}\left(\mathsf{OPT}(s, \boldsymbol{G}) \le \ell + \frac{s}{k}\right) + o_n(1) \,, \tag{26}$$

where $o_n(1) \to 0$ as $n \to \infty$. Similarly,

$$\mathbb{P}\left(\mathsf{OPT}(s, \boldsymbol{A}) \le \ell\right) \ge \mathbb{P}\left(\mathsf{OPT}(s, \boldsymbol{G}) \le \ell - \frac{s}{k}\right) - o_n(1) \,, \tag{27}$$

which implies

$$\mathbb{P}\left(\left|\frac{\mathsf{OPT}\left(s,\boldsymbol{A}\right)}{s}-\frac{\ell}{s}\right|\leq 2\epsilon\left(s\right)\right)\geq\mathbb{P}\left(\left|\frac{\mathsf{OPT}\left(s,\boldsymbol{G}\right)}{s}-\frac{\ell}{s}\right|\leq 2\epsilon\left(s\right)-\frac{1}{k}\right)-o_n\left(1\right). \tag{28}$$

Taking $k=\frac{1}{\epsilon(s)}$ and $\ell=\mathcal{L}^*+s\psi^*$, we obtain

$$\lim_{\mathsf{R}\to\infty}\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{\mathsf{OPT}\left(s,\boldsymbol{A}\right)}{s}-\frac{\mathcal{L}^*+s\psi^*}{s}\right|\leq 2\epsilon\left(s\right)\right)=1\,. \tag{29}$$

Similarly we have $\mathsf{OPT}\left(0,\boldsymbol{A}\right)$ converges to $\mathcal{L}^*$ in probability as $n\to\infty$ then $\mathsf{R}\to\infty$, using Lemma 8. Then applying Lemma 11 completes the proof. □

# 3 Proof of Lemmas 9, 11, and Proposition 10

## 3.1 Proof of Lemma 11

For brevity, we drop $\boldsymbol{A}$ from the notations $\hat{\boldsymbol{x}}\left(\boldsymbol{A}\right)$ and $\hat{\boldsymbol{x}}_s\left(\boldsymbol{A}\right)$. The KKT condition yields

$$\boldsymbol{A}^T\boldsymbol{A}\left(\hat{\boldsymbol{x}}-\boldsymbol{x}_0\right)-\boldsymbol{A}^T\boldsymbol{w}+\lambda\boldsymbol{u}+\rho\hat{\boldsymbol{x}}=0 \tag{30}$$

for some $\boldsymbol{u}\in\partial\left\|\hat{\boldsymbol{x}}\right\|_1$. Since $\hat{\boldsymbol{x}}\in\mathcal{B}$ with high probability by Lemmas 5 and 9, we then have for any $s$,

$$\begin{aligned}
0 &\leq \mathcal{C}\left(\hat{\boldsymbol{x}};s,\boldsymbol{A}\right)-\mathcal{C}\left(\hat{\boldsymbol{x}}_s;s,\boldsymbol{A}\right) &(31)\\
&=\frac{1}{2}\left[\boldsymbol{A}\left(\hat{\boldsymbol{x}}+\hat{\boldsymbol{x}}_s-2\boldsymbol{x}_0\right)-2\boldsymbol{w}\right]^T\boldsymbol{A}\left(\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right)+\lambda\left(\left\|\hat{\boldsymbol{x}}\right\|_1-\left\|\hat{\boldsymbol{x}}_s\right\|_1\right)\\
&\quad+\frac{\rho}{2}\left(\left\|\hat{\boldsymbol{x}}\right\|_2^2-\left\|\hat{\boldsymbol{x}}_s\right\|_2^2\right)+sn\left(\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}},\boldsymbol{x}_0\right)-\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}}_s,\boldsymbol{x}_0\right)\right) &(32)\\
&=-\frac{1}{2}\left\|\boldsymbol{A}\left(\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right)\right\|_2^2-\frac{\rho}{2}\left\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right\|_2^2+\lambda\left[\left\|\hat{\boldsymbol{x}}\right\|_1-\left\|\hat{\boldsymbol{x}}_s\right\|_1-\boldsymbol{u}^T\left(\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right)\right]\\
&\quad+sn\left(\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}},\boldsymbol{x}_0\right)-\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}}_s,\boldsymbol{x}_0\right)\right) &(33)\\
&\overset{(a)}{\leq}-\frac{\rho}{2}\left\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right\|_2^2+\left|s\right|n\left|\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}},\boldsymbol{x}_0\right)-\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}}_s,\boldsymbol{x}_0\right)\right| &(34)\\
&\overset{(b)}{\leq}-\frac{\rho}{2}\left\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right\|_2^2+\left|s\right|c\sqrt{n}\left\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right\|_2 &(35)
\end{aligned}$$

where we use convexity of $\left\|\cdot\right\|_1$ in $(a)$, and $(b)$ is by Lemma 3 and that $\hat{\boldsymbol{x}}_s\in\mathcal{B}$, with $c=L\sqrt{5\left(1+2\mathsf{T}+2\mathsf{M}_2+0.1\right)}$ for sufficiently large $n$. We thus have

$$\frac{1}{\sqrt{n}}\left\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right\|_2\leq\frac{2c}{\rho}\left|s\right|\,. \tag{36}$$

Next observe that

$$\mathsf{OPT}\left(0,\boldsymbol{A}\right)+s\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}},\boldsymbol{x}_0\right)\geq\mathsf{OPT}\left(s,\boldsymbol{A}\right) \tag{37}$$

and therefore, for $s>0$,

$$\frac{-\Delta\left(-s\right)}{s}\geq\psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}},\boldsymbol{x}_0\right)\geq\frac{\Delta\left(s\right)}{s}\,. \tag{38}$$

On the other hand,

$$\frac{-\Delta(-s)-\Delta(s)}{s} \leq \frac{1}{ns}\left(\mathcal{C}(\hat{\boldsymbol{x}}_s;0)+\mathcal{C}(\hat{\boldsymbol{x}}_{-s};0)-\mathcal{C}(\hat{\boldsymbol{x}}_s;s)-\mathcal{C}(\hat{\boldsymbol{x}}_{-s};-s)\right) \tag{39}$$

$$= \psi_{\mathrm{av}}(\hat{\boldsymbol{x}}_{-s},\boldsymbol{x}_0)-\psi_{\mathrm{av}}(\hat{\boldsymbol{x}}_s,\boldsymbol{x}_0) \tag{40}$$

$$\leq \frac{c}{\sqrt{n}}\left\|\hat{\boldsymbol{x}}_{-s}-\hat{\boldsymbol{x}}_s\right\|_2 \tag{41}$$

$$\leq \frac{c}{\sqrt{n}}\left(\left\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_s\right\|_2+\left\|\hat{\boldsymbol{x}}-\hat{\boldsymbol{x}}_{-s}\right\|_2\right) \tag{42}$$

where we use Lemma 3, $\hat{\boldsymbol{x}}_s, \hat{\boldsymbol{x}}_{-s} \in \mathcal{B}$, and $\frac{1}{n}\left\|\boldsymbol{x}_0\right\|_2^2$ is bounded eventually. Using Eq. (36) completes the proof.

### 3.2 Proof of Lemma 9

We state an auxiliary lemma.

**Lemma 12.** *Fix $S \subset [n]$ such that $|S| \leq \epsilon n$ for some $\epsilon \in (0,1)$. Suppose $\boldsymbol{z} \in \mathbb{R}^m$ is a (deterministic) function of $\boldsymbol{A}_{\bar{S}}$, such that $\|\boldsymbol{z}\|_2 \leq \sqrt{m}$. Then:*

$$\mathbb{P}\left(\left\|\boldsymbol{A}_S^T\boldsymbol{z}\right\|_2^2 \geq n\epsilon + 4c\sqrt{2n\epsilon t}+4ct\right) \leq e^{-t}$$

*for some constant $c > 0$ independent of $S$ and $\epsilon$.*

*Proof.* Since $\boldsymbol{A}_{\bar{S}}$ is independent of $\boldsymbol{A}_S$, we can prove the claim for a fixed vector $\boldsymbol{z}$. This follows from standard concentration arguments for sub-Gaussian random linear transformations, using Bernstein's inequality (see cf. [BLM13]). This is applicable since $\boldsymbol{A}$ has bounded entries by Lemma 4. The details are omitted. □

*Proof of Lemma 9.* Let $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\boldsymbol{A})$ for brevity. The strategy is to examine $\|\hat{\boldsymbol{x}}_S\|_\infty$ for each subset $S$ of $[n]$. If we can bound $\|\hat{\boldsymbol{x}}_S\|_\infty$ for all subsets $S$, we obtain a bound on $\|\hat{\boldsymbol{x}}\|_\infty$. To bound $\|\hat{\boldsymbol{x}}_S\|_\infty$, one can instead bound $\|\hat{\boldsymbol{x}}_S\|_2$. Here the idea is to consider a perturbation of the elastic net, which itself is an elastic net problem, in which the optimization domain is restricted to $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}_S = \boldsymbol{0}\}$. The details are as follows.

For some $\epsilon \in (0,1)$, consider a partition $\mathcal{P}$ of $[n]$ into $\frac{1}{\epsilon}$ subsets $S$, each of which has size $n\epsilon$. Here we assume without loss of generality that $\frac{1}{\epsilon}$ and $n\epsilon$ are integers. Let $\mathcal{G}$ be the event $\|\boldsymbol{A}\|_2 \leq 100\left(1+\frac{1}{\sqrt{\delta}}\right)$, and $\mathcal{E}_S$ be the event $\|\hat{\boldsymbol{x}}_S\|_\infty > c\sqrt{n}\epsilon^{0.4}$ for some constant $c$ independent of $S$ and $\epsilon$, to be chosen later. We bound $\mathbb{P}(\mathcal{E}_S \cap \mathcal{G})$. Let $\boldsymbol{u} = \mathrm{argmin}_{\boldsymbol{x}\in\mathbb{R}^n,\, \boldsymbol{x}_S=\boldsymbol{0}}\, \mathcal{C}_{\bar{S}}(\boldsymbol{x})$ in which

$$\mathcal{C}_{\bar{S}}(\boldsymbol{x}) = \frac{1}{2}\left\|\boldsymbol{A}_{\bar{S}}\left(\boldsymbol{x}_{\bar{S}}-\boldsymbol{x}_{0,\bar{S}}\right)-\boldsymbol{w}\right\|_2^2+\lambda\left\|\boldsymbol{x}_{\bar{S}}\right\|_1+\frac{\rho}{2}\left\|\boldsymbol{x}_{\bar{S}}\right\|_2^2 . \tag{43}$$

Then $\boldsymbol{u}_S = \boldsymbol{0}$, and $\boldsymbol{u}_{\bar{S}}$ is a function of $\boldsymbol{A}_{\bar{S}}$. By the KKT condition, for some $\boldsymbol{v} \in \partial\left\|\boldsymbol{u}_{\bar{S}}\right\|_1$,

$$\boldsymbol{A}_{\bar{S}}^T\left(\boldsymbol{A}_{\bar{S}}\boldsymbol{u}_{\bar{S}}-\boldsymbol{A}_{\bar{S}}\boldsymbol{x}_{0,\bar{S}}-\boldsymbol{w}\right)+\lambda\boldsymbol{v}+\rho\boldsymbol{u}_{\bar{S}} = 0 . \tag{44}$$

Therefore, similar to the proof of Lemma 5,

$$
\begin{aligned}
\|\boldsymbol{u}\|_2 &= \|\boldsymbol{u}_{\bar{S}}\|_2 && (45)\\
&= \left\| \left[\boldsymbol{A}_{\bar{S}}^T \boldsymbol{A}_{\bar{S}} + \rho \boldsymbol{I}\right]^{-1} \left(-\lambda \boldsymbol{v} + \boldsymbol{A}_{\bar{S}}^T \boldsymbol{A}_{\bar{S}} \boldsymbol{x}_{0,\bar{S}} + \boldsymbol{A}_{\bar{S}}^T \boldsymbol{w}\right) \right\|_2 && (46)\\
&\overset{(a)}{\leq} \frac{1}{\rho} \left(\lambda \|\boldsymbol{v}\|_2 + \|\boldsymbol{A}_{\bar{S}}\|_2^2 \left\|\boldsymbol{x}_{0,\bar{S}}\right\|_2 + \|\boldsymbol{A}_{\bar{S}}\|_2 \|\boldsymbol{w}\|_2\right) && (47)\\
&\overset{(b)}{\leq} \frac{1}{\rho} \left(\lambda \sqrt{n} + \|\boldsymbol{A}\|_2^2 \|\boldsymbol{x}_0\|_2 + \|\boldsymbol{A}\|_2 \|\boldsymbol{w}\|_2\right) && (48)
\end{aligned}
$$

where $(a)$ is due to the Cauchy-Schwarz inequality and the triangular inequality, and $(b)$ is because $|v_i| \leq 1$ for all $i \in [n]$. Since $\|\boldsymbol{A}\|_2 \leq 100 \left(1 + \frac{1}{\sqrt{\delta}}\right)$ under $\mathcal{G}$, $\frac{1}{n} \|\boldsymbol{x}_0\|_2^2$ and $\frac{1}{n} \|\boldsymbol{w}\|_2^2$ are bounded eventually, we obtain that the event $\mathcal{G}$ implies $\|\boldsymbol{u}\|_2^2 \leq n\tilde{\mathsf{T}}$ for some constant $\tilde{\mathsf{T}}$ independent of $S$ and $\epsilon$.

Next we have:

$$
\begin{aligned}
0 &\leq \mathcal{C}(\boldsymbol{u}; 0, \boldsymbol{A}) - \mathcal{C}(\hat{\boldsymbol{x}}; 0, \boldsymbol{A}) && (49)\\
&\overset{(a)}{=} \lambda \left[\|\boldsymbol{u}_{\bar{S}}\|_1 - \|\hat{\boldsymbol{x}}_{\bar{S}}\|_1 - \boldsymbol{v}^T (\boldsymbol{u}_{\bar{S}} - \hat{\boldsymbol{x}}_{\bar{S}})\right] - \lambda \|\hat{\boldsymbol{x}}_S\|_1 - \frac{1}{2} \|\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2^2\\
&\quad - \hat{\boldsymbol{x}}_S^T \boldsymbol{A}_S^T \left(\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}_{\bar{S}} \boldsymbol{x}_{0,\bar{S}} - \boldsymbol{w}\right) - \boldsymbol{x}_{0,S}^T \boldsymbol{A}_S^T (\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}\hat{\boldsymbol{x}}) - \frac{\rho}{2} \left(\|\boldsymbol{u}_{\bar{S}} - \hat{\boldsymbol{x}}_{\bar{S}}\|_2^2 + \|\hat{\boldsymbol{x}}_S\|_2^2\right) && (50)\\
&\overset{(b)}{\leq} \|\hat{\boldsymbol{x}}_S\|_2 \left\|\boldsymbol{A}_S^T \left(\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}_{\bar{S}} \boldsymbol{x}_{0,\bar{S}} - \boldsymbol{w}\right)\right\|_2 + \|\boldsymbol{A}\|_2 \|\boldsymbol{x}_{0,S}\|_2 \|\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2\\
&\quad - \frac{1}{2} \|\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2^2 - \frac{\rho}{2} \|\hat{\boldsymbol{x}}_S\|_2^2 && (51)
\end{aligned}
$$

where in $(a)$, we use Eq. (44), and in $(b)$, we use convexity of $\|\cdot\|_1$. Firstly, note that

$$
\|\boldsymbol{x}_{0,S}\|_2 \leq |S|^{0.4} \|\boldsymbol{x}_{0,S}\|_{10} \leq |S|^{0.4} \|\boldsymbol{x}_0\|_{10} \leq \left(\mathsf{M}_{10}^{0.1} + 0.1\right) \sqrt{n} \epsilon^{0.4} \qquad (52)
$$

for sufficiently large $n$. Secondly, since $\mathcal{G}$ implies $\|\boldsymbol{u}\|_2^2 \leq n\tilde{\mathsf{T}}$, it also implies

$$
\left\|\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}_{\bar{S}} \boldsymbol{x}_{0,\bar{S}} - \boldsymbol{w}\right\|_2 \leq \|\boldsymbol{A}\|_2 (\|\boldsymbol{u}\|_2 + \|\boldsymbol{x}_0\|_2) + \|\boldsymbol{w}\|_2 \leq c_1 \sqrt{n} \qquad (53)
$$

for some constant $c_1$ independent of $S$ and $\epsilon$. By Lemma 12, there exists a constant $c_2 > 0$, independent of $S$ and $\epsilon$, such that for

$$
\tilde{\mathcal{E}}_S = \left\{ \left\|\boldsymbol{A}_S^T \left(\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}_{\bar{S}} \boldsymbol{x}_{0,\bar{S}} - \boldsymbol{w}\right)\right\|_2^2 > c_2 n \epsilon \right\},
$$

we have

$$
\begin{aligned}
\mathbb{P}\left(\tilde{\mathcal{E}}_S \cap \mathcal{G}\right) &\leq \mathbb{P}\left(\tilde{\mathcal{E}}_S \cap \left\{ \left\|\boldsymbol{A}_{\bar{S}} \boldsymbol{u}_{\bar{S}} - \boldsymbol{A}_{\bar{S}} \boldsymbol{x}_{0,\bar{S}} - \boldsymbol{w}\right\|_2 \leq c_1 \sqrt{n} \right\}\right) && (54)\\
&\leq e^{-n\epsilon}. && (55)
\end{aligned}
$$

Then using Eq. (51) and (52), we have that the event $\tilde{\mathcal{E}}_S^{\complement} \cap \mathcal{G}$ implies

$$\frac{1}{4}\left(\|\hat{\boldsymbol{x}}_S\|_2 + \frac{1}{\sqrt{\rho}}\|\boldsymbol{A}_{\bar{S}}\boldsymbol{u}_{\bar{S}} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2\right)^2 \leq \frac{1}{2}\|\hat{\boldsymbol{x}}_S\|_2^2 + \frac{1}{2\rho}\|\boldsymbol{A}_{\bar{S}}\boldsymbol{u}_{\bar{S}} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2^2 \tag{56}$$

$$\leq \|\hat{\boldsymbol{x}}_S\|_2\sqrt{\frac{c_2}{\rho^2}n\epsilon} + \frac{1}{\sqrt{\rho}}\|\boldsymbol{A}_{\bar{S}}\boldsymbol{u}_{\bar{S}} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2\, c_3\sqrt{n}\epsilon^{0.4} \tag{57}$$

$$\leq \frac{c}{4}\sqrt{n}\epsilon^{0.4}\left(\|\hat{\boldsymbol{x}}_S\|_2 + \frac{1}{\sqrt{\rho}}\|\boldsymbol{A}_{\bar{S}}\boldsymbol{u}_{\bar{S}} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2\right) \tag{58}$$

for some $c_3 > 0$ independent of $S$ and $\epsilon$, and $c = 4\max\left\{\frac{\sqrt{c_2}}{\rho}, c_3\right\}$. This implies $\|\hat{\boldsymbol{x}}_S\|_\infty \leq \|\hat{\boldsymbol{x}}_S\|_2 \leq c\sqrt{n}\epsilon^{0.4}$. Note that $c$ is independent of $S$ and $\epsilon$, since $c_1$, $c_2$ and $c_3$ are independent of $S$ and $\epsilon$. Therefore $\mathbb{P}(\mathcal{E}_S \cap \mathcal{G}) \leq e^{-n\epsilon}$.

Finally, by the union bound,

$$\mathbb{P}\left(\|\hat{\boldsymbol{x}}\|_\infty > c\sqrt{n}\epsilon^{0.4}\right) = \mathbb{P}\left(\bigcup_{S\in\mathcal{P}}\mathcal{E}_S\right) \leq \mathbb{P}\left(\mathcal{G}^{\complement}\right) + \sum_{S\in\mathcal{P}}\mathbb{P}(\mathcal{E}_S \cap \mathcal{G}) \tag{59}$$

$$\leq \mathbb{P}\left(\mathcal{G}^{\complement}\right) + \frac{1}{\epsilon}e^{-n\epsilon}. \tag{60}$$

By the third property in Lemma 4, $\mathbb{P}\left(\mathcal{G}^{\complement}\right) \to 0$ as $n \to \infty$. Choosing $\epsilon = O\left(n^{-0.99}\right)$ completes the proof. □

### 3.3 Proof of Proposition 10

For some $\epsilon > 0$, consider a minimal $\epsilon\sqrt{n}$-net $\mathcal{X}_\epsilon \subset \mathcal{B}$ in which for any $\boldsymbol{x} \in \mathcal{B}$, there exists $\boldsymbol{u} \in \mathcal{X}_\epsilon$ such that $\|\boldsymbol{u} - \boldsymbol{x}\|_2^2 \leq n\epsilon^2$. Without loss of generality, assume that $\boldsymbol{x}_0 \in \mathcal{X}_\epsilon$, which is valid since $\mathsf{T} > \mathsf{M}_2$ as per Lemma 5. Since $\mathcal{B} \subset \left\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2^2 \leq n\mathsf{T}\right\}$, a standard argument from the epsilon-net method yields $|\mathcal{X}_\epsilon| \leq \left(1 + 2\frac{\sqrt{\mathsf{T}}}{\epsilon}\right)^n$. Let us define

$$\mathsf{OPT}_\epsilon(s, \boldsymbol{A}) = \min_{\boldsymbol{x}\in\mathcal{X}_\epsilon}\frac{1}{n}\mathcal{C}(\boldsymbol{x}; s, \boldsymbol{A}). \tag{61}$$

We have the following universality result.

**Proposition 13.** *For any $\ell \in \mathbb{R}$,*

$$\lim_{n\to\infty}\left|\mathbb{E}\left[h_k^-\left(\mathsf{OPT}_\epsilon(s, \boldsymbol{A}) - \ell\right)\right] - \mathbb{E}\left[h_k^-\left(\mathsf{OPT}_\epsilon(s, \boldsymbol{G}) - \ell\right)\right]\right| = 0 \tag{62}$$

*Proof.* We use the Lindeberg's method (see e.g. [Cha06, KM11]). Some of our steps are similar to [KM11, Proof of Theorems 3 and 5]; we present here the full proof for completeness. For $\beta > 0$, define the soft-max function

$$f(\epsilon, \beta, \boldsymbol{A}) = -\frac{1}{n\beta}\log\sum_{\boldsymbol{x}\in\mathcal{X}_\epsilon}e^{-\beta\mathcal{C}(\boldsymbol{x}; s, \boldsymbol{A})}. \tag{63}$$

It is easy to see that

$$\lim_{\beta\to\infty}f(\epsilon, \beta, \boldsymbol{A}) = \min_{\boldsymbol{x}\in\mathcal{X}_\epsilon}\frac{1}{n}\mathcal{C}(\boldsymbol{x}; s, \boldsymbol{A}) = \mathsf{OPT}_\epsilon(s, \boldsymbol{A}). \tag{64}$$

Then since $h_k^-$ is monotone and continuous, we have

$$\lim_{\beta\to\infty} h_k^-\left(f\left(\epsilon,\beta,\boldsymbol{A}\right)-\ell\right)=h_k^-\left(\mathsf{OPT}_\epsilon\left(s,\boldsymbol{A}\right)-\ell\right) \tag{65}$$

for any $\ell\in\mathbb{R}$. A standard argument on the derivative of the soft-max function (e.g. as in [KM11, Eq. (17)]) gives

$$0\le\frac{\partial}{\partial\beta}f\left(\epsilon,\beta,\boldsymbol{A}\right)\le\frac{1}{n\beta^2}\log\left|\mathcal{X}_\epsilon\right|\le\frac{1}{\beta^2}\log\left(1+\frac{2\sqrt{\mathsf{T}}}{\epsilon}\right). \tag{66}$$

Therefore, for any $\beta>0$,

$$\begin{aligned}
&\lim_{n\to\infty}\left|\mathbb{E}\left[h_k^-\left(\mathsf{OPT}_\epsilon\left(s,\boldsymbol{A}\right)-\ell\right)\right]-\mathbb{E}\left[h_k^-\left(\mathsf{OPT}_\epsilon\left(s,\boldsymbol{G}\right)-\ell\right)\right]\right|\\
&\le\lim_{n\to\infty}\left|\mathbb{E}\left[h_k^-\left(f\left(\epsilon,\beta,\boldsymbol{A}\right)-\ell\right)\right]-\mathbb{E}\left[h_k^-\left(f\left(\epsilon,\beta,\boldsymbol{G}\right)-\ell\right)\right]\right|+c\int_\beta^\infty\frac{1}{t^2}\log\left(1+\frac{2\sqrt{\mathsf{T}}}{\epsilon}\right)dt \qquad (67)\\
&=\lim_{n\to\infty}\left|\mathbb{E}\left[h_k^-\left(f\left(\epsilon,\beta,\boldsymbol{A}\right)-\ell\right)\right]-\mathbb{E}\left[h_k^-\left(f\left(\epsilon,\beta,\boldsymbol{G}\right)-\ell\right)\right]\right|+\frac{c}{\beta}\log\left(1+\frac{2\sqrt{\mathsf{T}}}{\epsilon}\right) \qquad (68)
\end{aligned}$$

for some $c=c\left(k,s\right)$ since $\left|\frac{d}{dx}h_k^-\left(x\right)\right|$ is bounded for any $x\in\mathbb{R}$. The proof is complete by showing that

$$\lim_{n\to\infty}\left|\mathbb{E}\left[h_k^-\left(f\left(\epsilon,\beta,\boldsymbol{A}\right)-\ell\right)\right]-\mathbb{E}\left[h_k^-\left(f\left(\epsilon,\beta,\boldsymbol{G}\right)-\ell\right)\right]\right|=0\,. \tag{69}$$

and subsequently letting $\beta\to\infty$.

In the following, let $f\left(\boldsymbol{A}\right)=f\left(\epsilon,\beta,\boldsymbol{A}\right)$, $h\left(\boldsymbol{A}\right)=h_k^-\left(f\left(\epsilon,\beta,\boldsymbol{A}\right)-\ell\right)$, and $\mathcal{C}\left(\boldsymbol{A}\right)=\mathcal{C}\left(\boldsymbol{x};s,\boldsymbol{A}\right)$ for brevity. Let $\boldsymbol{D}=\boldsymbol{D}\left(q,p,v\right)\in\mathbb{R}^{m\times n}$ be such that its $(i,j)$-th entry is

$$D_{ij}=\begin{cases}A_{ij}, & i<q \text{ or } \left(i=q \text{ and } j<p\right),\\ v, & i=q \text{ and } j=p,\\ G_{ij}, & \text{otherwise.}\end{cases} \tag{70}$$

We only consider $|v|\le\frac{3\mathsf{R}}{\sqrt{m}}$. Let $\boldsymbol{z}=\boldsymbol{x}_0-\boldsymbol{x}$, and $Q_i\left(\boldsymbol{x},\boldsymbol{D}\right)=w_i+\sum_{j=1}^n D_{ij}z_j$. Define the operators $\langle\cdot\rangle$ and $\langle\cdot\rangle_{\sim q}$ as introduced in [KM11, Section IV.B]

$$\langle\cdot\rangle=\frac{\sum_{\boldsymbol{x}\in\mathcal{X}_\epsilon}\cdot e^{-\beta\mathcal{C}(\boldsymbol{x};s,\boldsymbol{D})}}{\sum_{\boldsymbol{x}'\in\mathcal{X}_\epsilon}e^{-\beta\mathcal{C}(\boldsymbol{x}';s,\boldsymbol{D})}}, \tag{71}$$

$$\langle\cdot\rangle_{\sim q}=\frac{\sum_{\boldsymbol{x}\in\mathcal{X}_\epsilon}\cdot e^{-\beta\left[\mathcal{C}(\boldsymbol{x};s,\boldsymbol{D})-Q_q^2(\boldsymbol{x},\boldsymbol{D})/2\right]}}{\sum_{\boldsymbol{x}'\in\mathcal{X}_\epsilon}e^{-\beta\left[\mathcal{C}(\boldsymbol{x}';s,\boldsymbol{D})-Q_q^2(\boldsymbol{x}',\boldsymbol{D})/2\right]}}. \tag{72}$$

We also use $\partial$, $\partial^2$ and $\partial^3$ to denote $\frac{\partial}{\partial v}$, $\frac{\partial^2}{\partial v^2}$ and $\frac{\partial^3}{\partial v^3}$ respectively. We note that for $\boldsymbol{x}\in\mathcal{X}_\epsilon$, recalling $\mathcal{X}_\epsilon\subset\mathcal{B}$ and $\mathcal{B}$ is defined as per Eq. (19),

$$\|\boldsymbol{z}\|_\infty\le\|\boldsymbol{x}\|_\infty+\|\boldsymbol{x}_0\|_\infty\le\|\boldsymbol{x}\|_\infty+\|\boldsymbol{x}_0\|_{10}=O\left(n^{0.104}\right), \tag{73}$$

$$\|\boldsymbol{z}\|_2\le\|\boldsymbol{x}\|_2+\|\boldsymbol{x}_0\|_2\le\sqrt{n}\left(\mathsf{T}+\mathsf{M}_2\right)=O\left(\sqrt{n}\right). \tag{74}$$

We consider $\mathbb{E}\left[\left\langle Q_q^4\left(\boldsymbol{x},\boldsymbol{D}\right)\right\rangle\right]$:

$$\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle Q_q^4\left(\boldsymbol{x},\boldsymbol{D}\right)\right\rangle\right]=\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\frac{\left\langle Q_q^4\left(\boldsymbol{x},\boldsymbol{D}\right)e^{-Q_q^2(\boldsymbol{x},\boldsymbol{D})/2}\right\rangle_{\sim q}}{\left\langle e^{-Q_q^2(\boldsymbol{x},\boldsymbol{D})/2}\right\rangle_{\sim q}}\right] \tag{75}$$

$$\overset{(a)}{\leq}\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle Q_q^4\left(\boldsymbol{x},\boldsymbol{D}\right)\right\rangle_{\sim q}\right] \tag{76}$$

$$\leq 27\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[w_q^4+\left\langle v^4z_p^4+\left(\sum_{i\in[n]\backslash\{p\}}D_{qi}z_i\right)^4\right\rangle_{\sim q}\right] \tag{77}$$

$$\leq 27\left\{n\left\|\boldsymbol{w}\right\|_4^4+nmv^4\left\|\boldsymbol{z}\right\|_\infty^4+\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle\left(\sum_{i\in[n]\backslash\{p\}}D_{qi}z_i\right)^4\right\rangle_{\sim q}\right]\right\} \tag{78}$$

$$\overset{(b)}{=}27\left\{n\left\|\boldsymbol{w}\right\|_4^4+nmv^4\left\|\boldsymbol{z}\right\|_\infty^4+\sum_{p=1}^{n}\sum_{q=1}^{m}\sum_{i_1,i_2\in[n]\backslash\{p\}}\mathbb{E}\left[D_{qi_1}^2D_{qi_2}^2\right]\mathbb{E}\left[\left\langle z_{i_1}^2z_{i_2}^2\right\rangle_{\sim q}\right]\right\} \tag{79}$$

$$\overset{(c)}{\leq}27\left\{n\left\|\boldsymbol{w}\right\|_4^4+\frac{3^4\mathsf{R}^4}{\delta}\left\|\boldsymbol{z}\right\|_\infty^4+\frac{3^4\mathsf{R}^4}{m^2}\sum_{p=1}^{n}\sum_{q=1}^{m}\sum_{i_1,i_2\in[n]\backslash\{p\}}\mathbb{E}\left[\left\langle z_{i_1}^2z_{i_2}^2\right\rangle_{\sim q}\right]\right\} \tag{80}$$

$$\leq 27\left\{n\left\|\boldsymbol{w}\right\|_4^4+\frac{3^4\mathsf{R}^4}{\delta}\left\|\boldsymbol{z}\right\|_\infty^4+\frac{3^4\mathsf{R}^4}{m^2}\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle\left\|\boldsymbol{z}\right\|_2^4\right\rangle_{\sim q}\right]\right\} \tag{81}$$

$$\overset{(d)}{=}O\left(n^2\right) \tag{82}$$

where $(a)$ is because $Q_q^4\left(\boldsymbol{x},\boldsymbol{D}\right)$ and $e^{-Q_q^2(\boldsymbol{x},\boldsymbol{D})/2}$ are negatively correlated, $(b)$ is because $\mathbb{E}\left[D_{ij}\right]=0$, $(c)$ is because $\left|D_{ij}\right|\leq\frac{3\mathsf{R}}{\sqrt{m}}$ by Lemma 4 and $|v|\leq\frac{3\mathsf{R}}{\sqrt{m}}$, and $(d)$ is because $\left\|\boldsymbol{z}\right\|_2=O\left(\sqrt{n}\right)$ and $\frac{1}{m}\left\|\boldsymbol{w}\right\|_4^4$ is bounded eventually. We thus have:

$$\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle\left|\partial\mathcal{C}\left(\boldsymbol{D}\right)\right|^3\right\rangle\right]=\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle\left|z_p\right|^3\left|Q_q\left(\boldsymbol{x},\boldsymbol{D}\right)\right|^3\right\rangle\right] \tag{83}$$

$$\leq\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle\left\|\boldsymbol{z}\right\|_\infty^3\left|Q_q\left(\boldsymbol{x},\boldsymbol{D}\right)\right|^3\right\rangle\right] \tag{84}$$

$$=O\left(n^{0.312}\right)\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle\left|Q_q\left(\boldsymbol{x},\boldsymbol{D}\right)\right|^3\right\rangle\right] \tag{85}$$

$$\overset{(a)}{\leq}O\left(n^{0.312}\right)\sum_{p=1}^{n}\sum_{q=1}^{m}\left(\mathbb{E}\left[\left\langle\left|Q_q\left(\boldsymbol{x},\boldsymbol{D}\right)\right|^4\right\rangle\right]\right)^{3/4} \tag{86}$$

$$\overset{(b)}{\leq}O\left(n^{0.812}\right)\left(\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle\left|Q_q\left(\boldsymbol{x},\boldsymbol{D}\right)\right|^4\right\rangle\right]\right)^{3/4} \tag{87}$$

$$=O\left(n^{2.312}\right) \tag{88}$$

where $(a)$ is by Jensen's inequality, and $(b)$ is by Holder's inequality. Similarly, one can also obtain

$$\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|\rangle\right]=O\left(n^{2.104}\right), \tag{89}$$

$$\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|^{2}\right\rangle\right]=O\left(n^{2.208}\right). \tag{90}$$

Furthermore, we have

$$\left|\partial^{2}\mathcal{C}\left(\boldsymbol{D}\right)\right|=z_{p}^{2}=O\left(n^{0.208}\right). \tag{91}$$

Note that

$$\mathbb{E}\left[\left(\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|\rangle\right)^{3}\right]\leq\mathbb{E}\left[\left\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|^{3}\right\rangle\right] \tag{92}$$

$$\mathbb{E}\left[\left(\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|\rangle\right)^{2}\right]\leq\mathbb{E}\left[\left\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|^{2}\right\rangle\right] \tag{93}$$

$$\mathbb{E}\left[\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|\rangle\left\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|^{2}\right\rangle\right]\leq\left(\mathbb{E}\left[\left(\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|\rangle\right)^{3}\right]\right)^{1/3}\left(\mathbb{E}\left[\left(\left\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|^{2}\right\rangle\right)^{3/2}\right]\right)^{2/3} \tag{94}$$

$$\leq\mathbb{E}\left[\left\langle|\partial\mathcal{C}\left(\boldsymbol{D}\right)|^{3}\right\rangle\right] \tag{95}$$

by Jensen's inequality and Holder's inequality. Direct calculations then yield

$$\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[|\partial f\left(\boldsymbol{D}\right)|\right]=O\left(n^{1.104}\right), \tag{96}$$

$$\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left|\partial^{2} f\left(\boldsymbol{D}\right)\right|\right]=O\left(n^{1.208}\right), \tag{97}$$

$$\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left|\partial^{3} f\left(\boldsymbol{D}\right)\right|\right]=O\left(n^{1.312}\right). \tag{98}$$

Since $h_{k}^{-}$ is thrice continuously differentiable,

$$\sum_{p=1}^{n}\sum_{q=1}^{m}\mathbb{E}\left[\left|\partial^{3} h\left(\boldsymbol{D}\right)\right|\right]=O\left(n^{1.312}\right). \tag{99}$$

The final step is to apply the Lindeberg's principle. In particular, since $|A_{ij}|\leq\frac{3\mathsf{R}}{\sqrt{m}}$ and $|G_{ij}|\leq\frac{3\mathsf{R}}{\sqrt{m}}$ with probability 1 by Lemma 4, we obtain from Theorem 28 that

$$\left|\mathbb{E}\left[h\left(\boldsymbol{A}\right)\right]-\mathbb{E}\left[h\left(\boldsymbol{G}\right)\right]\right|=O\left(n^{-0.188}\right). \tag{100}$$

The proof is complete. □

*Proof of Proposition 10.* Let $\hat{\boldsymbol{x}}_s\in\operatorname{argmin}_{\boldsymbol{x}\in\mathcal{B}}\mathcal{C}\left(\boldsymbol{x};s,\boldsymbol{A}\right)$ and $\boldsymbol{x}_\epsilon\in\operatorname{argmin}_{\boldsymbol{x}\in\mathcal{X}_\epsilon}\|\hat{\boldsymbol{x}}_s-\boldsymbol{x}\|_2$. Then $\|\hat{\boldsymbol{x}}_s-\boldsymbol{x}_\epsilon\|_2^2\leq n\epsilon^2$. We consider $|s|\leq s_0$ for some constant $s_0>0$. By the Cauchy-Schwarz inequality, the triangular inequality and Lemma 3,

$$\begin{aligned}\frac{1}{n}\mathcal{C}\left(\boldsymbol{x}_\epsilon; s, \boldsymbol{A}\right) - \mathsf{OPT}\left(s, \boldsymbol{A}\right) &= \frac{1}{2n}\left(\boldsymbol{A}\left(\boldsymbol{x}_\epsilon + \hat{\boldsymbol{x}}_s - 2\boldsymbol{x}_0\right) - 2\boldsymbol{w}\right)^T \boldsymbol{A}\left(\boldsymbol{x}_\epsilon - \hat{\boldsymbol{x}}_s\right) + \frac{\lambda}{n}\left(\left\|\boldsymbol{x}_\epsilon\right\|_1 - \left\|\hat{\boldsymbol{x}}_s\right\|_1\right) \\ &\quad + \frac{\rho}{2n}\left(\left\|\boldsymbol{x}_\epsilon\right\|_2^2 - \left\|\hat{\boldsymbol{x}}_s\right\|_2^2\right) + s\left(\psi_{\mathrm{av}}\left(\boldsymbol{x}_\epsilon, \boldsymbol{x}_0\right) - \psi_{\mathrm{av}}\left(\hat{\boldsymbol{x}}_s, \boldsymbol{x}_0\right)\right) \qquad (101) \\ &\leq \frac{1}{2n}\left[\left\|\boldsymbol{A}\right\|_2\left(\left\|\boldsymbol{x}_\epsilon\right\|_2 + \left\|\hat{\boldsymbol{x}}_s\right\|_2 + 2\left\|\boldsymbol{x}_0\right\|_2\right) + 2\left\|\boldsymbol{w}\right\|_2\right]\left\|\boldsymbol{A}\right\|_2\left\|\boldsymbol{x}_\epsilon - \hat{\boldsymbol{x}}_s\right\|_2 \\ &\quad + \frac{\lambda}{\sqrt{n}}\left\|\boldsymbol{x}_\epsilon - \hat{\boldsymbol{x}}_s\right\|_2 + \frac{\rho}{2n}\left(\left\|\boldsymbol{x}_\epsilon\right\|_2 + \left\|\hat{\boldsymbol{x}}_s\right\|_2\right)\left\|\boldsymbol{x}_\epsilon - \hat{\boldsymbol{x}}_s\right\|_2 \\ &\quad + |s|\, L\sqrt{5}\sqrt{1 + \frac{1}{n}\left\|\boldsymbol{x}_\epsilon\right\|_2^2 + \frac{1}{n}\left\|\hat{\boldsymbol{x}}_s\right\|_2^2 + \frac{2}{n}\left\|\boldsymbol{x}_0\right\|_2^2}\,\frac{\left\|\boldsymbol{x}_\epsilon - \hat{\boldsymbol{x}}_s\right\|_2}{\sqrt{n}} \qquad (102) \\ &\leq \frac{c}{2\sqrt{n}}\left\|\hat{\boldsymbol{x}}_s - \boldsymbol{x}_\epsilon\right\|_2 \qquad (103)\end{aligned}$$

with high probability for some constant $c > 0$, in light of the fact that $\frac{1}{n}\left\|\hat{\boldsymbol{x}}_s\right\|_2^2$ and $\frac{1}{n}\left\|\boldsymbol{x}_\epsilon\right\|_2^2$ are bounded due to $\hat{\boldsymbol{x}}_s, \boldsymbol{x}_\epsilon \in \mathcal{B}$, $\left\|\boldsymbol{A}\right\|_2$ is bounded eventually by the third property in Lemma 4, and $\frac{1}{n}\left\|\boldsymbol{x}_0\right\|_2^2$ and $\frac{1}{n}\left\|\boldsymbol{w}\right\|_2^2$ are bounded eventually. Note that $c$ is independent of $s$ since we consider $|s| \leq s_0$. Since $\mathsf{OPT}\left(s, \boldsymbol{A}\right) \leq \mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right) \leq \frac{1}{n}\mathcal{C}\left(\boldsymbol{x}_\epsilon; s, \boldsymbol{A}\right)$, we then have

$$\lim_{n\to\infty}\mathbb{P}\left(\mathcal{E}_{n,\epsilon}\right) = 0, \quad \mathcal{E}_{n,\epsilon} = \left\{\left|\mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right) - \mathsf{OPT}\left(s, \boldsymbol{A}\right)\right| > c\epsilon\right\}. \qquad (104)$$

On the other hand, since $\mathsf{T} > \mathsf{M}_2$ as per Lemma 5, with probability 1,

$$\left|\mathsf{OPT}\left(s, \boldsymbol{A}\right)\right| \leq \left|\frac{1}{n}\mathcal{C}\left(\boldsymbol{x}_0; s, \boldsymbol{A}\right)\right| \leq \frac{1}{2n}\left\|\boldsymbol{w}\right\|_2^2 + \frac{\lambda}{n}\left\|\boldsymbol{x}_0\right\|_1 + \frac{\rho}{2n}\left\|\boldsymbol{x}_0\right\|_2^2 + |s|\left|\psi_{\mathrm{av}}\left(\boldsymbol{x}_0, \boldsymbol{x}_0\right)\right|. \qquad (105)$$

Note that $\left|\psi_{\mathrm{av}}\left(\boldsymbol{x}_0, \boldsymbol{x}_0\right)\right| \leq \left|\psi_{\mathrm{av}}\left(\boldsymbol{x}_0, \boldsymbol{x}_0\right) - \psi_{\mathrm{av}}\left(\mathbf{0}, \mathbf{0}\right)\right| + \psi\left(0, 0\right)$. Then from the fact that $\frac{1}{n}\left\|\boldsymbol{w}\right\|_2^2$ and $\frac{1}{n}\left\|\boldsymbol{x}_0\right\|_2^2$ are bounded, and Lemma 3, we have there exists a constant $M$, independent of $n$ and $\epsilon$, such that $\left|\mathsf{OPT}\left(s, \boldsymbol{A}\right)\right| \leq M$ with probability 1 for sufficiently large $n$. Similarly, recalling the fact that $\boldsymbol{x}_0 \in \mathcal{X}_\epsilon$, we have $\left|\mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right)\right| \leq M$ with probability 1 for sufficiently large $n$. Therefore,

$$\begin{aligned}\left|\mathbb{E}\left[\mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right) - \mathsf{OPT}\left(s, \boldsymbol{A}\right)\right]\right| &\leq \mathbb{E}\left[\left|\mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right) - \mathsf{OPT}\left(s, \boldsymbol{A}\right)\right|\right] \qquad (106) \\ &= \mathbb{E}\left[\left|\mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right) - \mathsf{OPT}\left(s, \boldsymbol{A}\right)\right|\mathbb{I}\left(\mathcal{E}_{n,\epsilon}^{\complement}\right)\right] \\ &\quad + \mathbb{E}\left[\left|\mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right) - \mathsf{OPT}\left(s, \boldsymbol{A}\right)\right|\mathbb{I}\left(\mathcal{E}_{n,\epsilon}\right)\right] \qquad (107) \\ &\leq c\epsilon + 2M\mathbb{P}\left(\mathcal{E}_{n,\epsilon}\right). \qquad (108)\end{aligned}$$

Taking $n \to \infty$ then $\epsilon \downarrow 0$, we have $\mathbb{E}\left[\mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right) - \mathsf{OPT}\left(s, \boldsymbol{A}\right)\right] \to 0$. Since $h_k^-$ is continuously differentiable,

$$\lim_{\epsilon\downarrow 0}\lim_{n\to\infty}\left|\mathbb{E}\left[h_k^-\left(\mathsf{OPT}_\epsilon\left(s, \boldsymbol{A}\right) - \ell\right)\right] - \mathbb{E}\left[h_k^-\left(\mathsf{OPT}\left(s, \boldsymbol{A}\right) - \ell\right)\right]\right| = 0 \qquad (109)$$

for any $\ell \in \mathbb{R}$. Note that this also applies to $\boldsymbol{G}$. Combing this result with Proposition 13, we obtain

$$\lim_{n\to\infty}\left|\mathbb{E}\left[h_k^-\left(\mathsf{OPT}\left(s, \boldsymbol{A}\right) - \ell\right)\right] - \mathbb{E}\left[h_k^-\left(\mathsf{OPT}\left(s, \boldsymbol{G}\right) - \ell\right)\right]\right| = 0 \qquad (110)$$

completing the proof. □

## 4 The Gaussian Case & AMP

Lemmas 7 and 8 follow from the following result.

**Theorem 14.** *For $\psi$ pseudo-Lipschitz, $\psi_{\text{av}}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{G}}\right), \boldsymbol{x}_0\right)$ converges to $\psi^*$ in probability as $n \to \infty$, where $\psi^*$ is a constant defined as per Eq. (5). Furthermore, $\frac{1}{n}\mathcal{C}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{G}}\right); 0, \bar{\boldsymbol{G}}\right)$ converges to $\mathcal{L}^*$ in probability as $n \to \infty$, where $\mathcal{L}^* = \mathcal{L}^*\left(\lambda, \rho, \sigma, \delta, p_{X_0}\right)$ is a (non-random) constant.*

*Proof of Lemma 7.* Immediate from Lemma 6 and Theorem 14. □

*Proof of Lemma 8.* Letting $\mathcal{B}' = \left\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2^2 \leq n\mathsf{T}\right\}$, we have with high probability,

$$\hat{\boldsymbol{x}}(\boldsymbol{G}) = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{B}'} \mathcal{C}(\boldsymbol{x}; 0, \boldsymbol{G}) = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{B}} \mathcal{C}(\boldsymbol{x}; 0, \boldsymbol{G}) \tag{111}$$

by Lemmas 5 and 9. Also by Lemma 5,

$$\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{G}}\right) = \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{B}'} \mathcal{C}\left(\boldsymbol{x}; 0, \bar{\boldsymbol{G}}\right) \tag{112}$$

with high probability. Hence,

$$\left|\mathsf{OPT}(0, \boldsymbol{G}) - \frac{1}{n}\mathcal{C}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{G}}\right); 0, \bar{\boldsymbol{G}}\right)\right| = \frac{1}{n}\left|\min_{\boldsymbol{x}\in\mathcal{B}'} \mathcal{C}(\boldsymbol{x}; 0, \boldsymbol{G}) - \min_{\boldsymbol{x}\in\mathcal{B}'} \mathcal{C}\left(\boldsymbol{x}; 0, \bar{\boldsymbol{G}}\right)\right| \tag{113}$$

$$\leq \frac{1}{n}\max_{\boldsymbol{x}\in\mathcal{B}'}\left|\mathcal{C}(\boldsymbol{x}; 0, \boldsymbol{G}) - \mathcal{C}\left(\boldsymbol{x}; 0, \bar{\boldsymbol{G}}\right)\right| \tag{114}$$

$$\leq \frac{1}{2n}\max_{\boldsymbol{x}\in\mathcal{B}'}\left[\left(\left\|\boldsymbol{G} + \bar{\boldsymbol{G}}\right\|_2 \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 + 2\|\boldsymbol{w}\|_2\right)\left\|\boldsymbol{G} - \bar{\boldsymbol{G}}\right\|_2 \|\boldsymbol{x} - \boldsymbol{x}_0\|_2\right] \tag{115}$$

By the second property in Lemma 4, $\left\|\boldsymbol{G} - \bar{\boldsymbol{G}}\right\|_2 \to 0$ in probability as $n \to \infty$ then $\mathsf{R} \to \infty$. The proof is then completed with Theorem 14, along with the facts that $\frac{1}{n}\|\boldsymbol{x}\|_2^2 \leq \mathsf{T}$ for $\boldsymbol{x} \in \mathcal{B}'$, $\frac{1}{n}\|\boldsymbol{x}_0\|_2^2$ and $\frac{1}{n}\|\boldsymbol{w}\|_2^2$ are bounded eventually, $\left\|\bar{\boldsymbol{G}}\right\|_2$ is bounded eventually by Theorem 26, and the third property in Lemma 4 for boundedness of $\|\boldsymbol{G}\|_2$. □

In the rest of this section, we establish Theorem 14. We use the AMP to construct an iterative algorithm which solves the elastic net (1), provably in the case of Gaussian matrix $\bar{\boldsymbol{G}}$. In Section 4.1, we describe the algorithm as well as the *state evolution* equation, which track the behavior of the AMP in the asymptotics $n \to \infty$. In Section 4.2, we describe the asymptotics of the state evolution, and show how to relate them to the parameters of the elastic net problem (1). We prove Theorem 14 in Section 4.3.

As a remark, the work [TAH16] proves a convergence result for the squared error of the regularized M-estimator with Gaussian sensing matrices, using the Gordon's Gaussian min-max theorem. The same line of work is extended in [ATH16] to Lipschitz error functions. While there are overlaps with Theorem 14, in which the elastic net is a special case of the M-estimator, we note a key difference in that our proof complements with an algorithm that provably solves the elastic net.

As another remark, the AMP we construct here is an extension to the one considered in [BM12], which obtains a result similar to Theorem 14 for the LASSO, and our proof also follows the same line. The fact $\rho > 0$ eases certain parts as compared to the proof in [BM12]. On the other hand, the bulk of our work here lies in establishing that it is possible to construct the AMP for every elastic net instances, i.e. for every sets of parameters $\delta$, $\lambda$ and $\rho$ (cf. Section 4.2).

### 4.1 AMP Recursions

Recall the definition of the soft-thresholding function $\eta$:

$$\eta\left(x;\chi\right)=\operatorname{sign}(x)\max(|x|-\chi,0)\,. \tag{116}$$

Fix $\gamma>0$. For each $\gamma$ and for a non-negative sequence of thresholds $\{\chi_t\}_{t\in\mathbb{N}}$, the AMP iterates are defined as follows:

$$\boldsymbol{x}^{t+1}=\frac{1}{\gamma+1}\eta_t\left(\bar{\boldsymbol{G}}^T\boldsymbol{z}^t+\boldsymbol{x}^t\right), \tag{117}$$

$$\boldsymbol{z}^t=\bar{\boldsymbol{y}}-\bar{\boldsymbol{G}}\boldsymbol{x}^t+\frac{1}{\delta\left(\gamma+1\right)}\mu^{t-1}\boldsymbol{z}^{t-1} \tag{118}$$

initialized with $\boldsymbol{x}^0=\boldsymbol{0}$ and $\boldsymbol{z}^0=\bar{\boldsymbol{y}}=\bar{\boldsymbol{G}}\boldsymbol{x}_0+\boldsymbol{w}$, in which

$$\mu^t=\frac{1}{n}\left\langle\boldsymbol{1},\eta_t'\left(\bar{\boldsymbol{G}}^T\boldsymbol{z}^t+\boldsymbol{x}^t\right)\right\rangle \tag{119}$$

and $\eta_t\left(\cdot\right)\equiv\eta\left(\cdot;\chi_t\right)$. With an abuse of notation, we will write $\eta_\chi\left(\cdot\right)$ for $\eta\left(\cdot;\chi\right)$ whenever the context is clear.

*Remark.* Given $\boldsymbol{w}\in\mathbb{R}^m$, $\boldsymbol{x}_0\in\mathbb{R}^n$ and $\boldsymbol{M}\in\mathbb{R}^{m\times n}$, the (more general) AMP iterates studied in [BM11] are defined by the iterations

$$\boldsymbol{h}^{t+1}=\boldsymbol{M}^T\boldsymbol{m}^t-\frac{1}{m}\left\langle\boldsymbol{1},g_t'\left(\boldsymbol{b}^t,\boldsymbol{w}\right)\right\rangle\boldsymbol{q}^t, \tag{120}$$

$$\boldsymbol{b}^t=\boldsymbol{M}\boldsymbol{q}^t-\frac{1}{m}\left\langle\boldsymbol{1},f_t'\left(\boldsymbol{h}^t,\boldsymbol{x}_0\right)\right\rangle\boldsymbol{m}^{t-1}, \tag{121}$$

$$\boldsymbol{m}^t=g_t\left(\boldsymbol{b}^t,\boldsymbol{w}\right), \tag{122}$$

$$\boldsymbol{q}^t=f_t\left(\boldsymbol{h}^t,\boldsymbol{x}_0\right), \tag{123}$$

with a given initialization $\boldsymbol{q}^0$, and $\boldsymbol{m}^{-1}=\boldsymbol{0}$. Here $\{g_t\}_{t\in\mathbb{N}}$ and $\{f_t\}_{t\in\mathbb{N}}$ are two sequences of Lipschitz continuous functions mapping from $\mathbb{R}^2\mapsto\mathbb{R}$, and $g_t'$ and $f_t'$ denote the respective derivatives w.r.t. the first argument. Our AMP iterates (117)-(118) fit this framework, in which we consider Gaussian matrix $\bar{\boldsymbol{G}}$ in place of $\boldsymbol{M}$, and specialize the iterates to the following:

$$\boldsymbol{h}^{t+1}=\boldsymbol{x}_0-\left(\bar{\boldsymbol{G}}^T\boldsymbol{z}^t+\boldsymbol{x}^t\right), \tag{124}$$

$$\boldsymbol{b}^t=\boldsymbol{w}-\boldsymbol{z}^t, \tag{125}$$

$$\boldsymbol{m}^t=-\boldsymbol{z}^t, \tag{126}$$

$$\boldsymbol{q}^t=\boldsymbol{x}^t-\boldsymbol{x}_0, \tag{127}$$

with the initialization $\boldsymbol{q}^0=-\boldsymbol{x}_0$. Correspondingly,

$$g_t\left(r,u\right)=r-u,\qquad f_t\left(r,u\right)=\frac{1}{\gamma+1}\eta_{t-1}\left(u-r\right)-u. \tag{128}$$

Hence results from [BM11] are applicable to our setting.

To track the behavior of the AMP iterates in the limit $n \to \infty$, we define the following (scalar) *state evolution* equation, which is a (scalar) recursion for a non-negative sequence $\{\tau_t\}_{t\in\mathbb{N}}$:

$$\tau_t^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left[\left(\frac{1}{\gamma+1}\eta_{t-1}\left(X_0 + \tau_{t-1}Z\right) - X_0\right)^2\right], \tag{129}$$

$$\tau_0^2 = \sigma^2 + \frac{1}{\delta}\mathsf{M}_2, \tag{130}$$

where $Z \sim \mathcal{N}(0,1)$ independent of $X_0$ and $W$. Recall that the distributions of $X_0$ and $W$ are those the empirical distributions of the sequences $\{\boldsymbol{x}_0\}_{n\in\mathbb{N}}$ and $\{\boldsymbol{w}\}_{n\in\mathbb{N}}$ converge weakly to. We specify the choice of thresholds $\{\chi_t\}_{t\in\mathbb{N}}$:

$$\chi_t = \alpha\tau_t \tag{131}$$

in which $\alpha > 0$ is a pre-specified parameter.

Further, define the following (scalar) recursion for $\{R_{s,t}\}_{s,t\in\mathbb{N}}$:

$$R_{s,t} = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left[\left(\frac{1}{\gamma+1}\eta_{s-1}\left(X_0 + Z_{s-1}\right) - X_0\right)\left(\frac{1}{\gamma+1}\eta_{t-1}\left(X_0 + Z_{t-1}\right) - X_0\right)\right], \tag{132}$$

$$R_{0,t} = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left[(-X_0)\left(\frac{1}{\gamma+1}\eta_{t-1}\left(X_0 + Z_{t-1}\right) - X_0\right)\right], \tag{133}$$

$$R_{0,0} = \sigma^2 + \frac{1}{\delta}\mathsf{M}_2, \tag{134}$$

where $(Z_s, Z_t) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} R_{s,s} & R_{s,t} \\ R_{s,t} & R_{t,t} \end{bmatrix}\right)$, independent of $X_0$ and $W$. Note that $R_{t,t} = \tau_t^2$ for any $t \in \mathbb{N}$.

We have some useful convergence results concerning the AMP.

**Lemma 15.** *For any $t > 0$, almost surely,*

$$\lim_{n\to\infty} \psi_{\mathrm{av}}\left(\boldsymbol{x}^t, \boldsymbol{x}_0\right) = \mathbb{E}\left[\psi\left(\frac{1}{\gamma+1}\eta_{t-1}\left(X_0 + \tau_{t-1}Z\right), X_0\right)\right], \tag{135}$$

$$\lim_{m\to\infty} \frac{1}{m}\left\|\boldsymbol{z}^t\right\|_2^2 = \tau_t^2, \tag{136}$$

$$\lim_{n\to\infty} \mu^t = \mathbb{P}\left(|X_0 + \tau_t Z| \geq \alpha\tau_t\right), \tag{137}$$

$$\lim_{m\to\infty} \frac{1}{m}\left\|\bar{\boldsymbol{G}}\boldsymbol{x}^t - \bar{\boldsymbol{y}}\right\|_2^2 = \tau_t^2 - \sigma^2 + \frac{\varUpsilon_t^2\left(\tau_{t-1}^2 - \sigma^2\right)}{\delta^2(\gamma+1)^2} + \left(1 - \frac{\varUpsilon_t}{\delta(\gamma+1)}\right)^2\sigma^2 - 2\frac{\varUpsilon_t\left(R_{t,t-1} - \sigma^2\right)}{\delta(\gamma+1)}, \tag{138}$$

*where $\varUpsilon_t = \mathbb{P}\left(|X_0 + \tau_{t-1}Z| \geq \alpha\tau_{t-1}\right)$ and $\psi$ is pseudo-Lipschitz.*

The following lemma states that we can unambiguously use a sequence of Gaussian random variables $\{Z_t\}_{t\in\mathbb{N}}$ to describe $\{R_{s,t}\}_{s,t\in\mathbb{N}}$ in the recursion (132).

**Lemma 16.** *For any $k > 0$, the $k \times k$ matrix $\mathsf{R}_{k\times k} \equiv \{R_{s,t}\}_{0\leq s,t<k}$ is positive definite.*

The proofs of the above lemmas are deferred to Section 5.

### 4.2 Asymptotics of the AMP State Evolution

Recall Eq. (2), (3) and (4) in Section 1.2. It should be recognized that $\tau_*$ is the limit of $\tau_t$ as $t \to \infty$, should it exist. In other words, $\tau_*$ is the solution to the fixed-point equation

$$\tau^2 = \mathsf{F}\left(\tau^2, \alpha, \gamma\right) \tag{139}$$

in which

$$\mathsf{F}\left(\tau^2, \alpha, \gamma\right) = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left[\left(\frac{1}{\gamma+1}\eta_{\alpha\tau}\left(X_0 + \tau Z\right) - X_0\right)^2\right]. \tag{140}$$

Recall that the AMP iterates are determined by $\alpha$ and $\gamma$. To specify $\alpha$ and $\gamma$, we have to calibrate the AMP with the the parameters $\lambda$ and $\rho$ of the elastic net (1), via Eq. (3) and (4). For this purpose, they are called the *calibration equations.*

Let $\alpha_{\min} = \max\{0, \alpha^*_{\min}\}$, where $\alpha^*_{\min} = \alpha^*_{\min}(\gamma, \delta)$ is the unique solution $u$ to

$$\left(1+u^2\right)\Phi(-u) - u\phi(u) = \frac{\delta(\gamma+1)^2}{2} \tag{141}$$

Its uniqueness is easy to see as follows. Let $g(u) = (1+u^2)\Phi(-u) - u\phi(u)$. Since $g'(u) = 2u\Phi(-u) - 2\phi(u) < 0$, $g$ is decreasing, proving the uniqueness of $\alpha^*_{\min}$. Note that only when $\delta(\gamma+1)^2 \in (0,1)$ would $\alpha^*_{\min}$ be positive, which necessarily requires $\delta < 1$.

**Lemma 17.** *For any $\alpha > \alpha_{\min}$, $\tau_*$ exists and is unique. Furthermore, for all $t \geq 0$, $\max\{\tau_0, \tau_*\} \geq \tau_t \geq \min\{\tau_0, \tau_*\} > 0$.*

**Lemma 18.** *For each $\delta > 0$, $\lambda > 0$ and $\gamma > 0$, there exists $\alpha > \alpha_{\min}$ that satisfies Eq. (3).*

**Lemma 19.** *Fix $\delta > 0$ and $\lambda > 0$. There exist $\gamma_0 > 0$ and continuously differentiable $\gamma \mapsto \tau_*(\gamma)$ and $\gamma \mapsto \alpha(\gamma)$ defined on $(-\gamma_0, \infty)$, such that at each $\gamma \in [0, \infty)$, Eq. (2) and (3) are satisfied.*

**Lemma 20.** *Fix $\delta > 0$ and $\lambda > 0$. There exist $\gamma_0 > 0$ and continuously differentiable $\gamma \mapsto \rho(\gamma)$ defined on $(-\gamma_0, \infty)$, such that at each $\gamma \in [0, \infty)$, Eq. (2), (3) and (4) are satisfied, $\rho \to 0$ as $\gamma \to 0$, and $\rho \to \infty$ as $\gamma \to \infty$.*

These lemmas show that given $\lambda$ and $\rho$ from the elastic net problem (1), one can construct a corresponding AMP algorithm and computes its high-dimensional behavior at convergence, i.e. as $n \to \infty$ then $t \to \infty$. The proofs of these lemmas are deferred to Section 5.

### 4.3 Proof of Theorem 14

For brevity, let $\boldsymbol{x}^t = \boldsymbol{x}^t\left(\bar{\boldsymbol{G}}\right)$ be the $t$-th AMP iterate, and $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}\left(\bar{\boldsymbol{G}}\right)$ be a minimizer to the elastic net problem (1). First define

$$\boldsymbol{u}^t = \frac{1}{\alpha\tau_{t-1}}\left[\bar{\boldsymbol{G}}^T\boldsymbol{z}^{t-1} + \boldsymbol{x}^{t-1} - (\gamma+1)\boldsymbol{x}^t\right]. \tag{142}$$

From Eq. (117), we see that $\boldsymbol{u}^t \in \partial\left\|\boldsymbol{x}^t\right\|_1$, since $\boldsymbol{x} = \eta(\boldsymbol{r}; \chi)$ if and only if there exists $\boldsymbol{u} \in \partial\|\boldsymbol{x}\|_1$ such that $\boldsymbol{x} + \chi\boldsymbol{u} = \boldsymbol{r}$. In addition, define

$$\boldsymbol{v}^t = \bar{\boldsymbol{G}}^T\left[\bar{\boldsymbol{G}}\left(\boldsymbol{x}^t - \boldsymbol{x}_0\right) - \boldsymbol{w}\right] + \lambda\boldsymbol{u}^t + \rho\boldsymbol{x}^t. \tag{143}$$

In other words, $\boldsymbol{v}^t$ is a subgradient of the elastic net's objective function at $\boldsymbol{x}^t$.

The following lemma shows that the AMP iterates converges as $n \to \infty$ then $t \to \infty$.

**Lemma 21.** *Almost surely,*

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{x}^t-\boldsymbol{x}^{t-1}\right\|_2^2=0, \tag{144}$$

$$\lim_{t\to\infty}\lim_{m\to\infty}\frac{1}{m}\left\|\boldsymbol{z}^t-\boldsymbol{z}^{t-1}\right\|_2^2=0. \tag{145}$$

The next lemma shows that the AMP converges to a point where $\boldsymbol{v}^t$ vanishes. This is essentially a consequence of the calibration equations.

**Lemma 22.** *Almost surely,* $\lim_{t\to\infty}\lim_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{v}^t\right\|_2^2=0$.

The proofs of these two lemmas are deferred to Section 5. The following proposition shows that the AMP solves the elastic net in the limit $n\to\infty$ then $t\to\infty$. This is key to establishing Theorem 14.

**Proposition 23.** *Almost surely,* $\lim_{t\to\infty}\lim_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{x}^t-\hat{\boldsymbol{x}}\right\|_2^2=0$.

*Proof.* We have:

$$0\le\mathcal{C}\left(\boldsymbol{x}^t;0,\bar{\boldsymbol{G}}\right)-\mathcal{C}\left(\hat{\boldsymbol{x}};0,\bar{\boldsymbol{G}}\right) \tag{146}$$

$$=\left(\boldsymbol{x}^t-\hat{\boldsymbol{x}}\right)^T\boldsymbol{v}^t+\lambda\left[\left\|\boldsymbol{x}^t\right\|_1-\left\|\hat{\boldsymbol{x}}\right\|_1-\left(\boldsymbol{x}^t-\hat{\boldsymbol{x}}\right)^T\boldsymbol{u}^t\right]-\frac{1}{2}\left\|\bar{\boldsymbol{G}}\left(\boldsymbol{x}^t-\hat{\boldsymbol{x}}\right)\right\|_2^2-\frac{\rho}{2}\left\|\boldsymbol{x}^t-\hat{\boldsymbol{x}}\right\|_2^2 \tag{147}$$

$$\le\left\|\boldsymbol{x}^t-\hat{\boldsymbol{x}}\right\|_2\left\|\boldsymbol{v}^t\right\|_2-\frac{\rho}{2}\left\|\boldsymbol{x}^t-\hat{\boldsymbol{x}}\right\|_2^2 \tag{148}$$

which yields

$$\frac{\rho}{2}\left\|\boldsymbol{x}^t-\hat{\boldsymbol{x}}\right\|_2\le\left\|\boldsymbol{v}^t\right\|_2 \tag{149}$$

where we use the fact $\boldsymbol{u}^t\in\partial\left\|\boldsymbol{x}^t\right\|_1$ and convexity of $\|\cdot\|_1$. The proof is completed in light of Lemma 22. ☐

*Proof of Theorem 14.* We have

$$\left|\psi_{\rm av}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{G}}\right),\boldsymbol{x}_0\right)-\psi_{\rm av}\left(\boldsymbol{x}^t,\boldsymbol{x}_0\right)\right|\to 0 \tag{150}$$

in probability as $n\to\infty$ then $t\to\infty$, immediately from Proposition 23, Lemma 3, along with the facts that $\frac{1}{n}\left\|\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{G}}\right)\right\|_2^2\le\mathsf{T}$ with high probability from Lemma 5, $\lim_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{x}^t\right\|_2^2<\infty$ by Lemma 15, and $\frac{1}{n}\left\|\boldsymbol{x}_0\right\|_2^2$ is bounded eventually. Furthermore, Proposition 23 also implies that

$$\left|\frac{1}{n}\mathcal{C}\left(\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{G}}\right);0,\bar{\boldsymbol{G}}\right)-\frac{1}{n}\mathcal{C}\left(\boldsymbol{x}^t;0,\bar{\boldsymbol{G}}\right)\right|\to 0 \tag{151}$$

in probability as $n\to\infty$ then $t\to\infty$. Then from Lemma 15, we obtain:

$$\psi^*=\mathbb{E}\left[\psi\left(\frac{1}{\gamma+1}\eta_{\alpha\tau_*}\left(X_0+\tau_*Z\right),X_0\right)\right], \tag{152}$$

$$\mathcal{L}^*=\frac{\tau_*^2}{2}\left(1-\frac{\varUpsilon_*}{\delta\left(\gamma+1\right)}\right)^2+\frac{\lambda}{\gamma+1}\mathbb{E}\left[\eta_{\alpha\tau_*}\left(\left|X_0+\tau_*Z\right|\right)\right]+\frac{\rho}{2\left(\gamma+1\right)^2}\mathbb{E}\left[\eta_{\alpha\tau_*}^2\left(X_0+\tau_*Z\right)\right], \tag{153}$$

in which $\varUpsilon_*=\mathbb{P}\left(\left|X_0+\tau_*Z\right|\ge\alpha\tau_*\right)$. ☐

# 5 Proof of Auxiliary Lemmas

## 5.1 Proof of Lemma 3

Using the definition of pseudo-Lipschitz functions and the Cauchy-Schwarz inequality, we get

$$
\begin{aligned}
\left|\psi_{\mathrm{av}}(\boldsymbol{u},\boldsymbol{v})-\psi_{\mathrm{av}}(\boldsymbol{r},\boldsymbol{t})\right| &\le \frac{1}{n}\sum_{i=1}^{n}\left|\psi(u_i,v_i)-\psi(r_i,t_i)\right| && (154)\\
&\le \frac{L}{n}\sum_{i=1}^{n}\left(|u_i-r_i|+|v_i-t_i|\right)\left(1+|u_i|+|v_i|+|r_i|+|t_i|\right) && (155)\\
&\le \frac{L}{n}\left(\|\boldsymbol{u}-\boldsymbol{r}\|_2+\|\boldsymbol{v}-\boldsymbol{t}\|_2\right)\sqrt{\sum_{i=1}^{n}\left(1+|u_i|+|v_i|+|r_i|+|t_i|\right)^2} && (156)\\
&\le L\sqrt{5}\frac{\|\boldsymbol{u}-\boldsymbol{r}\|_2+\|\boldsymbol{v}-\boldsymbol{t}\|_2}{\sqrt{n}}\sqrt{1+\frac{\|\boldsymbol{u}\|_2^2}{n}+\frac{\|\boldsymbol{v}\|_2^2}{n}+\frac{\|\boldsymbol{r}\|_2^2}{n}+\frac{\|\boldsymbol{t}\|_2^2}{n}}. && (157)
\end{aligned}
$$

## 5.2 Proof of Lemma 4

Let $\mathbb{E}\left[\left|\sqrt{m}\bar{A}_{ij}\right|^{\mathfrak{p}}\right]=\mathsf{K}_{\mathfrak{p}}$ and recall $\mathsf{K}_{\mathfrak{p}}<\infty$. To show the first property, since $\mathbb{E}\left[\bar{A}_{ij}\right]=0$ and $m\mathbb{E}\left[\bar{A}_{ij}^2\right]=1$, we have:

$$
\begin{aligned}
m\mathbb{E}\left[\tilde{A}_{ij}^2\right] &= m\mathbb{E}\left[\bar{A}_{ij}^2\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|\le\mathsf{R}\right)\right]-m\left(\mathbb{E}\left[\bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|\le\mathsf{R}\right)\right]\right)^2 && (158)\\
&= 1-m\mathbb{E}\left[\bar{A}_{ij}^2\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)\right]-m\left(\mathbb{E}\left[\bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)\right]\right)^2. && (159)
\end{aligned}
$$

Note that

$$
m\mathbb{E}\left[\bar{A}_{ij}^2\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)\right]\le\frac{\mathbb{E}\left[m^2\bar{A}_{ij}^4\right]}{\mathsf{R}^2}\le\frac{\mathsf{K}_{\mathfrak{p}}^{4/\mathfrak{p}}}{\mathsf{R}^2}. \tag{160}
$$

Similarly, $m\left(\mathbb{E}\left[\bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)\right]\right)^2\to 0$. As such, $m\mathbb{E}\left[\tilde{A}_{ij}^2\right]=1+o_{\mathsf{R}}(1)$, where $o_{\mathsf{R}}(1)\to 0$ as $\mathsf{R}\to\infty$ and $o_{\mathsf{R}}(1)$ is independent of $n$. Combining this with the fact $\sqrt{m}\left|\tilde{A}_{ij}\right|\le 2\mathsf{R}$ with probability 1 proves the first property.

To prove the second property, since $\mathbb{E}\left[\bar{A}_{ij}\right]=0$, considering $\boldsymbol{B}=\bar{\boldsymbol{A}}-\tilde{\boldsymbol{A}}$, we then have:

$$
\begin{aligned}
\mathbb{E}\left[B_{ij}^4\right] &= \mathbb{E}\left[\left(\bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)-\mathbb{E}\left[\bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)\right]\right)^4\right] && (161)\\
&\le 8\mathbb{E}\left[\bar{A}_{ij}^4\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)\right]+8\left(\mathbb{E}\left[\bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)\right]\right)^4 && (162)\\
&\overset{(a)}{\le} 16\mathbb{E}\left[\bar{A}_{ij}^4\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right|>\mathsf{R}\right)\right] && (163)\\
&\le 16\frac{\mathbb{E}\left[\left|\sqrt{m}\bar{A}_{ij}\right|^{\mathfrak{p}}\right]}{m^2\mathsf{R}^{\mathfrak{p}-4}} && (164)\\
&= \frac{16\mathsf{K}_{\mathfrak{p}}}{m^2\mathsf{R}^{\mathfrak{p}-4}} && (165)
\end{aligned}
$$

where $(a)$ is by Jensen's inequality. Consequently,

$$\mathbb{E}\left[B_{ij}^2\right] \leq \sqrt{\mathbb{E}\left[B_{ij}^4\right]} \leq \frac{4\sqrt{\mathsf{K}_{\mathfrak{p}}}}{m\mathsf{R}^{(\mathfrak{p}-4)/2}}. \tag{166}$$

From [Lat05], for some universal constant $c$,

$$\mathbb{E}\left[\|\boldsymbol{B}\|_2\right] \leq c\left(\max_{i\in[m]}\sqrt{\sum_{j=1}^{n}\mathbb{E}\left[B_{ij}^2\right]} + \max_{j\in[n]}\sqrt{\sum_{i=1}^{m}\mathbb{E}\left[B_{ij}^2\right]} + \sqrt[4]{\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{E}\left[B_{ij}^4\right]}\right) = o_{\mathsf{R}}(1). \tag{167}$$

Furthermore, by Jensen's inequality,

$$\begin{aligned}
\mathbb{E}\left[m^2\tilde{A}_{ij}^4\right] &\leq 8\mathbb{E}\left[m^2\bar{A}_{ij}^4\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right| \leq \mathsf{R}\right)\right] + 8\left(\mathbb{E}\left[\sqrt{m}\bar{A}_{ij}\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right| \leq \mathsf{R}\right)\right]\right)^4 && (168)\\
&\leq 16\mathbb{E}\left[m^2\bar{A}_{ij}^4\mathbb{I}\left(\sqrt{m}\left|\bar{A}_{ij}\right| \leq \mathsf{R}\right)\right] && (169)\\
&\leq 16\mathbb{E}\left[m^2\bar{A}_{ij}^4\right] && (170)\\
&\leq 16\mathsf{K}_{\mathfrak{p}}^{4/\mathfrak{p}} && (171)
\end{aligned}$$

and hence again, $\mathbb{E}\left[\left\|\tilde{\boldsymbol{A}}\right\|_2\right]$ is bounded. By the triangular inequality,

$$\begin{aligned}
\mathbb{E}\left[\left\|\boldsymbol{A} - \bar{\boldsymbol{A}}\right\|_2\right] &\leq \mathbb{E}\left[\|\boldsymbol{B}\|_2\right] + \mathbb{E}\left[\left\|\boldsymbol{A} - \tilde{\boldsymbol{A}}\right\|_2\right] && (172)\\
&= o_{\mathsf{R}}(1) + \left|1 - \left(m\mathbb{E}\left[\tilde{A}_{ij}^2\right]\right)^{-1/2}\right|\mathbb{E}\left[\left\|\tilde{\boldsymbol{A}}\right\|_2\right] = o_{\mathsf{R}}(1). && (173)
\end{aligned}$$

The property then follows immediately from Markov's inequality.

The third property follows immediately from the first property and Theorem 26, recalling that $\mathbb{E}\left[A_{ij}\right] = 0$, $\mathbb{E}\left[mA_{ij}^2\right] = 1$, and

$$\mathbb{E}\left[m^2A_{ij}^4\right] = \frac{\mathbb{E}\left[m^2\tilde{A}_{ij}^4\right]}{\left(\mathbb{E}\left[m\tilde{A}_{ij}^2\right]\right)^2} = \frac{16\mathsf{K}_{\mathfrak{p}}^{4/\mathfrak{p}}}{\left(1 + o_{\mathsf{R}}(1)\right)^2} \tag{174}$$

which is bounded for sufficiently large $\mathsf{R}$.

### 5.3 Proof of Lemma 5

The proof is the same for both $\bar{\boldsymbol{A}}$ and $\boldsymbol{A}$. Let $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\boldsymbol{A})$ for brevity. Since the problem is convex, by the KKT condition, there exists $\boldsymbol{u} \in \partial\|\hat{\boldsymbol{x}}\|_1$ such that with high probability,

$$\begin{aligned}
\|\hat{\boldsymbol{x}}\|_2 &= \left\|\left[\boldsymbol{A}^T\boldsymbol{A} + \rho\boldsymbol{I}\right]^{-1}\left(-\lambda\boldsymbol{u} + \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{A}^T\boldsymbol{w}\right)\right\|_2 && (175)\\
&\leq \frac{1}{\rho}\left(\lambda\|\boldsymbol{u}\|_2 + \|\boldsymbol{A}\|_2^2\|\boldsymbol{x}_0\|_2 + \|\boldsymbol{A}\|_2\|\boldsymbol{w}\|_2\right) && (176)\\
&\leq \frac{1}{\rho}\left[\lambda\sqrt{n} + \left(1 + \frac{1}{\sqrt{\delta}}\right)^2\sqrt{n\mathsf{M}_2} + \left(1 + \frac{1}{\sqrt{\delta}}\right)\sigma\sqrt{m} + 0.01\sqrt{n}\right] && (177)
\end{aligned}$$

where the last step is by the third property in Lemma 4, and the fact that $|u_i| \leq 1$ for all $i \in [n]$. This yields a lower bound on $\mathsf{T}$, and we choose $\mathsf{T}$ to be the maximum between this bound and $100\mathsf{M}_2$.

### 5.4 Proof of Lemmas 15 and 16

We state a useful convergence result concerning the AMP iterates (124)-(127).

**Proposition 24.** *Let $\psi$ be a pseudo-Lipschitz function. Consider the AMP iterates (124)-(127). Then for all $s, t \in \mathbb{N}$,*

$$\lim_{n\to\infty} \psi_{\mathrm{av}}\left(\boldsymbol{h}^{t+1}, \boldsymbol{x}_0\right) = \mathbb{E}\left[\psi\left(Z_t, X_0\right)\right], \tag{178}$$

$$\lim_{m\to\infty} \frac{1}{m}\left\langle \boldsymbol{b}^t, \boldsymbol{b}^s \right\rangle = R_{s,t} - \sigma^2, \tag{179}$$

$$\lim_{m\to\infty} \frac{1}{m}\left\langle \boldsymbol{b}^t, \boldsymbol{w} \right\rangle = 0, \tag{180}$$

*almost surely.*

*Proof.* The first and last equations are immediate from Lemma 27.1. To see the second equation, note that almost surely,

$$\lim_{n\to\infty} \frac{1}{m}\left\langle \boldsymbol{b}^t, \boldsymbol{b}^s \right\rangle \overset{(a)}{=} \lim_{n\to\infty} \frac{1}{m}\left\langle \boldsymbol{m}^t, \boldsymbol{m}^s \right\rangle + \lim_{n\to\infty} \frac{1}{m}\left\langle \boldsymbol{b}^t + \boldsymbol{b}^s, \boldsymbol{w} \right\rangle - \sigma^2 \tag{181}$$

$$\overset{(b)}{=} \lim_{n\to\infty} \frac{1}{n}\left\langle \boldsymbol{h}^{t+1}, \boldsymbol{h}^{s+1} \right\rangle - \sigma^2 \tag{182}$$

where $(a)$ is by Eq. (125) and (126), and $(b)$ is by Lemma 27. One can prove by induction that $\lim_{n\to\infty} \frac{1}{n}\left\langle \boldsymbol{h}^{t+1}, \boldsymbol{h}^{s+1} \right\rangle = R_{s,t}$ almost surely. This is similar to [BM12, Theorem 4.2] and hence omitted. □

*Proof of Lemma 15.* The first equation is a direct application of Proposition 24 to a pseudo-Lipschitz mapping that maps $\left(h_i^{t+1}, x_{0,i}\right)$ to $\psi\left(\frac{1}{\gamma+1}\eta_{t-1}\left(x_{0,i} - h_i^{t+1}\right), x_{0,i}\right)$. To see the second equation, notice that

$$\lim_{m\to\infty} \frac{1}{m}\left\|\boldsymbol{z}^t\right\|_2^2 = \lim_{m\to\infty} \frac{1}{m}\left\|\boldsymbol{m}^t\right\|_2^2 = \lim_{n\to\infty} \frac{1}{n}\left\|\boldsymbol{h}^t\right\|_2^2 \tag{183}$$

almost surely by Lemma 27.2, and the result follows from Proposition 24 applied to $\left(h_i^{t+1}, x_{0,i}\right) \mapsto \left(h_i^{t+1}\right)^2$.

To see the third equation, note that Proposition 24 implies, almost surely, the empirical distribution of $\left\{x_{0,i} - h_i^{t+1}\right\}_{i\in[n]}$ converges weakly to the distribution of $X_0 + \tau_t Z$, which admits a density since $\tau_t \geq \sigma > 0$. As such, almost surely,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(\left|x_{0,i} - h_i^{t+1}\right| \geq \alpha\tau_t\right) = \mathbb{P}\left(\left|X_0 + \tau_t Z\right| \geq \alpha\tau_t\right) \tag{184}$$

which yields the third equation. Finally, almost surely,

$$\lim_{m\to\infty}\frac{1}{m}\left\|\bar{\boldsymbol{G}}\boldsymbol{x}^t-\bar{\boldsymbol{y}}\right\|_2^2 \overset{(a)}{=} \lim_{m\to\infty}\frac{1}{m}\left\|\boldsymbol{z}^t-\frac{1}{\delta(\gamma+1)}\Upsilon_t\boldsymbol{z}^{t-1}\right\|_2^2 \tag{185}$$

$$= \lim_{m\to\infty}\frac{1}{m}\left\|\boldsymbol{b}^t-\frac{1}{\delta(\gamma+1)}\Upsilon_t\boldsymbol{b}^{t-1}-\left(1-\frac{1}{\delta(\gamma+1)}\Upsilon_t\right)\boldsymbol{w}\right\|_2^2 \tag{186}$$

$$\begin{aligned}= &\lim_{m\to\infty}\frac{1}{m}\left\|\boldsymbol{b}^t\right\|_2^2+\frac{\Upsilon_t^2}{\delta^2(\gamma+1)^2}\lim_{m\to\infty}\frac{1}{m}\left\|\boldsymbol{b}^{t-1}\right\|_2^2+\left(1-\frac{\Upsilon_t}{\delta(\gamma+1)}\right)^2\sigma^2\\ &-2\frac{\Upsilon_t}{\delta(\gamma+1)}\lim_{m\to\infty}\left\langle\boldsymbol{b}^t,\boldsymbol{b}^{t-1}\right\rangle-2\left(1-\frac{\Upsilon_t}{\delta(\gamma+1)}\right)\lim_{m\to\infty}\frac{1}{m}\left\langle\boldsymbol{b}^t,\boldsymbol{w}\right\rangle\\ &+2\frac{\Upsilon_t}{\delta(\gamma+1)}\left(1-\frac{\Upsilon_t}{\delta(\gamma+1)}\right)\lim_{m\to\infty}\frac{1}{m}\left\langle\boldsymbol{b}^{t-1},\boldsymbol{w}\right\rangle\end{aligned} \tag{187}$$

$$\overset{(b)}{=} \tau_t^2-\sigma^2+\frac{\Upsilon_t^2\left(\tau_{t-1}^2-\sigma^2\right)}{\delta^2(\gamma+1)^2}+\left(1-\frac{\Upsilon_t}{\delta(\gamma+1)}\right)^2\sigma^2-2\frac{\Upsilon_t\left(R_{t,t-1}-\sigma^2\right)}{\delta(\gamma+1)} \tag{188}$$

where $(a)$ is by the third equation and Eq. (118), and $(b)$ is by Proposition 24. □

*Proof of Lemma 16.* By Proposition 24, almost surely,

$$R_{s,t}=\lim_{m\to\infty}\frac{1}{m}\left\langle\boldsymbol{b}^t,\boldsymbol{b}^s\right\rangle+\sigma^2=\lim_{m\to\infty}\frac{1}{m}\left\langle\boldsymbol{m}^t,\boldsymbol{m}^s\right\rangle. \tag{189}$$

The thesis follows from [BM12, Lemma 5.2]. □

## 5.5 Proof of Lemma 17

We have:

$$\begin{aligned}\delta(\gamma+1)^2\frac{\partial\mathsf{F}}{\partial(\tau^2)}= &\left(1+\alpha^2\right)\mathbb{E}\left[\Phi\left(\frac{X_0}{\tau}-\alpha\right)+\Phi\left(-\frac{X_0}{\tau}-\alpha\right)\right]\\ &-\mathbb{E}\left[\left(\frac{X_0}{\tau}+\alpha\right)\phi\left(\frac{X_0}{\tau}-\alpha\right)-\left(\frac{X_0}{\tau}-\alpha\right)\phi\left(\frac{X_0}{\tau}+\alpha\right)\right]\\ &+\frac{\gamma}{\tau}\mathbb{E}\left[-X_0\phi\left(\frac{X_0}{\tau}-\alpha\right)+X_0\phi\left(\frac{X_0}{\tau}+\alpha\right)\right]\\ &+\alpha\frac{\gamma}{\tau}\mathbb{E}\left[X_0\Phi\left(\frac{X_0}{\tau}-\alpha\right)-X_0\Phi\left(-\frac{X_0}{\tau}-\alpha\right)\right]\end{aligned} \tag{190}$$

which then yields

$$\begin{aligned}\delta(\gamma+1)^2\frac{\partial^2\mathsf{F}}{\partial(\tau^2)^2}= &-\frac{1+\gamma}{2\tau^2}\mathbb{E}\left[\frac{X_0^3}{\tau^3}\left(\phi\left(\frac{X_0}{\tau}-\alpha\right)-\phi\left(\frac{X_0}{\tau}+\alpha\right)\right)\right]\\ &-\frac{\gamma}{2\tau^3}\mathbb{E}\left[\alpha X_0\Phi\left(\frac{X_0}{\tau}-\alpha\right)-X_0\phi\left(\frac{X_0}{\tau}-\alpha\right)\right]\\ &+\frac{\gamma}{2\tau^3}\mathbb{E}\left[\alpha X_0\Phi\left(-\frac{X_0}{\tau}-\alpha\right)-X_0\phi\left(\frac{X_0}{\tau}+\alpha\right)\right].\end{aligned} \tag{191}$$

We have $u^3 \left[\phi(u-\alpha) - \phi(-u-\alpha)\right] \geq 0$ for any $u$ and $\alpha \geq 0$, since $\phi(u-\alpha) \geq \phi(-u-\alpha)$ if $u \geq 0$ and $\phi(u-\alpha) \leq \phi(-u-\alpha)$ if $u \leq 0$. Also, letting $f(u) = \alpha\Phi(u-\alpha) - \phi(u-\alpha)$, we have

$$f(u) - f(-u) = \int_{-u}^{u} f'(t)\, dt = \int_{0}^{u} \left(t\phi(t-\alpha) - t\phi(-t-\alpha)\right) dt. \tag{192}$$

From this, it is easy to see that for $\alpha \geq 0$, with $u \geq 0$, $f(u) \geq f(-u)$, and with $u \leq 0$, $f(u) \leq f(-u)$, which implies $uf(u) - uf(-u) \geq 0$ for any $u$. Together with the fact $\gamma > 0$, we have $\frac{\partial^2 \mathsf{F}}{\partial(\tau^2)^2} \leq 0$, i.e. $\mathsf{F}$ is concave in $\tau^2$.

Next it is easy to see that

$$\lim_{\tau\to\infty} \frac{\partial \mathsf{F}}{\partial(\tau^2)} = \frac{2}{\delta(\gamma+1)^2}\left[\left(1+\alpha^2\right)\Phi(-\alpha) - \alpha\phi(\alpha)\right]. \tag{193}$$

Let $g(\alpha) = \left(1+\alpha^2\right)\Phi(-\alpha) - \alpha\phi(\alpha)$. Since $g'(\alpha) = 2\alpha\Phi(-\alpha) - 2\phi(\alpha) < 0$, $g$ is decreasing. Since $\alpha > \alpha^*_{\min}$ and recalling that $g(\alpha^*_{\min}) = \frac{1}{2}\delta(\gamma+1)^2$, we then have

$$0 < \lim_{\tau\to\infty} \frac{\partial \mathsf{F}}{\partial(\tau^2)} < 1. \tag{194}$$

Combining with the proven concavity, we have $\mathsf{F}$ is increasing in $\tau^2$. Thereby $\tau_*$ exists uniquely and $\max\{\tau_0, \tau_*\} \geq \tau_t \geq \min\{\tau_0, \tau_*\}$ for all $t \geq 0$. Also, $\min\{\tau_0, \tau_*\} \geq \sigma > 0$.

## 5.6 Proof of Lemma 18

We claim the following:

- $\alpha \mapsto \tau_*^2(\alpha)$ is continuously differentiable on $(\alpha_{\min}, \infty)$, and therefore so is $\alpha \mapsto \lambda(\alpha)$,
- $\lambda(\alpha) \to +\infty$ as $\alpha \to \infty$,
- if $\delta < 1$ and $\alpha^*_{\min} > 0$, $\lambda(\alpha) \to -\infty$ as $\alpha \downarrow \alpha^*_{\min}$,
- if $\delta < 1$ and $\alpha^*_{\min} < 0$, $\lambda(\alpha) \to 0$ as $\alpha \downarrow 0$,
- if $\delta \geq 1$, $\lambda(\alpha) \to 0$ as $\alpha \downarrow 0$,
- if $\delta < 1$ and $\alpha^*_{\min} = 0$, $\limsup_{\alpha\downarrow 0} \lambda(\alpha) \leq 0$.

where $\alpha \mapsto \lambda(\alpha)$ is defined via Eq. (3). The thesis follows from these claims.

To see the first claim, note that $\alpha \mapsto \tau_*^2(\alpha)$ is a well-defined mapping by uniqueness of $\tau_*$ for each $\alpha$, from Lemma 17. As shown in the proof of Lemma 17, we have that $\tau^2 \mapsto \mathsf{F}(\tau^2, \alpha, \gamma)$ is concave and increasing, and $0 < \lim_{\tau\to\infty} \frac{\partial \mathsf{F}}{\partial(\tau^2)} < 1$ for any $\alpha > 0$. Combining with the fact that $\tau_* > 0$ by Lemma 17, we have

$$0 < \frac{\partial \mathsf{F}}{\partial(\tau^2)}\left(\tau_*^2, \alpha, \gamma\right) < 1 \tag{195}$$

for any $\alpha > 0$. Also, $(\tau^2, \alpha) \mapsto \mathsf{F}(\tau^2, \alpha, \gamma)$ is continuously differentiable. Then the implicit function theorem applied to $(\tau^2, \alpha) \mapsto \tau^2 - \mathsf{F}(\tau^2, \alpha, \gamma)$ shows continuous differentiability of $\alpha \mapsto \tau_*^2(\alpha)$ on $(0, \infty)$ and hence on $(\alpha_{\min}, \infty)$.

The second claim is immediate: as $\alpha \to \infty$, we have $\tau_*^2 \to \sigma^2 + \frac{1}{\delta}\mathsf{M}_2 > 0$ and consequently $\mathbb{P}(|X_0 + \tau_* Z| \geq \alpha\tau_*) \to 0$. It is then easy to see that $\lambda(\alpha) \to +\infty$ in this limit.

We show the third claim. We first show that $\tau_* \to +\infty$ as $\alpha \downarrow \alpha^*_{\min}$. Indeed, by concavity of $\tau^2 \mapsto \mathsf{F}(\tau^2, \alpha, \gamma)$,

$$\tau_*^2 \geq \mathsf{F}(0, \alpha, \gamma) + \tau_*^2 \lim_{\tau\to\infty} \frac{\partial \mathsf{F}}{\partial (\tau^2)} \tag{196}$$

$$= \sigma^2 + \frac{\gamma^2 \mathsf{M}_2}{\delta(\gamma+1)^2} + \tau_*^2 \lim_{\tau\to\infty} \frac{\partial \mathsf{F}}{\partial (\tau^2)}. \tag{197}$$

It was proven in the proof of Lemma 17 that as $\alpha \downarrow \alpha^*_{\min}$, we have $\lim_{\tau\to\infty} \frac{\partial \mathsf{F}}{\partial(\tau^2)} \uparrow 1$. Since $\sigma > 0$, we have $\tau_* \to +\infty$. We thus have in the limit $\alpha \downarrow \alpha^*_{\min}$, $\mathbb{P}(|X_0 + \tau_* Z| \geq \alpha\tau_*) \to 2\Phi(-\alpha^*_{\min})$. Using the definition of $\alpha^*_{\min}$,

$$\frac{\delta(\gamma+1)^2}{2} = \left(1 + (\alpha^*_{\min})^2\right)\Phi(-\alpha^*_{\min}) - \alpha^*_{\min}\phi(\alpha^*_{\min}) \leq \frac{\phi(\alpha^*_{\min})}{\alpha^*_{\min}}, \tag{198}$$

we then have

$$\lim_{\alpha\downarrow\alpha^*_{\min}} \lambda(\alpha) = \lim_{\alpha\downarrow\alpha^*_{\min}} \alpha^*_{\min}\tau_* \left(1 - \frac{2\Phi(-\alpha^*_{\min})}{\delta(\gamma+1)}\right) \tag{199}$$

$$= \lim_{\alpha\downarrow\alpha^*_{\min}} \alpha^*_{\min}\tau_* \left[1 - \frac{2}{\delta\left(1 + (\alpha^*_{\min})^2\right)}\left(\frac{\delta(\gamma+1)}{2} + \frac{\alpha^*_{\min}\phi(\alpha^*_{\min})}{\gamma+1}\right)\right] \tag{200}$$

$$\leq \lim_{\alpha\downarrow\alpha^*_{\min}} \alpha^*_{\min}\tau_* \left[1 - (\gamma+1)\right] = -\infty \tag{201}$$

since $\gamma > 0$. This completes the proof of the third claim.

We show the fourth and fifth claims. The solution to the fixed-point equation $\tau^2 = \mathsf{F}(\tau^2, 0, \gamma)$ satisfies

$$\tau^2 = \sigma^2 + \frac{\gamma^2\mathsf{M}_2}{\delta(\gamma+1)^2} + \frac{\tau^2}{\delta(\gamma+1)^2}. \tag{202}$$

In the fourth claim, $\alpha^*_{\min} < 0$ , which implies $\delta(\gamma+1)^2 > 1$. In the fifth claim, since $\delta \geq 1$ and $\gamma > 0$, we also have $\delta(\gamma+1)^2 > 1$. Then as $\alpha \downarrow 0$,

$$\tau_* \to \left(1 - \frac{1}{\delta(\gamma+1)^2}\right)^{-1}\left(\sigma^2 + \frac{\gamma^2\mathsf{M}_2}{\delta(\gamma+1)^2}\right) < \infty$$

by continuity of $\alpha \mapsto \tau_*^2(\alpha)$, shown in the first claim, and since $\tau_* > 0$ by Lemma 17. As such, $\lambda(\alpha) \to 0$.

To see the sixth claim, a similar argument to the third claim shows that $\tau_* \to \infty$ as $\alpha \downarrow 0$, since $\alpha^*_{\min} = 0$. Furthermore $\alpha^*_{\min} = 0$ implies $\delta(\gamma+1)^2 = 1$. As such,

$$\limsup_{\alpha\downarrow 0} \lambda(\alpha) = \left(1 - \frac{1}{\sqrt{\delta}}\right)\limsup_{\alpha\downarrow 0}(\alpha\tau_*). \tag{203}$$

For $\delta < 1$, this shows $\limsup_{\alpha\downarrow 0} \lambda(\alpha) \leq 0$.

### 5.7 Proof of Lemma 19

Let us define

$$D_1 = \mathbb{E}\left[\Phi\left(\frac{X_0}{\tau_*} - \alpha\right) + \Phi\left(-\frac{X_0}{\tau_*} - \alpha\right)\right], \tag{204}$$

$$D_2 = \mathbb{E}\left[\phi\left(\frac{X_0}{\tau_*} - \alpha\right) + \phi\left(\frac{X_0}{\tau_*} + \alpha\right)\right], \tag{205}$$

$$D_3 = \mathbb{E}\left[\frac{X_0}{\tau_*}\phi\left(\frac{X_0}{\tau_*} - \alpha\right) - \frac{X_0}{\tau_*}\phi\left(\frac{X_0}{\tau_*} + \alpha\right)\right], \tag{206}$$

$$D_4 = \mathbb{E}\left[\frac{X_0}{\tau_*}\Phi\left(\frac{X_0}{\tau_*} - \alpha\right) - \frac{X_0}{\tau_*}\Phi\left(-\frac{X_0}{\tau_*} - \alpha\right)\right], \tag{207}$$

and rewrite Eq. (3) as $f\left(\tau_*^2, \alpha, \gamma\right) = 0$ in which

$$f\left(\tau^2, \alpha, \gamma\right) = \frac{\lambda}{\alpha\tau} + \frac{1}{\delta(\gamma+1)}\mathbb{P}\left(|X_0 + \tau Z| \geq \alpha\tau\right) - 1. \tag{208}$$

Note that $D_1, D_2 > 0$ and $D_3, D_4 \geq 0$. To see why $D_3 \geq 0$, we have for $\alpha > 0$, $\phi(u-\alpha) > \phi(-u-\alpha)$ if $u > 0$ and $\phi(u-\alpha) < \phi(-u-\alpha)$ if $u < 0$. And to see why $D_4 \geq 0$, consider $g(u) = u\Phi(u-\alpha) - u\Phi(-u-\alpha)$. We have $g'(u) = u\phi(u-\alpha) + u\phi(u+\alpha) + \Phi(u-\alpha) - \Phi(-u-\alpha)$. When $u < 0$, $g'(u) < 0$, and when $u \geq 0$, $g'(u) \geq 0$. Then since $g(0) = 0$, we have $g(u) \geq 0$ for any $u$.

We have:

$$\delta(\gamma+1)^2 \frac{\partial \mathsf{F}}{\partial \alpha}\left(\tau_*^2, \alpha, \gamma\right) = 2\alpha\tau_*^2 D_1 - 2\tau_*^2 D_2 + 2\gamma\tau_*^2 D_4, \tag{209}$$

$$\delta(\gamma+1)^2 \frac{\partial \mathsf{F}}{\partial(\tau^2)}\left(\tau_*^2, \alpha, \gamma\right) = \left(1+\alpha^2\right) D_1 - D_3 - \alpha D_2 - \gamma D_3 + \alpha\gamma D_4, \tag{210}$$

$$\frac{\partial f}{\partial(\tau^2)}\left(\tau_*^2, \alpha, \gamma\right) = -\frac{\lambda}{2\alpha\tau_*^3} - \frac{1}{2\delta(\gamma+1)\tau_*^2} D_3, \tag{211}$$

$$\frac{\partial f}{\partial \alpha}\left(\tau_*^2, \alpha, \gamma\right) = -\frac{\lambda}{\alpha^2\tau_*} - \frac{1}{\delta(\gamma+1)} D_2. \tag{212}$$

Consider the following:

$$\Delta = \delta(\gamma+1)^2 \left\{\frac{\partial f}{\partial(\tau^2)}\left(\tau_*^2, \alpha, \gamma\right)\left[-\frac{\partial \mathsf{F}}{\partial \alpha}\left(\tau_*^2, \alpha, \gamma\right)\right] - \frac{\partial f}{\partial \alpha}\left(\tau_*^2, \alpha, \gamma\right)\left[1 - \frac{\partial \mathsf{F}}{\partial(\tau^2)}\left(\tau_*^2, \alpha, \gamma\right)\right]\right\}. \tag{213}$$

We claim that $\Delta > 0$ for any $\gamma \geq 0$. The thesis then follows from the implicit function theorem.

To show the claim, we first consider $\gamma > 0$. Note that from the proof of Lemma 18, $0 < \frac{\partial \mathsf{F}}{\partial(\tau^2)}\left(\tau_*^2, \alpha, \gamma\right) < 1$. Also note that $\frac{\partial f}{\partial \alpha}\left(\tau_*^2, \alpha, \gamma\right) < 0$ and $\frac{\partial f}{\partial(\tau^2)}\left(\tau_*^2, \alpha, \gamma\right) \leq 0$. In the case $\alpha D_1 - D_2 + \gamma D_4 \geq 0$, we have $\frac{\partial \mathsf{F}}{\partial \alpha}\left(\tau_*^2, \alpha, \gamma\right) \geq 0$, which then leads to $\Delta > 0$. Consider the case $\alpha D_1 - D_2 + \gamma D_4 < 0$. We express $\Delta$ as

$$\Delta = \frac{\lambda}{\alpha^2\tau_*}\left[\delta(\gamma+1)^2 - D_1 + \gamma D_3\right] + \frac{1}{\delta(\gamma+1)}\left[D_3\left(\alpha D_1 + \gamma D_4\right) + D_2\left(\delta(\gamma+1)^2 - D_1 + \gamma D_3 - \alpha\left(\alpha D_1 - D_2 + \gamma D_4\right)\right)\right]. \tag{214}$$

Note that since $\lambda > 0$, we have $D_1 < \delta(\gamma+1)^2$. Then it is easy to see that $\Delta > 0$ in this case.

When $\gamma = 0$, the state evolution is then reduced to the one defined in [BM12]. We still have that there exists $\alpha > \alpha_{\min}$ that satisfies Eq. (3), similar to Lemma 18, and $0 < \frac{\partial \mathsf{F}}{\partial(\tau^2)}\left(\tau_*^2, \alpha, 0\right) < 1$. The above argument for the case $\gamma > 0$ is then applicable.

## 5.8 Proof of Lemma 20

The existence of continuously differentiable $\gamma \mapsto \rho(\gamma)$ is immediate from Lemma 19.

Consider $\gamma = 0$. The state evolution is then reduced to the one defined in [BM12]. We note some relevant facts in this case (which can also be seen from the arguments in the proof of Lemma 19):

- $\alpha \mapsto \tau_*^2(\alpha)$ is continuously differentiable on $(\alpha_{\min}, \infty)$, and therefore so is $\alpha \mapsto \lambda(\alpha)$,
- $\lambda(\alpha) \to +\infty$ and $\tau_*$ tends to a finite non-zero constant as $\alpha \to \infty$,
- $\tau_* \to +\infty$ as $\alpha \downarrow \alpha_{\min}^*$,
- if $\delta < 1$ (hence $\alpha_{\min}^* > 0$), $\lambda(\alpha) \to -\infty$ as $\alpha \downarrow \alpha_{\min}^*$,
- if $\delta > 1$ (hence $\alpha_{\min}^* < 0$), $\lambda(\alpha) \to 0$ as $\alpha \downarrow 0$, and $\tau_*$ tends to a finite non-zero constant as $\alpha \downarrow \alpha_{\min}^*$,
- if $\delta = 1$ (hence $\alpha_{\min}^* = 0$), $\limsup_{\alpha \downarrow 0} \lambda(\alpha) \le 0$.

Furthermore the mapping $\lambda \mapsto \alpha(\lambda)$ is well-defined [BM12, Corollary 1.7]. Therefore, with a given $\lambda > 0$, $\alpha$ and $\tau_*$ are non-zero and finite, such that $\alpha > \alpha_{\min}$, when $\gamma = 0$. Then $\rho \to 0$ in the limit $\gamma \to 0$.

Consider $\gamma \to \infty$. In this limit, $\mathsf{F}\left(\tau^2, \alpha, \gamma\right) \to \sigma^2 + \frac{1}{\delta}\mathsf{M}_2$, and hence $\tau_*^2 \to \sigma^2 + \frac{1}{\delta}\mathsf{M}_2$ non-zero and finite. We also see from Eq. (3) that $\alpha\tau_* \to \lambda$, and consequently, $\alpha$ tends to a finite constant. Therefore $\rho \to \infty$.

## 5.9 Proof of Lemma 21

First, notice that $\boldsymbol{b}^t - \boldsymbol{b}^{t-1} = \boldsymbol{z}^t - \boldsymbol{z}^{t-1}$ and $\boldsymbol{q}^t - \boldsymbol{q}^{t-1} = \boldsymbol{x}^t - \boldsymbol{x}^{t-1}$. By Lemma 27.2,

$$\lim_{n\to\infty} \frac{1}{n\delta} \left\| \boldsymbol{q}^t - \boldsymbol{q}^{t-1} \right\|_2^2 = \lim_{m\to\infty} \frac{1}{m} \left\| \boldsymbol{b}^t - \boldsymbol{b}^{t-1} \right\|_2^2 \tag{215}$$

almost surely. By Proposition 24, almost surely,

$$\lim_{n\to\infty} \frac{1}{n\delta} \left\| \boldsymbol{b}^t - \boldsymbol{b}^{t-1} \right\|_2^2 = \left(R_{t,t} - \sigma^2\right) + \left(R_{t-1,t-1} - \sigma^2\right) - 2\left(R_{t,t-1} - \sigma^2\right) = R_{t,t} + R_{t-1,t-1} - 2R_{t,t-1}. \tag{216}$$

A simple modification of the proof of [BM12, Lemma 5.7] yields that $R_{t,t} + R_{t-1,t-1} - 2R_{t,t-1} \to 0$ as $t \to \infty$. This completes the proof.

### 5.10 Proof of Lemma 22

Since $\lambda > 0$, by Eq. (3) and the fact that

$$\lim_{t\to\infty}\lim_{n\to\infty}\mu^t = \mathbb{P}\left(|X_0 + \tau_* Z| \geq \alpha\tau_*\right) \tag{217}$$

from Lemma 15, for sufficiently large $n$ and $t$, we have $1 - \frac{1}{\delta(\gamma+1)}\mu^{t-1} > 0$. Consequently,

$$\begin{aligned}
\boldsymbol{v}^t &\overset{(a)}{=} \frac{\lambda}{\alpha\tau_{t-1}}\left[\bar{\boldsymbol{G}}^T\boldsymbol{z}^{t-1} + \boldsymbol{x}^{t-1} - (\gamma+1)\,\boldsymbol{x}^t\right] + \bar{\boldsymbol{G}}^T\bar{\boldsymbol{G}}\left(\boldsymbol{x}^t - \boldsymbol{x}_0\right) - \bar{\boldsymbol{G}}^T\boldsymbol{w} + \frac{\lambda\gamma}{\alpha\tau_*}\boldsymbol{x}^t && (218)\\
&\overset{(b)}{=} \frac{\lambda}{\alpha\tau_{t-1}}\left(1 - \frac{1}{\delta(\gamma+1)}\mu^{t-1}\right)^{-1}\bar{\boldsymbol{G}}^T\left(\bar{\boldsymbol{G}}\boldsymbol{x}_0 + \boldsymbol{w} - \bar{\boldsymbol{G}}\boldsymbol{x}^{t-1} + \frac{\mu^{t-2}\boldsymbol{z}^{t-2} - \mu^{t-1}\boldsymbol{z}^{t-1}}{\delta(\gamma+1)}\right)\\
&\quad + \frac{\lambda}{\alpha\tau_{t-1}}\left[\boldsymbol{x}^{t-1} - (\gamma+1)\,\boldsymbol{x}^t\right] + \bar{\boldsymbol{G}}^T\bar{\boldsymbol{G}}\left(\boldsymbol{x}^t - \boldsymbol{x}_0\right) - \bar{\boldsymbol{G}}^T\boldsymbol{w} + \frac{\lambda\gamma}{\alpha\tau_*}\boldsymbol{x}^t && (219)\\
&= \left[\frac{\lambda}{\alpha\tau_{t-1}}\left(1 - \frac{1}{\delta(\gamma+1)}\mu^{t-1}\right)^{-1} - 1\right]\bar{\boldsymbol{G}}^T\left(\bar{\boldsymbol{G}}\boldsymbol{x}_0 + \boldsymbol{w} - \bar{\boldsymbol{G}}\boldsymbol{x}^{t-1}\right)\\
&\quad + \frac{\lambda}{\alpha\tau_{t-1}\delta(\gamma+1)}\left(1 - \frac{\mu^{t-1}}{\delta(\gamma+1)}\right)^{-1}\bar{\boldsymbol{G}}^T\left[\left(\mu^{t-2} - \mu^{t-1}\right)\boldsymbol{z}^{t-2} + \mu^{t-1}\left(\boldsymbol{z}^{t-2} - \boldsymbol{z}^{t-1}\right)\right]\\
&\quad + \left(\frac{\lambda}{\alpha\tau_{t-1}}\boldsymbol{I} + \bar{\boldsymbol{G}}^T\bar{\boldsymbol{G}}\right)\left(\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right) + \left(\frac{\lambda\gamma}{\alpha\tau_*} - \frac{\lambda\gamma}{\alpha\tau_{t-1}}\right)\boldsymbol{x}^t && (220)
\end{aligned}$$

where $(a)$ is from Eq. (142) and (4), and $(b)$ is from Eq. (118). Therefore,

$$\begin{aligned}
\left\|\boldsymbol{v}^t\right\|_2 &\leq \left|\frac{\lambda}{\alpha\tau_{t-1}}\left(1 - \frac{1}{\delta(\gamma+1)}\mu^{t-1}\right)^{-1} - 1\right|\left\|\bar{\boldsymbol{G}}^T\left(\bar{\boldsymbol{G}}\boldsymbol{x}_0 + \boldsymbol{w} - \bar{\boldsymbol{G}}\boldsymbol{x}^{t-1}\right)\right\|_2\\
&\quad + \frac{\lambda}{\alpha\tau_{t-1}\delta(\gamma+1)}\left(1 - \frac{\mu^{t-1}}{\delta(\gamma+1)}\right)^{-1}\left\|\bar{\boldsymbol{G}}\right\|_2\left(\left|\mu^{t-2} - \mu^{t-1}\right|\left\|\boldsymbol{z}^{t-2}\right\|_2 + \mu^{t-1}\left\|\boldsymbol{z}^{t-2} - \boldsymbol{z}^{t-1}\right\|_2\right)\\
&\quad + \left\|\frac{\lambda}{\alpha\tau_{t-1}}\boldsymbol{I} + \bar{\boldsymbol{G}}^T\bar{\boldsymbol{G}}\right\|_2\left\|\boldsymbol{x}^t - \boldsymbol{x}^{t-1}\right\|_2 + \frac{\lambda\gamma}{\alpha}\left|\frac{1}{\tau_*} - \frac{1}{\tau_{t-1}}\right|\left\|\boldsymbol{x}^t\right\|_2. && (221)
\end{aligned}$$

Note that almost surely, by Theorem 26 and Lemma 15,

$$\lim_{n\to\infty}\left\|\frac{\lambda}{\alpha\tau_{t-1}}\boldsymbol{I} + \bar{\boldsymbol{G}}^T\bar{\boldsymbol{G}}\right\|_2 \leq \frac{\lambda}{\alpha\tau_{t-1}} + \lim_{n\to\infty}\left\|\bar{\boldsymbol{G}}\right\|_2^2 < \infty, \tag{222}$$

$$\lim_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{x}^t\right\|_2^2 < \infty, \tag{223}$$

$$\lim_{n\to\infty}\frac{1}{n}\left\|\boldsymbol{x}_0\right\|_2^2 = \mathsf{M}_2 < \infty, \tag{224}$$

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{\sqrt{n}}\left\|\bar{\boldsymbol{G}}^T\left(\bar{\boldsymbol{G}}\boldsymbol{x}_0 + \boldsymbol{w} - \bar{\boldsymbol{G}}\boldsymbol{x}^{t-1}\right)\right\|_2 &\leq \lim_{n\to\infty}\left\|\bar{\boldsymbol{G}}\right\|_2^2\left(\sqrt{\mathsf{M}_2} + \lim_{n\to\infty}\frac{1}{\sqrt{n}}\left\|\boldsymbol{x}^{t-1}\right\|_2\right) +\\
&\quad + \sigma\lim_{n\to\infty}\left\|\bar{\boldsymbol{G}}\right\|_2 < \infty, && (225)
\end{aligned}$$

$$\lim_{m\to\infty}\frac{1}{n}\left\|\boldsymbol{z}^t\right\|_2^2 < \infty. \tag{226}$$

The proof is complete with the calibration equation (3), the fact that $\tau_{t-1} \to \tau_*$, along with Lemma 21.

### 5.11 Uniqueness of $\alpha$, $\tau_*$ and $\gamma$

Recall Eq. (2), (3) and (4) which relate $\rho$ and $\lambda$ of the elastic net problem to $\alpha$, $\tau_*$ and $\gamma$.

**Corollary 25.** *Fix $\lambda > 0$ and $\rho > 0$. There exist uniquely $\alpha$, $\tau_*$ and $\gamma$ that satisfy Eq. (2), (3) and (4).*

*Proof.* This is a consequence of Theorem 1. Indeed, the existence follows from the lemmas in Section 4.2, so we only need to prove uniqueness. By Theorem 1, we have:

$$\frac{1}{n}\left\|\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)-\boldsymbol{x}_0\right\|_2^2 \to \delta\left(\tau_*^2-\sigma^2\right) \tag{227}$$

in probability as $n\to\infty$. Since $\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)$ is unique and is determined by the elastic net's parameters, we deduce that $\tau_*$ must be unique. We also have:

$$\frac{1}{n}\left\|\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{A}}\right)\right\|_1 \to \mathbb{E}\left[\left|\frac{1}{\gamma+1}\eta\left(X_0+\tau_* Z;\alpha\tau_*\right)\right|\right] \tag{228}$$

$$= \mathbb{E}\left[\left|\frac{1}{\frac{\rho}{\lambda}\alpha\tau_*+1}\eta\left(X_0+\tau_* Z;\alpha\tau_*\right)\right|\right] \tag{229}$$

in probability, where we use Eq. (4). Given a fixed $\tau_* > 0$, recalling that the support of the distribution of $Z$ is $\mathbb{R}$ since $Z\sim\mathcal{N}(0,1)$, we have the right-hand side decreases as $\alpha\tau_*$ increases. Therefore $\alpha\tau_*$ and consequently $\alpha$ must be unique. Finally, since $\gamma=\frac{\rho}{\lambda}\alpha\tau_*$, this implies uniqueness of $\gamma$, completing the proof. □

## 6 Useful Facts

We state three useful known results. The first concerns with the Bai-Yin law on the convergence of the maximum singular values of random matrices. The second concerns with the convergence of the general AMP iterates (120)-(123). The third is a variant of the Lindeberg's principle, which can be established following [Cha06, KM11]. A proof for the third result is included for reference.

**Theorem 26** (Bai-Yin law [BY93])**.** *Let $\boldsymbol{M}\in\mathbb{R}^{m\times n}$ be a random matrix with i.i.d. entries in which $\mathbb{E}\left[M_{ij}\right]=0$, $\mathbb{E}\left[M_{ij}^2\right]=1$, $\mathbb{E}\left[M_{ij}^4\right]<\infty$ and $\frac{m}{n}\to\delta>0$ as $n\to\infty$. Then $\|\boldsymbol{M}\|_2$ and $\sigma_{\min}(\boldsymbol{M})$ (the minimum singular value of $\boldsymbol{M}$) converge almost surely to $\frac{1}{\sqrt{\delta}}+1$ and $\max\left\{1-\frac{1}{\sqrt{\delta}},0\right\}$ respectively, as $n\to\infty$.*

**Lemma 27** ([BM11, Part of Lemma 1])**.** *Let $\{\boldsymbol{M}(n)\}_{n\in\mathbb{N}}$ be a sequence of matrices in $n$, in which $\boldsymbol{M}\in\mathbb{R}^{m\times n}$ with i.i.d. entries $M_{ij}\sim\mathcal{N}\left(0,\frac{1}{m}\right)$, and $\frac{m}{n}\to\delta>0$. Let $\{\boldsymbol{x}_0(n)\}_{n\in\mathbb{N}}$, $\{\boldsymbol{w}(n)\}_{n\in\mathbb{N}}$ and $\{\boldsymbol{q}^0(n)\}_{n\in\mathbb{N}}$ be sequences of vectors whose empirical distributions converge weakly to probability measures $p_{X_0}$, $p_W$ and $p_Q$ on $\mathbb{R}$, in which $\mathbb{E}\left[X_0^{2k-2}\right]<\infty$, $\mathbb{E}\left[W^{2k-2}\right]<\infty$ and $\mathbb{E}\left[Q^{2k-2}\right]<\infty$, for a given integer $k\geq 2$.*

*Recall the general AMP iterates (120)-(123). Consider sequences of non-negative scalars $\{\tau_t\}_{t\in\mathbb{N}}$ and $\{\sigma_t\}_{t\in\mathbb{N}}$ which follow the recursions:*

$$\tau_t^2=\mathbb{E}\left[g_t\left(\sigma_t Z,W\right)^2\right],\qquad \sigma_t^2=\frac{1}{\delta}\mathbb{E}\left[f_t\left(\tau_{t-1}Z,X_0\right)^2\right], \tag{230}$$

*where $Z \sim \mathcal{N}(0,1)$ independent of $X_0$ and $W$, with the initialization $\sigma_0^2 = \frac{1}{\delta}\mathbb{E}\left[Q^2\right]$. Then for all $t \in \mathbb{N}$:*

1. *for all pseudo-Lipschitz functions $\phi_h, \phi_b : \mathbb{R}^{t+2} \to \mathbb{R}$ of order $k$,*

$$\frac{1}{n}\sum_{i=1}^{n} \phi_h\left(h_i^1, ..., h_i^{t+1}, x_{0,i}\right) \to \mathbb{E}\left[\phi_h\left(\tau_0 Z_0, ..., \tau_t Z_t, X_0\right)\right], \tag{231}$$

$$\frac{1}{n}\sum_{i=1}^{n} \phi_b\left(b_i^0, ..., b_i^t, w_i\right) \to \mathbb{E}\left[\phi_b\left(\sigma_0 \hat{Z}_0, ..., \sigma_t \hat{Z}_t, W\right)\right] \tag{232}$$

   *almost surely, where $(Z_0, ..., Z_t)$ and $\left(\hat{Z}_0, ..., \hat{Z}_t\right)$ are independent of $X_0$ and $W$, and $Z_i, \hat{Z}_i \sim \mathcal{N}(0,1)$,*

2. *for all $r, s \in \{0, 1, ..., t\}$,*

$$\lim_{n\to\infty} \frac{1}{n}\left\langle \boldsymbol{h}^{r+1}, \boldsymbol{h}^{s+1}\right\rangle \overset{\text{a.s.}}{=} \lim_{n\to\infty}\frac{1}{m}\left\langle \boldsymbol{m}^r, \boldsymbol{m}^s\right\rangle, \tag{233}$$

$$\lim_{n\to\infty} \frac{1}{m}\left\langle \boldsymbol{b}^{r}, \boldsymbol{b}^{s}\right\rangle \overset{\text{a.s.}}{=} \frac{1}{\delta}\lim_{n\to\infty}\frac{1}{n}\left\langle \boldsymbol{q}^r, \boldsymbol{q}^s\right\rangle. \tag{234}$$

**Theorem 28** (Lindeberg's principle for bounded distributions)**.** *Consider $\boldsymbol{A}, \boldsymbol{G} \in \mathbb{R}^{m\times n}$ two random matrices with independent entries such that $|A_{ij}| \leq C_1$ and $|G_{ij}| \leq C_1$ with probability 1, and $\mathbb{E}\left[A_{ij}\right] = \mathbb{E}\left[G_{ij}\right]$, $\mathbb{E}\left[A_{ij}^2\right] = \mathbb{E}\left[G_{ij}^2\right]$ for any $i \in [m]$, $j \in [n]$. Let $\boldsymbol{D}(q,p,v) \in \mathbb{R}^{m\times n}$ be such that its $(i,j)$-th entry is*

$$D_{ij}(q,p,v) = \begin{cases} A_{ij}, & i < q \text{ or } (i = q \text{ and } j < p), \\ v, & i = q \text{ and } j = p, \\ G_{ij}, & \text{otherwise.} \end{cases} \tag{235}$$

*Consider a function $h: \mathbb{R}^{m\times n} \to \mathbb{R}$ such that $h$ is thrice-differentiable w.r.t. each coordinate and*

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{E}\left[\left|\frac{\partial^3}{\partial v^3} h\left(\boldsymbol{D}(i,j,v)\right)\right|\right] \leq C_2 \tag{236}$$

*for $|v| \leq C_1$. Then:*

$$\left|\mathbb{E}\left[h(\boldsymbol{A})\right] - \mathbb{E}\left[h(\boldsymbol{G})\right]\right| \leq \frac{1}{3}C_1^3 C_2. \tag{237}$$

*Proof.* We use $\partial$, $\partial^2$ and $\partial^3$ as short-hand notations for $\frac{\partial}{\partial v}$, $\frac{\partial^2}{\partial v^2}$ and $\frac{\partial^3}{\partial v^3}$. We have:

$$\left|\mathbb{E}\left[h(\boldsymbol{A})\right] - \mathbb{E}\left[h(\boldsymbol{G})\right]\right| = \left|\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{E}\left[h\left(\boldsymbol{D}(i,j,A_{ij})\right) - h\left(\boldsymbol{D}(i,j,G_{ij})\right)\right]\right| \tag{238}$$

$$\overset{(a)}{=} \left|\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{6}\mathbb{E}\left[\partial^3 h\left(\boldsymbol{D}\left(i,j,v_A^{ij}\right)\right)A_{ij}^3 - \partial^3 h\left(\boldsymbol{D}\left(i,j,v_G^{ij}\right)\right)G_{ij}^3\right]\right| \tag{239}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{6}\left\{\mathbb{E}\left[\left|\partial^3 h\left(\boldsymbol{D}\left(i,j,v_A^{ij}\right)\right)\right|\left|A_{ij}\right|^3\right] + \mathbb{E}\left[\left|\partial^3 h\left(\boldsymbol{D}\left(i,j,v_G^{ij}\right)\right)\right|\left|G_{ij}\right|^3\right]\right\} \tag{240}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{6}C_1^3\left\{\mathbb{E}\left[\left|\partial^3 h\left(\boldsymbol{D}\left(i,j,v_A^{ij}\right)\right)\right|\right]+\mathbb{E}\left[\left|\partial^3 h\left(\boldsymbol{D}\left(i,j,v_G^{ij}\right)\right)\right|\right]\right\} \tag{241}$$

for some $v_A^{ij}$ between 0 and $A_{ij}$ and some $v_G^{ij}$ between 0 and $G_{ij}$. Here to derive step $(a)$, we use Taylor's theorem for $h\left(\boldsymbol{D}\left(i,j,A_{ij}\right)\right)$, in particular,

$$\begin{aligned} h\left(\boldsymbol{D}\left(i,j,A_{ij}\right)\right) = h\left(\boldsymbol{D}\left(i,j,0\right)\right) + A_{ij}\partial h\left(\boldsymbol{D}\left(i,j,0\right)\right) + \frac{1}{2}A_{ij}^2\partial^2 h\left(\boldsymbol{D}\left(i,j,0\right)\right) \\ + \frac{1}{6}A_{ij}^3\partial^3 h\left(\boldsymbol{D}\left(i,j,v_{ij}^A\right)\right) \end{aligned} \tag{242}$$

and similarly for $h\left(\boldsymbol{D}\left(i,j,G_{ij}\right)\right)$. Observe that $A_{ij}$ and $G_{ij}$ are independent of $\boldsymbol{D}\left(i,j,0\right)$. Using the fact $\mathbb{E}\left[A_{ij}\right]=\mathbb{E}\left[G_{ij}\right]$ and $\mathbb{E}\left[A_{ij}^2\right]=\mathbb{E}\left[G_{ij}^2\right]$, we arrive at $(a)$. The proof is complete if $\left|v_A^{ij}\right|\leq C_1$ and $\left|v_G^{ij}\right|\leq C_1$. But this is immediate from that $\left|A_{ij}\right|\leq C_1$ and $\left|G_{ij}\right|\leq C_1$. □

## References


[ATH16] Ehsan Abbasi, Christos Thrampoulidis, and Babak Hassibi, *General performance metrics for the lasso*, Information Theory Workshop (ITW), 2016 IEEE, IEEE, 2016, pp. 181–185.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration inequalities: A nonasymptotic theory of independence*, Oxford University Press, 2013.

[BLM+15] Mohsen Bayati, Marc Lelarge, Andrea Montanari, et al., *Universality in polytope phase transitions and message passing algorithms*, Ann. Appl. Probabil. **25** (2015), no. 2, 753–822.

[BM11] M. Bayati and A. Montanari, *The dynamics of message passing on dense graphs, with applications to compressed sensing*, IEEE Trans. Inf. Theory **57** (2011), no. 2, 764–785.

[BM12] ———, *The LASSO risk for gaussian matrices*, IEEE Trans. Inf. Theory **58** (2012), no. 4, 1997–2017.

[BY93] ZD Bai and YQ Yin, *Limit of the smallest eigenvalue of a large dimensional sample covariance matrix*, Ann. Probabil. (1993), 1275–1294.

[Cha06] Sourav Chatterjee, *A generalization of the lindeberg principle*, Ann. Probabil. **34** (2006), no. 6, 2061–2076.

[Don05] David L Donoho, *Neighborly polytopes and sparse solutions of underdetermined linear equations*, Tech. report, Stanford Uni., Stanford, CA, 2005.

[DT09] David Donoho and Jared Tanner, *Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing*, Philos. Trans. R. Soc. Lond. Ser. A Math., Phys. Eng. Sci. **367** (2009), no. 1906, 4273–4293.

[HTW15] Trevor Hastie, Robert Tibshirani, and Martin Wainwright, *Statistical learning with sparsity: the lasso and generalizations*, CRC Press, 2015.

[KM11] S. B. Korada and A. Montanari, *Applications of the Lindeberg principle in communications and statistical learning*, IEEE Trans. Inf. Theory **57** (2011), no. 4, 2440–2450.

[Lat05] Rafał Latała, *Some estimates of norms of random matrices*, Proceedings of the American Mathematical Society **133** (2005), no. 5, 1273–1282.

[MN17] Andrea Montanari and Phan-Minh Nguyen, *Universality of the elastic net error*, 2017 IEEE International Symposium on Information Theory (ISIT), IEEE, 2017, pp. 2338–2342.

[OT15] Samet Oymak and Joel A Tropp, *Universality laws for randomized dimension reduction, with applications*, arXiv preprint arXiv:1511.09433 (2015).

[OTH13] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi, *The squared-error of generalized lasso: A precise analysis*, arXiv preprint arXiv:1311.0830 (2013).

[Sto13] Mihailo Stojnic, *A framework to characterize performance of lasso algorithms*, arXiv preprint arXiv:1303.7291 (2013).

[SWM14] Weiwei Shen, Jun Wang, and Shiqian Ma, *Doubly regularized portfolio with risk minimization*, AAAI, 2014, pp. 1286–1292.

[TAH16] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi, *Precise error analysis of regularized m-estimators in high-dimensions*, arXiv preprint arXiv:1601.06233 (2016).

[Tao12] Terence Tao, *Topics in random matrix theory*, vol. 132, American Mathematical Society Providence, RI, 2012.

[Tib13] Ryan J. Tibshirani, *The LASSO problem and uniqueness*, Electron. J. Statist. **7** (2013), 1456–1490.

[TOH15] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi, *Regularized linear regression: A precise analysis of the estimation error*, Proc. 28th Conf. Learning Theory, 2015, pp. 1683–1709.

[TPH15] Christos Thrampoulidis, Ashkan Panahi, and Babak Hassibi, *Asymptotically exact error analysis for the generalized $\ell_2^2$-lasso*, arXiv preprint arXiv:1502.06287 (2015).

[Vil03] Cédric Villani, *Topics in optimal transportation*, no. 58, American Mathematical Soc., 2003.

[WZZ06] Li Wang, Ji Zhu, and Hui Zou, *The doubly regularized support vector machine*, Statistica Sinica (2006), 589–615.

[ZH05] Hui Zou and Trevor Hastie, *Regularization and variable selection via the elastic net*, J. R. Statist. Soc. Ser. B **67** (2005), no. 2, 301–320.