ALL-OR-NOTHING PHENOMENA: FROM SINGLE-LETTER TO HIGH DIMENSIONS

Galen Reeves[*] *Jiaming Xu*[†] *Ilias Zadik* [‡]

ABSTRACT

We consider the problem of estimating a p-dimensional vector β from n observations $Y = X\beta + W$, where $\beta_j \overset{\text{i.i.d.}}{\sim} \pi$ for a real-valued distribution π with zero mean and unit variance, $X_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, and $W_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2)$. In the asymptotic regime where $n/p \to \delta$ and $p/\sigma^2 \to \mathsf{snr}$ for two fixed constants $\delta, \mathsf{snr} \in (0,\infty)$ as $p \to \infty$, the limiting (normalized) minimum mean-squared error (MMSE) has been characterized by a single-letter (additive Gaussian scalar) channel.

In this paper, we show that if the MMSE function of the single-letter channel converges to a step function, then the limiting MMSE of estimating β converges to a step function which jumps from 1 to 0 at a critical threshold. Moreover, we establish that the limiting mean-squared error of the (MSE-optimal) approximate message passing algorithm also converges to a step function with a larger threshold, providing evidence for the presence of a computational-statistical gap between the two thresholds.

1. INTRODUCTION

Consider the classical linear regression model

$$Y = X\beta + W \tag{1}$$

where $X \in \mathbb{R}^{n\times p}$ with $X_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $\beta \in \mathbb{R}^p$ with $\beta_j \overset{\text{i.i.d.}}{\sim} \pi$ for a distribution π with zero mean and unit variance, and $W \in \mathbb{R}^n$ with $W_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2)$. We are interested in estimating β from observation of (X,Y). For a given estimator $\widehat{\beta}(X,Y)$, the normalized mean squared-error of estimating β is given by

$$\mathsf{MSE}\left(\widehat{\beta}\right) := \frac{1}{p}\mathbb{E}\left[\left\|\beta - \widehat{\beta}\right\|^2\right].$$

Let MMSE denote the minimum of $\mathsf{MSE}\left(\widehat{\beta}\right)$ among all possible estimators $\widehat{\beta}$, or equivalently,

$$\mathsf{MMSE} := \frac{1}{p}\mathbb{E}\left[\|\beta - \mathbb{E}[\beta \mid X, Y]\|^2\right]. \tag{2}$$

In this paper, we focus on the asymptotic regime:

$$\frac{n}{p} \to \delta \quad \text{and} \quad \frac{p}{\sigma^2} \to \mathsf{snr}, \quad \text{as } p \to \infty, \tag{3}$$

for two fixed constants $\delta, \mathsf{snr} \in (0,\infty)$. Note that δ is the under-sampling ratio and snr is the signal-to-noise ratio in view of $\mathbb{E}\left[\|X\beta\|^2\right]/\mathbb{E}\left[\|W\|^2\right] = p/\sigma^2$.

Recent work [1–3] proves that under certain structural assumptions in terms of $(\pi, \delta, \mathsf{snr})$, the limiting MMSE in the asymptotic regime (3) is characterized by the *replica-symmetric* (RS) formula through a single letter channel

$$y = \sqrt{s}\beta_0 + N, \tag{4}$$

where $s > 0$, $\beta_0 \sim \pi$ and $N \sim \mathcal{N}(0,1)$ are independent. However, often the RS formula is too complicated to extract structural behavior of the limiting MMSE.

In this work, we discover that the limiting MMSE exhibits an all-or-nothing phenomena. More precisely, consider a family $(\pi_\epsilon, \delta_\epsilon, \mathsf{snr}_\epsilon)$ indexed by a positive parameter ϵ where π_ϵ has finite entropy $H_\epsilon := H(\pi_\epsilon)$. We show that if the MMSE of the single letter channel (4) as a function of s converges to a step function as $\epsilon \to 0$, then the limiting MMSE of the linear regression model (1) also converges to a step function, which jumps from 1 to 0 at a critical threshold $\delta_\epsilon = \delta_{\epsilon,\mathsf{MMSE}}$, where

$$\delta_{\epsilon,\mathsf{MMSE}} := \frac{2H_\epsilon}{\log(1+\mathsf{snr}_\epsilon)}. \tag{5}$$

In other words, an all-or-nothing phenomena in the single letter channel implies an all-or-nothing phenomena in the high-dimensional linear regression model. Moreover, we establish that the limiting MSE of the (MSE-optimal) approximate message passing (AMP) algorithm also converges to a step function, which jumps from 1 to 0 at a larger threshold $\delta_\epsilon = \delta_{\epsilon,\mathsf{AMP}}$, where

$$\delta_{\epsilon,\mathsf{AMP}} := \frac{2H_\epsilon(1+\mathsf{snr}_\epsilon)}{\mathsf{snr}_\epsilon}. \tag{6}$$

An important application of our general result is the binary linear regression model where $\beta_j \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(\epsilon)$. In this case, we show that the MMSE function of the single letter channel converges to a step function as the sparsity $\epsilon \to 0$. Then we obtain from our general result that the limiting MMSE of the binary linear regression model converges to a step function which

[*] G. Reeves is with the Department of Electrical and Computer Engineering and the Department of Statistical Science, Duke University, Durham, NC 27708 USA; e-mail: galen.reeves@duke.edu.

[†] J. Xu is with the Fuqua School of Business, Duke University, Durham, NC 27708 USA; e-mail: jiamingxu.868@duke.edu.

[‡] I. Zadik is with the Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139 USA; e-mail: izadik@mit.edu.

jumps from 1 to 0 at the critical threshold $\delta_\epsilon = \frac{2\epsilon\log(1/\epsilon)}{\log(1+\mathsf{snr}_\epsilon)}$. This coincides with the all-or-nothing phenomena established in [4] for the binary linear regression model where β is chosen uniformly at random from the set of binary k-sparse vectors, in the highly sparse and high signal-to-noise ratio regime where $k/\sqrt{p} \to 0$ and k/σ^2 is above a sufficiently large constant. Furthermore, we deduce from our general result that the limiting MSE of the (MSE-optimal) AMP converges to a step function which jumps from 1 to 0 at the critical threshold $\delta_\epsilon = \frac{2\epsilon\log(1/\epsilon)(1+\mathsf{snr}_\epsilon)}{\mathsf{snr}_\epsilon}$. This coincides with the computational threshold for a number of computationally efficient methods in the literature such as LASSO or Orthogonal Matching Pursuit. In particular, our result adds to the existing evidence for the presence of a computational-statistical gap (see [5, 6] for an extended discussion and literature review on the presence of this computational-statistical gap).

2. PRELIMINARIES

2.1. The Replica Symmetric Formulas

To describe the RS formulas, we first define the mutual information and MMSE functions for the single letter channel (4):

$$I(s) := I(\beta_0; \sqrt{s}\beta_0 + N), \quad s > 0 \tag{7}$$

$$M(s) := \mathsf{mmse}(\beta_0 \mid \sqrt{s}\beta_0 + N), \quad s > 0 \tag{8}$$

where $\beta_0 \sim \pi$ and $N \sim \mathcal{N}(0,1)$ are independent. Both of these functions are non-negative and the unit variance assumption on π means that for any $s > 0$, (see [7] for details)

$$I(s) \leq \frac{1}{2}\log(1+s) \leq \frac{s}{2}, \tag{9}$$

$$M(s) \leq \frac{1}{1+s} \leq 1. \tag{10}$$

Next, we define the potential function $\mathcal{F} : [0,\infty) \to [0,\infty)$ according to

$$\mathcal{F}(s) := I(s) + \frac{\delta}{2}\phi\Big(\frac{s}{\delta\,\mathsf{snr}}\Big), \tag{11}$$

where $\phi(x) = x - \log x - 1$, and δ, snr are respectively the undersampling ratio and the signal-to-noise ratio of our original model. Note that $\phi(x)$ is convex and non-negative on $(0,\infty)$.

Lemma 1. *All stationary points of $\mathcal{F}(s)$ lie on the open interval between $\delta\,\mathsf{snr}/(1+\mathsf{snr})$ and $\delta\,\mathsf{snr}$.*

Proof. By differentiation with respect to s and the I-MMSE relation for the single-letter channel [7], we have that for any $s > 0$

$$\mathcal{F}'(s) \propto M(s) + 1/\mathsf{snr} - \delta/s.$$

The fact that $M(s) > 0$ for all s implies that $\mathcal{F}'(s)$ is strictly positive for all $s \geq \delta\,\mathsf{snr}$, and thus $\mathcal{F}(s)$ is strictly increasing on $[\delta\,\mathsf{snr}, \infty)$. Alternatively, the fact that $M(s) < 1$ for all $s > 0$ implies that $\mathcal{F}'(s)$ is strictly negative for all $s \leq \delta\,\mathsf{snr}/(1+\mathsf{snr})$, and thus $\mathcal{F}(s)$ is strictly decreasing on $(0, \delta\,\mathsf{snr}/(1+\mathsf{snr})]$. □

In view of Lemma 1, the minimum of the potential function and the smallest and largest minimizers can be defined as follows:

$$\mathcal{F}^* := \min_s \mathcal{F}(s), \tag{12}$$

$$\underline{s}^* := \min\{s : \mathcal{F}(s) = \mathcal{F}^*\}, \tag{13}$$

$$\overline{s}^* := \max\{s : \mathcal{F}(s) = \mathcal{F}^*\}. \tag{14}$$

Note that $\underline{s}^* = \overline{s}^*$ if and only if the minimum is attained at a unique point.

Proposition 2 (RS MMSE [2, 3, 8])**.** *For any $(\delta, \mathsf{snr}, \pi)$ for which (snr, π) satisfies the single-crossing property [2] and π has finite fourth moment* [1]*, the mutual information and MMSE satisfy*

$$\lim_{p\to\infty} \frac{1}{p} I(\beta; X, Y) = \mathcal{F}^*, \tag{15}$$

$$\limsup_{p\to\infty} \frac{1}{p}\mathbb{E}\Big[\|\beta - \mathbb{E}[\beta \mid y, X]\|^2\Big] \leq M(\overline{s}^*), \tag{16}$$

$$\liminf_{p\to\infty} \frac{1}{p}\mathbb{E}\Big[\|\beta - \mathbb{E}[\beta \mid y, X]\|^2\Big] \geq M(\underline{s}^*), \tag{17}$$

where the limits are taken as $(n = n_p, p, \sigma^2 = \sigma_p^2)$ scale to infinity with $p \to +\infty$, $n/p \to \delta$ and $p/\sigma^2 \to \mathsf{snr}$.

Next, we turn to the family of *approximate message passing* (AMP) [9, 10] algorithms and specifically to the case of MMSE-AMP which is proven in to be optimal among AMP algorithms in minimizing the MSE of the recovery problem of interest [11]. For simplicity from now on when we say AMP we refer to the AMP-MMSE algorithm. We show how a related formula to the one described in Proposition 2 describes the asymptotic MSE associated with AMP.

The smallest stationary point is defined as

$$s^{\mathsf{AMP}} := \inf\{s : \mathcal{F}'(s) = 0\}. \tag{18}$$

It is rather straightforward to check that s^{AMP} is attained by some positive value s and therefore its a stationary point of $\mathcal{F}(s)$. In particular, by Lemma 1 we have $s^{\mathsf{AMP}} \in (\delta\,\mathsf{snr}/(1+\mathsf{snr}), \delta\,\mathsf{snr})$.

For the next result, for $T \in \mathbb{N}$ let $\widehat{\beta}_{\mathsf{AMP},T}(Y,X)$ the output of the AMP estimator [11, Section II.C] with input data (Y, X) after T iterations.

[1] Different set of assumptions on $(\delta, \mathsf{snr}, \pi)$ for which the Proposition holds can be found in [3, 8]

Proposition 3 (AMP, [10, 11])**.** *For any* $(\delta, \mathsf{snr}, \pi)$ *where* π *has a finite fourth moment, AMP satisfies*

$$\lim_{T\to+\infty}\lim_{p\to+\infty}\frac{1}{p}\mathbb{E}\left[\left\|\beta-\widehat{\beta}_{\mathsf{AMP},T}(Y,X)\right\|^2\right]=M(s^{\mathsf{AMP}}) \tag{19}$$

where the outer limit is taken as $(n = n_p, p, \sigma^2 = \sigma_p^2)$ *scale to infinity with* $p \to +\infty$, $n/p \to \delta$ *and* $p/\sigma^2 \to \mathsf{snr}$.

Remark 1. The results stated above imply that AMP is optimal whenever $\underline{s}^* = \overline{s}^* = s^{\mathsf{AMP}}$.

Remark 2. For a proof of Proposition 3 we refer the reader to the statement and proof of [11, Theorem 6].

3. MAIN RESULTS

Let us consider now a family of coefficient distributions $(\pi_\epsilon)_{\epsilon>0}$ indexed by a positive-valued parameter $\epsilon > 0$. We assume throughout the section that for each $\epsilon > 0$ the distributions π_ϵ has zero mean, unit variance and finite entropy H_ϵ. Our results are all based on the following assumption on the family π_ϵ.

Assumption 1. Let $(\pi_\epsilon)_{\epsilon>0}$ be a family of distributions with unit variance and finite entropy H_ϵ. The MMSE function $M_\epsilon(s)$ of the single letter channel, as defined in (8), for π_ϵ coefficient distribution is assumed to converge pointwise to a step function as $\epsilon \to 0$ in the following sense

$$\lim_{\epsilon\to 0} M_\epsilon(2H_\epsilon\, t) = \begin{cases} 1, & t \in [0,1) \\ 0, & t \in (1,\infty). \end{cases} \tag{20}$$

Remark 3. It can be straightforwardly checked using the I-MMSE relation for the single-letter channel [7] that the rescaling in the argument of M_ϵ by twice the entropy term, i.e. by $2H_\epsilon$, is necessary for the convergence of M_ϵ to the step function.

Remark 4. As we establish later in the section, Assumption 1 is satisfied for the family of (normalized) Bernoulli distributions with probability ϵ. We expect the assumption to hold under greater generality.

We now present our two main results assuming the family of distributions $(\pi_\epsilon)_{\epsilon>0}$ satisfies Assumption 1.

Theorem 4. *Let* $(\pi_\epsilon)_{\epsilon>0}$ *satisfying Assumption 1. Given a number* $r \in (0,1) \cup (1,\infty)$*, let* $(\delta_\epsilon, \mathsf{snr}_\epsilon, \pi_\epsilon)_{\epsilon>0}$ *be a family of triplets such that*

$$\lim_{\epsilon\to 0}\frac{\delta_\epsilon}{\delta_{\epsilon,\mathsf{MMSE}}} = r. \tag{21}$$

Then, the minimizers of the RS potential function exhibit the all-or-nothing behavior in the small-ϵ *limit depending on whether* r *is greater than or less than one:*

$$r \in (0,1) \implies M_\epsilon(\overline{s}_\epsilon^*) \to 0 \tag{22}$$
$$r \in (1,\infty) \implies M_\epsilon(\underline{s}_\epsilon^*) \to 1. \tag{23}$$

Combining Theorem 4 with Proposition 2 we obtain the following Corollary.

Corollary 5 (All-or-nothing MMSE behavior)**.** *Let* $r \in (0,1) \cup (1,\infty)$*. For any family of triplets* $(\delta_\epsilon, \mathsf{snr}_\epsilon, \pi_\epsilon)_{\epsilon>0}$*, suppose that for any* $\epsilon > 0$*,* $(\mathsf{snr}_\epsilon, \pi_\epsilon)$ *satisfies the single-crossing property,* π_ϵ *has finite fourth moment,* $(\pi_\epsilon)_{\epsilon>0}$ *satisfies Assumption 1, and* (21) *holds. Then it holds that*

$$\lim_{\epsilon\to 0}\lim_{p\to\infty}\frac{1}{p}\mathbb{E}\left[\|\beta-\mathbb{E}[\beta \mid Y,X]\|^2\right] = \begin{cases} 1, & r \in [0,1) \\ 0, & r \in (1,\infty). \end{cases} \tag{24}$$

where the inner limits are taken as $(n = n_p, p, \sigma^2 = \sigma_p^2)$ *scale to infinity with* $p \to +\infty$, $n/p \to \delta_\epsilon$ *and* $p/\sigma^2 \to \mathsf{snr}_\epsilon$.

We next present our second main result.

Theorem 6. *Let* $(\pi_\epsilon)_{\epsilon>0}$ *satisfying Assumption 1. Given a number* $r \in (0,1) \cup (1,\infty)$*, let* $(\delta_\epsilon, \mathsf{snr}_\epsilon, \pi_\epsilon)$ *be such that*

$$\lim_{\epsilon\to 0}\frac{\delta_\epsilon}{\delta_{\epsilon,\mathsf{AMP}}} = r \tag{25}$$

Then, the smallest stationary point s^{AMP} *exhibits the all-or-nothing behavior in the small-*ϵ *limit depending on whether* r *is greater than or less than one:*

$$r \in (0,1) \implies M_\epsilon\big(s_\epsilon^{\mathsf{AMP}}\big) \to 1 \tag{26}$$
$$r \in (1,\infty) \implies M_\epsilon\big(s_\epsilon^{\mathsf{AMP}}\big) \to 0. \tag{27}$$

For the result, for $T \in \mathbb{N}$ let $\widehat{\beta}_{\mathsf{AMP},T}(Y,X)$ the output of the AMP estimator [11, Section II.C] with input data (Y,X) after T iterations. Combining Theorem 6 with Proposition 3 we obtain the following Corollary on the performance of AMP.

Corollary 7 (All-or-nothing AMP behavior)**.** *Let* $r \in (0,1) \cup (1,\infty)$*. For any family of triplets* $(\delta_\epsilon, \mathsf{snr}_\epsilon, \pi_\epsilon)_{\epsilon>0}$*, suppose that each* π_ϵ *has a finite fourth moment, the family* $(\pi_\epsilon)_{\epsilon>0}$ *satisfies Assumption 1, and* (25) *holds. Then it holds that*

$$\lim_{\epsilon\to 0}\lim_{T\to+\infty}\lim_{p\to\infty}\frac{1}{p}\mathbb{E}\left[\left\|\beta-\widehat{\beta}_{\mathsf{AMP},T}(Y,X)\right\|^2\right] = \begin{cases} 1, & r \in [0,1) \\ 0, & r \in (1,\infty). \end{cases} \tag{28}$$

where the inner limits are taken as $(n = n_p, p, \sigma^2 = \sigma_p^2)$ *scale to infinity with* $p \to +\infty$, $n/p \to \delta_\epsilon$ *and* $p/\sigma^2 \to \mathsf{snr}_\epsilon$.

4. APPLICATION: SPARSE BINARY REGRESSION

We now present our main application of our two results to sparse binary regression, where $\beta_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(\epsilon)$. To this end, we first consider the case where β_i is i.i.d. drawn from the following two-point distribution:

$$\pi_\epsilon = (1-\epsilon)\,\delta_{\mu_1} + \epsilon\,\delta_{\mu_2}, \tag{29}$$

where δ_x denotes a Dirac distribution with mass at $x \in \mathbb{R}$, and $\mu_1 = -\sqrt{\epsilon/(1-\epsilon)}$ and $\mu_2 = \sqrt{(1-\epsilon)/\epsilon}$ are chosen such that π_ϵ has zero mean and unit variance. The following Lemma holds for the family of MMSE functions $(M_\epsilon(s))_{\epsilon>0}$:

Lemma 8. *The distribution π_ϵ in* (29) *has entropy $H_\epsilon = -\epsilon \log \epsilon - (1-\epsilon)\log(1-\epsilon)$ and MMSE function*

$$M_\epsilon(s) = \mathbb{E}\left[\frac{1}{1-\epsilon+\epsilon\,\exp\left(\frac{s}{2\epsilon(1-\epsilon)}+\sqrt{\frac{s}{\epsilon(1-\epsilon)}}N\right)}\right], \tag{30}$$

where $N \sim \mathcal{N}(0,1)$. Furthermore, the distribution π_ϵ satisfies the single-crossing condition [2] for all $\mathsf{snr} > 0$ *and*

$$\lim_{\epsilon\to 0} H_\epsilon/(\epsilon \log 1/\epsilon) = 1 \tag{31}$$

$$\lim_{\epsilon\to 0}\sup_{s>0}\left|M_\epsilon(s) - Q\left(\frac{s-2\epsilon\log(1/\epsilon)}{2\sqrt{s\epsilon}}\right)\right| = 0, \tag{32}$$

where $Q(z) = \int_z^\infty (2\pi)^{-1/2}\exp(-t^2/2)\,dt$.

An immediate implication of the result is that the family of distributions $(\pi_\epsilon)_{\epsilon>0}$ satisfies Assumption 1 as well as the conditions of Corollaries 5 and 7. Hence, all-or-nothing phase transitions hold for the limiting MMSE around $\delta_{\epsilon,\mathsf{MMSE}}$ and for the MSE of the AMP around $\delta_{\epsilon,\mathsf{AMP}}$. Using that $\lim_{\epsilon\to 0} H_\epsilon/(\epsilon\log 1/\epsilon) = 1$, we can further simplify the phase transition points given in (5), (6) by observing

$$\lim_{\epsilon\to 0}\delta_{\epsilon,\mathsf{MMSE}}\Big/\left(\frac{2\epsilon\log(1/\epsilon)}{\log(1+\mathsf{snr}_\epsilon)}\right) = 1 \tag{33}$$

and

$$\lim_{\epsilon\to 0}\delta_{\epsilon,\mathsf{AMP}}\Big/\left(\frac{2(1+\mathsf{snr}_\epsilon)\epsilon\log(1/\epsilon)}{\mathsf{snr}_\epsilon}\right) = 1. \tag{34}$$

Next, we extend the above results to the sparse binary regression problem of interest, where $\beta_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(\epsilon)$. We denote by $k = \epsilon p$ the (expected) number of non-zero coordinates of β. Define

$$\widetilde{\beta} = \frac{\beta - \mathbb{E}[\beta]}{\sqrt{\epsilon(1-\epsilon)}}.$$

Then $\widetilde{\beta}_i \overset{\text{i.i.d.}}{\sim} \pi_\epsilon$ as given in (29). Moreover, define

$$\widetilde{Y} = \frac{Y - X\mathbb{E}[\beta]}{\sqrt{\epsilon(1-\epsilon)}}, \quad \widetilde{W} = \frac{W}{\sqrt{\epsilon(1-\epsilon)}}.$$

Then it follows that $\widetilde{Y} = X\widetilde{\beta} + \widetilde{W}$. Since $\widetilde{W}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\widetilde{\sigma}^2)$ with $\widetilde{\sigma} = \sigma/\sqrt{\epsilon(1-\epsilon)}$, it follows that

$$\mathsf{snr}_\epsilon = \frac{p}{\widetilde{\sigma}^2} = \frac{p\epsilon(1-\epsilon)}{\sigma^2}. \tag{35}$$

Hence, according to Corollary 5, (33), (35), we obtain that the limiting MMSE exhibits an all-or-nothing behavior at

$$\delta_{\epsilon,\mathsf{MMSE}} = 2\epsilon\log(1/\epsilon)/\log(1+\epsilon(1-\epsilon)p/\sigma^2),$$

which using $k = \epsilon p$ as $\epsilon \to 0$ simplifies with negligible multiplicative error to

$$\delta_{\epsilon,\mathsf{MMSE}} = 2(k/p)\log(p/k)/\log(1+k/\sigma^2).$$

Note that this is the exact information-theoretic threshold for which an all-or-nothing phenomenon has been proven to hold when $\limsup_p \log k/\log p < 0.5$ in [4].

Similarly, according to Corollary 7, (34), and (35), the limiting MSE of the AMP exhibits an all-or-nothing behavior at:

$$\delta_{\epsilon,\mathsf{AMP}} = 2\left(1+\frac{\sigma^2}{p\epsilon(1-\epsilon)}\right)\epsilon\log(1/\epsilon),$$

which using $k = \epsilon p$ as $\epsilon \to 0$ simplifies with negligible multiplicative error to

$$\delta_{\epsilon,\mathsf{AMP}} = 2\big(k+\sigma^2\big)\log(p/k)/p.$$

Note that this is the exact computational threshold for a number of computationally efficient methods in the literature such as LASSO or Orthogonal Matching Pursuit (see [5, 6] for references). Our result suggest that the threshold corresponds to a barrier also for AMP in a strong sense.

5. REFERENCES

[1] Dongning Guo and S. Verdu, "Randomly spread cdma: asymptotics via statistical physics," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[2] Galen Reeves and Henry D. Pfister, "The replica-symmetric prediction for random linear estimation with Gaussian matrices is exact," *IEEE Trans. Inform. Theory*, vol. 65, no. 4, pp. 2252–2283, Apr. 2019.

[3] Jean Barbier, Mohamad Dia, Nicolas Macris, and Florent Krzakala, "The mutual information in random linear estimation," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, 2016.

[4] Galen Reeves, Jiaming Xu, and Ilias Zadik, "The all-or-nothing phenomenon in sparse linear regression," in *Conference On Learning Theory (COLT)*, 2019, [Online]. Available `https://arxiv.org/abs/1903.05046`.

[5] David Gamarnik and Ilias Zadik, "High dimensional regression with binary coefficients. estimating squared error and a phase transtition," in *COLT*, 2017.

[6] David Gamarnik and Ilias Zadik, "Sparse high-dimensional linear regression. algorithmic barriers and a local search algorithm," 2017, [Online]. Available `https://arxiv.org/abs/1711.04952`.

[7] Dongning Guo, S. Shamai, and S. Verdu, "Mutual information and mmse in gaussian channels," in *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, June 2004, pp. 349–349.

[8] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, pp. 5451–5460, 2019.

[9] David L. Donoho, Arian Maleki, and Andrea Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[10] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[11] Galen Reeves and Michael Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inf. Theor.*, vol. 58, no. 5, pp. 3065–3092, May 2012.