We thank the reviewers for their thorough comments and suggestions. Let us first address the main weakness of the submission: all the reviewers point out that the discussion on the asymptotic MMSE and the all-or-nothing phenomenon is nonrigorous. We are now able to resolve this point. The main difficulty in order to make the discussion on the MMSE rigorous is to compute the limit of the variational formula in Theorem 1 when ρ_n vanishes. Since the submission deadline we have proved that the latter limit is given by the minimum of a finite set of values whenever P_0 is a discrete distribution with a finite support. We can combine this limit with Theorem 1 to obtain the result:

Theorem 2. *Suppose that* $\Delta > 0$ *and that* $P_{0,n} := (1-\rho_n)\delta_0 + \rho_n P_0$ *where* P_0 *is a discrete distribution with finite support* $\mathrm{supp}(P_0) \subseteq \{-v_K, -v_{K-1}, \dots, -v_1, v_1, v_2, \dots, v_K\}$ *where* $0 < v_1 < v_2 < \cdots < v_K < v_{K+1} := +\infty$. *Further assume that the hypotheses (H2), (H3) of Theorem 1 hold. Let* $\rho_n = \Theta(n^{-\lambda})$ *with* $\lambda \in (0, 1/9)$ *and* $\alpha_n = \gamma\rho_n|\ln\rho_n|$ *with* $\gamma > 0$. *Then (in what follows* $X \sim P_0$*):*

$$\lim_{n\to+\infty} \frac{I(\boldsymbol{X}^*;\boldsymbol{Y}|\boldsymbol{\Phi})}{m_n} = \min_{1\le k\le K+1} \left\{ I_{P_{\mathrm{out}}}\big(\mathbb{E}[X^2 \mathbb{1}_{\{|X|\ge v_k\}}], \mathbb{E}[X^2]\big) + \frac{\mathbb{P}(|X|\ge v_k)}{\gamma} \right\} . \tag{1}$$

Once we have Theorem 2 we can obtain the asymptotic MMSE with a little more work. First we have to introduce a modified inference problem where in addition to the observations $\boldsymbol{Y}$ we are given $\widetilde{\boldsymbol{Y}}^{(\tau)} = \sqrt{\alpha_n\tau/\rho_n}\,\boldsymbol{X}^* + \widetilde{\boldsymbol{Z}}$. When τ is close enough to 0, i.e., $\tau < 2/\gamma v_K^2$, the analysis yielding Theorem 2 can be adapted to obtain the limit

$$\lim_{n\to+\infty} \frac{I(\boldsymbol{X}^*;\boldsymbol{Y},\widetilde{\boldsymbol{Y}}^{(\tau)}|\boldsymbol{\Phi})}{m_n} = \min_{1\le k\le K+1} \left\{ I_{P_{\mathrm{out}}}\big(\mathbb{E}[X^2 \mathbb{1}_{\{|X|\ge v_k\}}], \mathbb{E}[X^2]\big) + \frac{\mathbb{P}(|X|\ge v_k)}{\gamma} + \frac{\tau\mathbb{E}[X^2\mathbb{1}_{\{|X|<v_k\}}]}{2} \right\} .$$

We can then apply the I-MMSE identity[1] to obtain the asymptotic MMSE:

Theorem 3. *Suppose that* $\Delta > 0$ *and that* $P_{0,n} := (1-\rho_n)\delta_0 + \rho_n P_0$ *where* P_0 *is a discrete distribution with finite support* $\mathrm{supp}(P_0) \subseteq \{-v_K, -v_{K-1}, \dots, -v_1, v_1, v_2, \dots, v_K\}$ *where* $0 < v_1 < v_2 < \cdots < v_K < v_{K+1} := +\infty$. *Further assume that the hypotheses (H2), (H3) of Theorem 1 hold. Let* $\rho_n = \Theta(n^{-\lambda})$ *with* $\lambda \in (0, 1/9)$ *and* $\alpha_n = \gamma\rho_n|\ln\rho_n|$ *with* $\gamma > 0$.

If the minimization problem on the right-hand side of (1) *has a unique solution* $k^* \in \{1, \dots, K+1\}$ *then*

$$\lim_{n\to+\infty} \frac{\mathbb{E}\|\boldsymbol{X}^* - \mathbb{E}[\boldsymbol{X}^*|\boldsymbol{Y},\boldsymbol{\Phi}]\|^2}{k_n} = \mathbb{E}\big[X^2 \mathbb{1}_{\{|X|<v_{k^*}\}}\big] . \tag{2}$$

When $P_{0,n}$ is a Bernoulli or Bernoulli-Rademacher distribution we have $K = 1$ and Theorem 3 corresponds to the all-or-nothing phenomenon. If $K > 1$ the phenomenology is different and the asymptotic MMSE exhibits up to K sharp phase transitions.

If our submission is accepted we shall use the additional content page to state both Theorems 2 and 3 in Section 2.2. We shall refer the reader to the supplementary material for the proofs. We shall slightly rewrite the introduction and Section 3 to take into account that the discussion on the asymptotic MMSE is now rigorous and not only heuristic.

Let us now address the main other points raised by the reviewers.

- About the figures. The horizontal lines in Figure 1 indeed correspond to the value of the potential at $q = 1$. We plot these lines to highlight how the minimum of the potential shifts from being close to $q = 0$ to being close to $q = 1$ at the all-or-nothing phase transition. We will make this clear in the caption. The inset plots in Figures 2 and 3 are zooms on tight intervals around the all-or-nothing phase transitions. We propose to get rid of these inset plots and instead to pick a tighter interval for the x-axis of the main plots (in the submission $\alpha_n/\alpha_c(\rho_n) \in [0.1, 10]$). In Figure 2 and 3 the dotted lines correspond to the performance of GAMP algorithms as predicted by the state evolution equations (see the paragraph "Further remarks" in Section 3). The departure of a dotted line from the thicker line having the same color puts in evidence the algorithmic-to-stastical gap mentioned in the paragraph "Further remarks". When this gap is small the dotted line is not visible on the figures as it matches the MMSE curve. Our captions will be corrected.
- The scaling $\rho_n = \Theta(n^{-\lambda})$ with $\lambda < 1/9$ is indeed a proof artifact. To be more precise, we have to lower bound the vanishing rate of ρ_n by $n^{-1/9}$ if we want that the upper bound of Lemma 10 in the supplementary material vanishes.
- The value $\gamma_c = 0$ for noiseless linear models in Table 1 indeed shows that there is no all-or-nothing phenomenon in this case, and is in agreement with the existing literature. E.g., Reeves, Xu, and Zadik[2] locate the transition at a sample size n^* that vanishes when the noise variance approaches zero.

Finally, we take note of the additional reference as well as the rephrasings and grammatical corrections suggested by the reviewers in order to improve our manuscript.

[1] The derivative of $I(\boldsymbol{X}^*;\boldsymbol{Y},\widetilde{\boldsymbol{Y}}^{(\tau)}|\boldsymbol{\Phi})/m_n$ with respect to τ at $\tau = 0$ is equal to half the MMSE of the original problem.

[2] Galen Reeves, Jiaming Xu, and Ilias Zadik. "The all-or-nothing phenomenon in sparse linear regression". In: *Proceedings of the Thirty-Second Conference on Learning Theory* (Phoenix, USA). Vol. 99. PMLR, 2019, pp. 2652–2663.