We would like to thank the reviewers for their insightful comments. Responses below:

Common Points:

- We want to emphasize that our work is primarily theoretical, building on [27] (published in IEEE Transaction on Information Theory) by improving sample complexity bounds and more importantly eliminating the need for strong assumptions. We believe this significantly enhances our understanding of this fundamental problem.
- Given this, we feel that experiments do not add much to the paper. Synthetic experiments will simply confirm our theorems. We urge the reviewers to evaluate the paper on the basis of the theoretical results and techniques.
- We also note that many published NeurIPS do not contain empirical evaluations, and this is not a barrier to publication or even NeurIPS awards. For example one of last year's best paper awards went to *Nearly Tight Sample Complexity Bounds for Learning Mixtures of Gaussians via Sample Compression Schemes*, a purely theoretical contribution, on a problem quite close to what we study here.

Reviewer 1:

- Thanks for pointing out the relevant papers. The second paper is quite relevant as it studies mixture model parameter recovery in a compressive sensing framework. We will add the citations.
- An illustration is a great suggestion! We will add one to the final version to help explain our algorithms.

Reviewer 2:

- Exponential SNR is common in statistical problems involving mixtures. Examples include trace reconstruction and learning binomial mixture where the dependence is known to be optimal. The exponential dependence in our result arises from a reduction to learning Gaussian mixtures. This latter problem is extremely well studied (and out of scope for our paper), and while we are not aware of better results for Gaussian mixtures, any improvement would immediately yield improvements for our setting.
- As in [27], we assume L is a constant. We can extract the dependence on L from our proof for the final version, but note that it is fairly complicated as L is implicit in the definitions of the constants in Eqs (A), (B), (C).
- ϵ-precision means that all coordinates are integer-multiples of ϵ (this assumption is also used in [27]). We can add the definition to the camera ready.
- The set of responses recovered after using 2k rows of the Vandermonde matrix is provably unique, and it is possible to recover the unknown vector uniquely using efficient decoding algorithms borrowing from literature of coding theory (such as, Berlekamp-Welch Decoder, see, Arora and Barak, Computational Complexity, Sec. 19.3).
- There are at most L unique responses (fixed sensing vector, only L possible choices for β), and by a coupon collector phenomenon $L\log(Lk^2)$ measurements suffices to ensure there are exactly L, with high probability.
- The interval must belong to the set of positive real numbers so that we can apply Lemma 1. If the difference polynomials have no roots in an interval, then by continuity, the ordering of the polynomials $f^1, f^2, \ldots, f^L$ does not change in that interval. Thus the ordering is consistent.
- We will add the definition of the minimum distance estimator to the final version.
- We will clarify the paragraph after Lemma 2. In short, the β^js are on an ϵ-grid, but $\langle v, \beta^j \rangle$ may not be if we naively run the noiseless algorithm. (The ϵ-grid property is crucial for our mixture learning algorithm.) This motivates the new decoding strategy, which ensures $\langle v, \beta^j \rangle$ is on the ϵ-grid.

Reviewer 3:

- You are right that Vandermonde matrices are ill-conditioned – this is precisely why we do not use them for the noisy case. That is also the point of having different treatments for noisyless and noisy cases, with the later being more involved. This is similar to classical results in compressed sensing.
- Due to space restrictions we cannot include the entire proof in the body. But all the details are in the appendix.
- The constants are simply artifacts of the proof.
- Note that every triple (good or bad) contains vectors with integral entries. Thus we can learn the parameters $\langle v, \beta^j \rangle$ for the three vectors. Then to check if the condition holds, note that for fixed β and our choice of vectors, we have $\langle v_1, \beta \rangle + \langle v_2, \beta \rangle = \langle v_3, \beta \rangle$, so we can match up the learned parameters. If we find two such identities involving a single parameter, we know that triplet is not good.