We appreciate the detailed, insightful, and encouraging comments from the revewiers, as well as the constructive criticism. We first highlight the novelty of the results and analyses and discuss an use case for adaptive learning where the results will be drectly applicable. Subsequently, we respond to specific comments from individual reviewers.

New results and analyses. As the reveiwers noted, the main technical results (Theorems 1 and 2) are new. The fact that the results match the corresponding results for the i.i.d. case is desirable, as this will allow extension of existing results from the i.i.d. to the MDS case in a seamless manner. However, the technical details rely on several new results for MDSs. First, for decoupling, the i.i.d. case is somewhat straightforward since one has to handle products of independent random variables (r.v.s). For MDSs (Appendix A), we worked with a quadratic form of dependent r.v.s, had to first show distributional equivalence of two such forms, and finally got the decoupling result by using a *decoupled tangent sequence* to the original MDS, but *not an independent MDS* (Theorem 3). Second, for uniformly bounding the L_p norms of r.v.s from the MDS quadratic form, we showed that or both sub-Gaussian and sub-exponential tails $E[\sup \cdots]$ can be upper bounded by $\sup E[\cdot]$, which is easier to handle, and an additive term which depends on Talagrand's γ_2 function. The analyses (Lemma 7 and 8) utilize the core argument in generic chaining along with concentration bounds (Azuma-Hoeffding and Azuma-Bernstein) for MDSs, leading to new results on uniform bounds on L_p norms of quadratic forms of MDSs (Theorems 4 and 5). We relegated most of such technical details to the appendix, but based on the constructive feedback we will discuss the core ideas behind these results and proofs in the main paper.

Use case for adaptive learning: Linear contextual bandits. Our results on Restricted Isometry Property (RIP) have direct applications to many adaptive learning problems, including linear contextual bandit (LCB) learning. As a concrete example, consider the smoothed analysis model for LCB proposed in [1, 2]. The parameter estimation step involves solving an ordinary least squares (OLS) problem with a design matrix whose rows are sub-Gaussian MDS. The parameter estimation error rate depends on the minimum eigenvalue of the design matrix. Our results significantly simplifies the minimum eigenvalue analysis (Lemma B.1 of [1]), which currently uses cumbersome high-probability boundedness arguments to satisfy the boundedness requirement in [3]. In fact, our results provide a tighter high-probability bound on minimum eigenvalue of the design matrix. Moreover, our RIP results can be directly applied to the high-dimensional LCB setting where the latent parameter is assumed to have structure (such as sparsity) and parameter recovery requires RIP of the design matrix. We will expand on this LCB application in Sec. 3.

R1. We will include a brief sketch of how the terms in (6) are bounded and give some details on what a, b, c are in (7). We appreciate the detailed comments and will update the draft to address these. Brief responses for some of the points raised: for compact sets, one can indeed use a covering argument along with Hanson-Wright, but generic chaining gives a sharper bound by using a hierarchical covering argument; lines 156, 164, its a typo, the expectation should be a conditional expectation, the analysis in Appendix B uses the correct form, we will fix it; line 199, we will update it to be $\forall u \in A$, the $\inf$-$\sup$ form is sometimes used in high-d statistics; line 227, you are correct, we will fix them.

R3. First, we give a concrete example above on LCB [1] where the main results can be directly used. Second, while the results for the i.i.d. and MDS cases match, note that the MDS results needed new tools and results which we developed as part of the work. For example, consider the decoupling results in Appendix A. Recall that for decoupling for the i.i.d. setting, one just needs to consider an independent copy of the random vector. However, for the MDS setting, an independent copy of the MDS *does not* lead to decoupling, so we had to develop the MDS decoupling result based on a decoupled tangent sequence. Similar new results were developed in the context of generic chaining. We appreciate the detailed comments, we will make a pass on the paper to incorporate these. Brief responses for some of the points raised: for (7), we plan to bring the technical results (Theorems 4 and 5 in Appendix B) in the main paper; a, b, c are independent of p; yes, we needed to introduce the γ_β function because a, b in (9) depend on the γ_2 function, this can be seen by comparing (9) with Theorems 1 and 2, but we will work on the writeup to make these clear; we in fact now have the sharper analysis which drops the extra $\log n$ term; for Corollary 1, as we discuss above, the vec-wise MDS shows up for LCB [1]; existing results on RIP for Toeplitz matrices rely on the $(2p-1)$ elements being drawn i.i.d., and we extend the result to MDSs, but we plan to replace the Toeplitz example with the arguably more compelling LCB example; for CountSketch, you are right, each η_{ij} is not sampled sequentially, but $\text{vec}(X)$ is still a MDS since the Rademacher r.v.s δ_{ij} are independent of η_{ij}, the conditional expectation $E[\delta_{ij}\eta_{ij}|.] = E[\delta_{ij}]E[\eta_{ij}|.] = 0$.

R4. We have discussed a concrete example on linear contextual bandits [1] where the main results can be directly used. Brief responses for some of the points raised: we will state (6) as a Lemma, and briefly sketch how the terms in (6) are bounded (in Appendix B); we agree with the historical remark, we will switch the RIP and JL sub-sections, make the presentation uniform (all Corollaries), and also add additional remarks on the countsketch example. We appreciate the detailed comments and will update the draft based on these.

[1] Kannan et al., A Smoothed Analysis of the Greedy Algorithm for the Linear Contextual Bandit Problem, 2018.

[2] Raghavan et al., The Externalities of Exploration and How Data Diversity Helps Exploitation, 2018.

[3] Tropp, J. A. User-friendly Tail Bounds for Sums of Random Matrices, 2012.