We thank all reviewers for their valuable comments. We thank R1 for positive comments on the writing style, R2 for recognizing the novelty of our work, R3 for raising insightful questions and concerns, and R4 for encouraging us to clarify differences from previous literature.

1. (R1, R2, R3, R4) Uniform improvement. We point out that, up to the logarithmic factor, **our bound uniformly improves upon previous results of the form (10)**. To see this, let M' be the realization of $\Phi_{2k,+}(M)$ with most columns (M' will have at least k columns since adding a column to M' will not decrease its maximum singular value). Note that $k\Phi_{2k,-}(M)/\Phi_{2k,+}(M) \le k\sigma_{\min}^2(M')/\sigma_{\max}^2(M')$ by definition of Φ. We may assume that M' has full column rank, otherwise the proof is trivial. Then, for any submatrix $Q \in \mathbb{R}^{n\times k}$ of M', we have $k\sigma_{\min}^2(M')/\sigma_{\max}^2(M') \le k\sigma_{\min}^2(Q)/\sigma_{\max}^2(Q)$ (see, e.g., 2.2.33. of [Hanson–Lawson, "Extensions and Applications of the Householder Algorithm for Solving Linear Least Squares Problems", 1969]). Next, $k\sigma_{\min}^2(Q)/\sigma_{\max}^2(Q) \le \|\Lambda_Q\|_1^2/\|\Lambda_Q\|_2^2$ holds because $\sigma_{\max}^2(Q)\|\Lambda_Q\|_1^2 \ge k\sigma_{\min}^2(Q)(\sigma_{\max}^2(Q)\|\Lambda_Q\|_1) \ge k\sigma_{\min}^2\|\Lambda_Q\|_2^2$. Combining these with Corollary 4 gives the claim. Our result, although without a logarithmic term, elegantly captures the spectral dependence and can provide much sharper bounds when the minimum and maximum constrained eigenvalues differ in order. Our results also give interesting bounds when the minimum constrained eigenvalue is degenerate (i.e., equal to 0).

2. (R3) Tightness of result. Previous result (10) is tight *only* when *all* the eigenvalues of the spectrum are of the same order, while our bound is tight when the top $\Theta(d)$ eigenvalues are of the same order *or* when the lower $\Theta(d)$ eigenvalues are of the same order. An example is when the minimum constrained eigenvalue of M is close to zero (say, $1/t$ for some large t), while the maximum constrained eigenvalue is equal to t. Then, (10) would scale as d/nt^2, ignoring log terms. With our result, if the largest $d/2$ eigenvalues are of the same order (say, proportional to t), our result scales as d/n, *regardless of how small the remaining eigenvalues are*. On the other hand, if the smallest $d/2$ eigenvalues are of the same order (say, proportional to $1/t$), our result also scales as d/n, *regardless of the remaining eigenvalues*. In general, our result tolerates extremal eigenvalues in the spectrum, as shown in Table 1. **Log factor**. We would argue that the improved spectral dependence takes precedence over the log factor, which is insignificant in many cases. For example, when k is not very sparse, say $k = \Theta(d^c)$ for some $c \in (0, 1]$, the log factor is not significant and our results can be orders better than (10) when matrix M is mildly ill-conditioned. In the very sparse case, say $k = O(\log d)$, our results still provide meaningful bounds for M with minimum constrained eigenvalue close to 0.

3. (R3) Dimensionality reduction. For the sparse model with M having repeated columns, Reviewer 3's point on reducing the dimensionality via ignoring certain components of θ is valid. However, the result of (10) still depends on the ratio between minimum and maximum constrained eigenvalues of the new "effective" matrix, and leads to a poor lower bound when the minimum and maximum constrained eigenvalues are of a different order. Another point is that for a general matrix M with zero singular values, it is not immediately clear if the same dimensionality reduction trick can be used. This is because of the sparsity constraint on θ (namely, $\|\theta\|_0 \le k$); for example, an approach one may try is to use an orthogonal matrix U and map $\theta \mapsto U\theta$ and $M \mapsto MU^\top$ so that the matrix $MU^\top$ has repeated columns or zero columns, and proceed to ignore certain components of $U\theta$. Yet, in general restricting $U\theta$ to satisfy $\|U\theta\|_0 \le k$ does not necessarily imply $\|\theta\|_0 \le k$, and hence, we cannot map the problem to one where we can easily use the same dimensionality reduction method on $(M, \theta) \mapsto (MU^\top, U\theta)$. Moreover, in general when the spectrum of M is all positive (with divergent large/small eigenvalues), one cannot use dimensionality reduction to improve the result of (10).

4. (R3) Case $n < k$. It should be noted that the result of (10) depends on the ratio of restricted eigenvalues and hence cannot be used to give a meaningful lower bound when $n < k$ due to a degenerate minimum restricted eigenvalue of 0. From this perspective, it is still interesting to have reasonable lower bounds for this situation. Our results accomplish this—despite the trivialness of the error, our method successfully provides a tight lower bound for the Gaussian design.

5. (R3) Novelty with respect to [23]. We consider the prediction scheme as opposed to the estimation scheme considered in [23]. In terms of technical novelty, the key step in [23] requires a single-letterization bound of the Fisher information (i.e., Lemma 9 of [23]). The prediction problem is fundamentally different (and often most important in practice) and the same single-letterization cannot be applied. In our setting, we developed new bounds for the expected Fisher information (i.e., Lemmas 10 & 11), with proofs essentially orthogonal to [23]; see supplementary for details.

6. (R4) Improvement with respect to [19]. We would like to note that the results of [19], like (10), depend on the ratio between smallest and largest constrained eigenvalues (defined as "$\tau[\cdot]$" in Theorem 2 of [19]). Hence, the results of [19] suffer from the same (undesirable) structural dependencies as (10). Our bounds overcome this dependence, and we have showed evidence that our bound provides nontrivial improvements in the above points 1. to 4. and Section 2.1.

7. (R2) Other extensions and intuition. In general, one may derive minimax bounds for other settings such as subexponential designs and heavy-tailed designs with spectral concentration results. The gap of Lemma 1 comes from L_2 bounds arising from bounding entropy with the entropy of a Gaussian having same variance; in fact, our entropic minimax bounds can be extended to other norms (not only L_2) via tools from rate distortion theory (see, e.g., [17]). We will add remarks on this point in any revision.