We would like to thank all the reviewers for their incredibly thorough reading and many comments. We are encouraged to see that all agree the paper is on an important problem, well-written, and the method itself is highly effective.

General Comments: R1 asks about the theoretical comparison with KLSU19, whose bounds we match. Indeed, it is impossible to surpass the sample complexity bounds of KLSU19, as they prove matching lower bounds (see also [5]). That said, our focus is going beyond the theory, to make practical and realizable tools for this setting. Our algorithms are a novel approach based on shrinking confidence sets, and this new approach was crucial to achieve this goal. KLSU's mean estimation algorithm is essentially the KV baseline, which we beat by substantial margins. Their covariance estimation algorithm was designed for theory, and we found it impossible to achieve non-trivial accuracy (line 248). As such, ours is the first effective algorithm for private covariance estimation.

Several reviewers ask about the role of the parameter t, the number of iterations, and R4 asked about splitting the privacy budget across these rounds. Roughly speaking, a larger t allows for weaker "prior knowledge" (i.e., a larger R or K), at the cost of spending more privacy budget before the crucial final iteration. The theoretically principled way to choose t is $\Theta(\log R)$ or $\Theta(\log K)$. Of course in practice, the picture is not as clear-cut. For mean estimation, we found choosing t to be rather large (i.e., 10) was robustly effective in all scenarios. For covariance estimation, the effective choice of t seems to be more setting-dependent. The effect of t is explored in depth in the supplement, throughout, but with a particular focus on the effect of increasing R and K (bottom of page 11 and 15). As for the privacy budget, the majority of the privacy budget should be allocated to the final round as it plays a special role in the algorithm – it provides the point estimate which is returned, while previous rounds only shrink the confidence interval. We explored several splits and found this was adequate for strong performance, practitioners may tune further for improved accuracy. We note that, given a t and split of the privacy budget, the width of the confidence intervals for mean estimation are computable in advance. Therefore, one can optimize over these parameters in advance, we will describe this in the final version (along with a caveat that this doesn't account for bias introduced from aggressive clipping). We will address readability of figures in the final version. (In particular, we did not have access to a printer due to stay-at-home policies.) If anything is unclear beyond this, we would be happy to add more discussion/exploration.

More specific responses follow.

R1: We note that $t = 1$ corresponds to the private baseline (commonly known as "Analyze Gauss," see also line 246, caption of Figure 8), and there are no other effective approaches known.

R2: See above for discussion of lower bounds and setting t. Sheffet's paper assumes a bound B on the data and pays linearly in this parameter – as our focus is minimizing the cost when B might be large, Sheffet's method would not be competitive in our setting. We will add discussion to this effect.

R3: Thanks to R3 for the numerous comments, due to space restrictions we address a subset here. We are glad the reviewer found appreciated the intuition and ideas conveyed in the body.

Our method is not restricted to Normal distributions (in fact, the algorithm, theorem, and proof for the sub-Gaussian case are unchanged), or even sub-Gaussian ones – all we need is the ability to derive tail bounds or confidence intervals for the class of interest. Given this, only superficial modifications are required (in particular, changing γ and r' in lines 2 and 4 of MVM). The algorithms, theorems, and proofs are otherwise effectively the same (but would give different rates depending on the form of the tail bounds). We will add discussion on this. On other topics: "None of these points" is indeed a high probability statement. The tilde notation disregards log factors. Details related to clipping thresholds and aggressive shrinking are documented in our submitted code. For our non-Gaussian experiments, yes, we used the same Gaussian bounds (testing the effect of model misspecification). Regarding the exclusion of $t = 1$ for clarity in some of the plots, in those cases, the $t = 1$ line is so far apart from the ones for higher t, such that the latter become difficult to tell apart. On the private projection: the principal component vectors (the "structure" we extracted) are output privately, the points are plotted only for visualization (releasing them would not be private, as correctly identified).

R3 and **R4**: SYMQ advantaged: The second half of line 213 says: with the same n, SYMQ fails catastrophically (Figure 4 of supplement). While we could declare superiority at this point, we allow SYMQ twice as many samples to permit further comparison (which we still outperform).

R4: Style comments: We respectfully prefer the current structure, as we believe the theorem statements placed upfront would be confusing or uninformative. This style of argument leading to a theorem is common in mathematical prose. Regarding large enough n, this refers to the bound in the sample complexity. Details of how this bound is applied are spelled out in the proof in the supplement.

Computation: The running time is linear in t, we can add a comment to this effect. Values of t which are too small will result in poor accuracy, see theorem statements and Figure 4 of the supplement.

Non-Gaussian non-symmetric distributions: See first comment to R3.