Thank you to all the reviewers for their detailed comments. We will make sure to address all minor feedback in the final version of the paper.

R1

- *What do the authors conjecture is happening in the wine data set with $\epsilon = 0.01$?* For small datasets like wine, when using small values of epsilon the output contains almost no information about the input. Instead, utility is determined mostly by coincidental alignment between the mechanisms noise distribution and the specific dataset at hand.
- *If the data truly is sparse, do the authors conjecture that one should use a technique like Chaudhuri et al. or Wang and Xu, that is tailored to sparse data?* Yes, if there is public knowledge that the matrix will be sparse, one should use the techniques of Wang and Xu. However, it is worth mentioning that in practice covariance matrices are generally not sparse (although they may be low rank).
- *Why is it the case that the observed number of samples is much greater than the bound given in Kent et al.?* This is probably explained by the fact that the scale of d is not large enough. Kent et al. provide only an asymptotic bound on the number of samples.
- *Im unsure what the authors mean by optimal in line 58.* We meant that the parameters optimize our error bound, but youre right that this is not optimal in general, and the wording is misleading. We will make this clearer in the final version of the paper.

R2

- *When n is somewhat large, in particular, the new guarantees are worse than that of Analyze Gauss.* As touched upon in the introduction, we view large n as coinciding with low sensitivity and therefore easily attainable privacy. Indeed, for a fixed epsilon as n becomes large all reasonable mechanisms begin to perform well, and the theoretical distinctions vanish in practice. In contrast, practitioners can face difficulties when attempting to deploy mechanisms on hundreds or thousands of examples, for instance in the social sciences or medicine. We will make this fact more clear in the paper.
- *The assumption that the columns have bounded L_2 norm is quite restrictive and somewhat unrealistic.* We emphasize that a bound B on the column norm can easily be surfaced to the theorems (rather than assumed equal to one), and will do so in the final version of the paper. Some kind of scale dependence is, of course, fundamental. We agree that an analysis incorporating the conditioning of the matrix could make for interesting future work.
- *When the data matrix is low rank (or approximately low rank), it appears that this algorithm gives weak guarantees.* If the rank k is known beforehand then our algorithm is easily adapted by terminating after k (rather than d) iterations. Furthermore, rank is 1-sensitive and therefore can be easily estimated in a differentially private manner by adding Laplace noise to the rank of the true matrix.

R3

- *Lemma 1 has a utility bound with RHS depending on $\widehat{w}_\alpha$* A dependence on either $\|w_\alpha\|$ or $\widehat{w}_\alpha\|$ is necessary since, intuitively, the scale of w_α will affect the scale of the error. The easiest way to see this formally is to inspect the equations after line 399 in the Appendix. The bound is in fact symmetric, and can be written in terms of $\|w_\alpha\|$; however, we chose the stated bound because $\|\widehat{w}_\alpha\|$ can be computed from the private matrix and therefore is known to the practitioner.
- *Publicly available code may help the significance of this paper.* We will be releasing an open source version of our code soon.