Addressing all reviewers: We thank the reviewers for their careful review of our paper. Following the reviewers comments, we: (1) have made the write-up of the paper better, (2) fixed all typos and comments from the additional feedback sections of the review, (3) added more graphs that compare our result to prior work, and additional graphs that compare time vs accuracy against uniform sampling to emphasize our improvements as requested, (4) added an explanation to our definitions, theorems and figures to make things more clear, specifically speaking Definition 1 and Figure 1, (5) discussed the differences between our result and previous works.

Reviewer 1: We thank the reviewer for his constructive review and detailed comments.

Q: Many such contributions focus on one specific problem, but in this paper we deal with a fairly general class. This is challenging non-trivial work. **A**: This is the main contribution of the paper. We thank the reviewer for pointing this out.

Q: I am missing a critical discussion of the necessities of the assumptions in Definition 1 and the limitations that they impose. **A**: Added to the introduction. In short: Assumptions (i)-(iii) in Definition 1 are used to reduce the problem to dealing with an “easier” pair of functions where the first is a convex “bi-log-log-Lipschitz” function “g” and the second function “h” being independent of the input points. Finally, assumption (iv) ensures that the ellipsoid which encloses the level set of g (the convex function) exists and to be centered at the origin to avoid dealing with the center throughout the technical proofs. Combining the properties associated with the level set of g (the convex function) and Assumptions (i)-(iv), allow us to bound the loss function from above and below by the mahalanobis distance with respect to the enclosing ellipsoid. This is due to the fact that the level set encloses a contracted version of the ellipsoid that encloses the level set of g. As future work, we aim to relax the aforementioned assumptions and use the center of the enclosing ellipsoid, since we think that this step will widen the applicability of our framework. What do you think?

Q: The SVM result depends on regularization and additionally on the structure of the data which seems quite restrictive to rely on both relaxations. **A**: The SVM problem formulation presented quite an obstacle when trying to show that the loss function is near convex. As presented in the proof of Lemma 29 in the supplementary material, we had to split the query space into two parts, where the first lead to having a bound on the sensitivity as a function of the structure of the data, while the latter was handled using our f-SVD. We think that relaxing our assumptions in Definition 1, will yield a tighter upper bound on the sensitivity of each input point in P and in turn on the total sensitivity.

Q: Can you explain your solution to this intuitively, especially why it is sufficient to use the convex ℓ_1 level set instead and "outsource" the power of z. **A**: First let $\tilde{P} \in \mathbb{R}^{|P| \times d}$ be a matrix such that every point in P is a row in $\tilde{P}$. Now observe that for any $z \in (0,1)$, $\sum_{p\in P} \left|p^T x\right|^z = \left\|\tilde{P}x\right\|_z^z \geq \left\|\tilde{P}x\right\|_1^z$, where the equality holds by definition of $\tilde{P}$ and the inequality follows from properties of norms. Using this we can bound the denominator of the sensitivity for each $p \in P$ from below by $\left(\sum_{p\in P} \left|p^T x\right|\right)^z$. This allows us to use the f-SVD with respect to the ℓ_1-regression problem as done.

Q: Corrections are needed for related work. **A**: All suggested corrections were accepted and added to the related work, we thank the reviewer for pointing this out.

Reviewer 2: We thank the reviewer for the constructive suggestions. We incorporate them in the new version as explained below, and hope to raise the final scoring.

Q: Do we assume that g and h are given? **A**: As for the applications specified in our framework, g and h were established theoretically, and we assume they are given. As part of our future work, we aim to generalize this framework as well as having a polynomial algorithm for finding the proper g, h for each “near-convex” function (see Definition 1).

Q: Moreover, some more explanation on finding matrices D and V would be helpful. **A**: We have added intuition and explanation regarding this. Basically, we explain (in a more detailed fashion) that D is a diagonal matrix where its diagonal entries are proportional to the Löwner ellipsoid’s axis lengths, while V represent the basis of the ellipsoid.

Q: It might be helpful to discuss an example after Definition 1. **A**: Added for the logistic regression problem.

Reviewer 3: We thank the reviewer for his supportive comments and very helpful suggestions.

Q: Figure 1 is nice. It would be good if they bring it to the beginning of the paper and explain the intuition better. **A**: We thank the reviewer for this very nice suggestion. We changed its place and explained the intuition better.

Q: They can sell better this nice results. **A**: We are not so good at selling, but hope that the strong result will help us here. We thank the reviewer also for the selling suggestions.

Reviewer 4: We thank the reviewer for the supportive scoring and for the detailed review.

Q: The Figure captions are in a font that is too small!!!. **A**: We thank the reviewer for pointing this out. We changed the fonts as requested.