Relation to prior work, novelty: • To answer a major concern of all reviewers: we study approximation of $\hat{\beta}$ that are an order of magnitude more precise than the commonly studied risk bounds. Lasso risk bounds are of order $s\log(p/s)/n$ while our results are $(s\log(p/s)/n)^{3/2}$; similarly for group-lasso in Section 3.3. Results at such scale are not available in prior work, at the sole exception of [21, Thm 5.1] which is restricted to Lasso and squared loss, cf. lines 190-198 for comparison. • Results are such scale **are thus novel**. Proof techniques are also novel: a careful application of powerful generic chaining results from [15, 31] is needed to obtain results at this scale, cf. the theorems of Section 6 and their proofs. • Results at such finer scale are useful to characterize the risk exactly (as opposed to upper/lower bounds up to multiplicative constants), cf. Section 4; or to construct uncertainty quantification results in the form of confidence intervals, cf. Section 5. Uncertainty quantification is a major challenge in high-dimensions and calls for results at a finer scale such as those from the submission. • We invite the reviewers to revisit their scores in light of this.

Reviewer 1: *"(a) tools and techniques used in the proofs are pretty standard, and do not contain novel ideas."* ⇒ cf. line 1-10 above. *"(b) - Even though the authors ..., I am unable to imagine more general applications of their results in Machine Learning. (...) for any new problem of interest (...) need to think from scratch to quantifying the set T and controlling its rademacher complexity "* ⇒ Sets T and their gaussian complexity have already been studied for most high-dimensional estimators by many authors: (Group-)Lasso, Slope [6], Nuclear norm [32], tensor norms, etc, see surveys [4, 16, 23, 32]. For all these examples, set T is already available and extension of our results to such penalty is straightforward–we'll clarify this. *"Further the estimator in (4) may not be computationally easier"*: the quantity in (4) is not an estimator since it depends on β^*; it is an approximation of $\hat{\beta}$ that can be used for applications in Secitons 4-5, including confidence intervals. *"(c) Writing: (...)"* ⇒ we'll clarify the writing and fix the typos as suggested.

Reviewer 2: *"main results of the paper is novel (...) applications on lasso, group lasso are already studied in literature. (...) unknown whether there are applications s.t. the theory leads to new results beyond those in literature."* ⇒ cf. line 1-10. *"The formula of the first order expansion is not computable except in some special situations. Applications that can lead to new results beyond those already in the literature will be useful to illustrate the value of such a formula."* ⇒ cf. line 18. *"In Prop 5, T_n exists but its computational formula is not available. How could it be used for inference?"* ⇒ Proving $\sqrt{n}(\hat{\theta} - a^\top \beta^*) - T_n \to 0$ in probability and T_n has t-distribution with n degrees-of-freedom yields confidence intervals: $\mathbb{P}(\sqrt{n}|\hat{\theta} - a^\top\beta^*| \leq 1.96) \approx 0.95$, hence $a^\top\beta^* \in [\hat{\theta} \pm 1.96 n^{-1/2}]$ (asymptotically) with probability 0.95.

Reviewer 3: *"(...) results are only usable under the strong assumption of the existence of a function ψ (...)"* ⇒ Our construction η generalizes $\beta^* + \frac{1}{n}\sum_i \psi(X_i, Y_i)$ in high-dimensions, proving that such approximation exists for several $\hat{\beta}$ in line 15 above. *"technically speaking, (...) third item of (A1) which seems to constrain ℓ to be quadratic (see my comments below). (...) not clear how this assumption is really "called" in the main results (...)"* ⇒ ℓ need not be quadratic, cf. line 52-54 below. Third item of (A1) is the Restricted Strong Convexity (RSC) assumption from [32], required to obtain risk bounds for logistic lasso of order $s\log(p/s)/n$. It is used in the main theorems in Section 6 to bound certain empirical processes. The constant B_3 is explicit in these proofs, which allows to track where third item of (A1) is used. *"(...) difficult to assess the limitations and width of application of the work: which problems can it handle, which not? What would be next hurdles to break? Are there definitely problematic issues for follow ups?"* ⇒ cf. lines 14-17 above. *""certain smoothness assumptions" → such as?"* ⇒ Differentiability of the loss in [18,24] and stochastic equicontinuity (a weaker form of differentiability) in [35, 36]. We'll clarify this. *"please specify the definition of the derivatives of $\ell(.,.)$. This in passing imposes that ℓ be properly differentiable. This is not the case for the l1 norm. How is that dealt with?"* ⇒ Derivatives of $\ell(y, u)$ are always with respect to u. We'll clarify this. The data-fitting loss (squared, logistic) is required to be differentiable, but not the penalty h (ℓ_1-norm, ...). *"the overall setting also assumes a concentation effect of the argument of ℓ as $n \to \infty$. This is known not to be the case when $n, p \to \infty$ together for e.g., $X_i \sim N(0, I)$. Thus Taylor expansions are to no avail in this case. Since it is proposed here to let p grow possibly large, it would be worth discussing what scaling for 'p' is allowed and how it relates to the data statistics."* ⇒ The submission does allow for $n, p \to \infty$ together: Lasso requires $r_n \asymp (s\log(p/s)/n)^{1/2} \to 0$ cf. (23), Group-Lasso requires $r_n \asymp \{(sd + s\log(M/s))/n\}^{1/2} \to 0$, cf. Lemma 3.5. The required concentration is obtained by a careful application of powerful generic chaining results from Dirksen [15] and Mendelson [31] that let us obtain concentration results uniformly over T; cf. Section 6 and the corresponding proofs. *"what is T in third display of (A1)?"* ⇒ Third ineq. in (A1) is the Restricted Strong Convexity of [32], T is the restricted cone. *"(...) the last [assumption in (A1)] possibly stringent. (...) isn't the denominator simply $\|u\|_K^2$? (...) this implies (...) that no eigenvalue of K vanishes, thus essentially that ℓ'' is bounded below (...) equivalent to saying that only quadratic costs are allowed? This would be a major issue"* ⇒ $\|u\|_K = \|K^{1/2}u\|$ and $K = E[\frac{1}{n}\sum_i \ell''(Y_i, X_i^T\beta^*)X_i X_i^T]$ defined in line 72 is an expected (population) quantity. ℓ'' needs **not** be bounded below, only the population matrix K. For logistic loss, the assumptions hold (Prop 2.2) although ℓ'' is not bounded from below. Third inequality in (A1) is Restricted Strong Convexity of [32], a common assumption for analysing logistic lasso/group-lasso. *"(...) appropriate to comment on Th2.1 (...) differ from prior work, what's new/interesting in it?"* ⇒ cf. line 1-10 above. *"I do not understand in Prop2.2 why (8) holds for some $B_3 > 0$. Doesn't $l''(y, u)$ tend to 0 with $u \to \infty$ for instance?"* ⇒ Third ineq in (8) involves K and Σ which are both expectations. $\ell(y, u) \to 0$ as $u \to \infty$ is OK as long as $\|\Sigma^{1/2}\beta^*\| \leq 1$ (or $\leq C$) (cf line 109). This is common assumption in logistic lasso, e.g. Prop 6.2 in [1]; though our proof is not restricted to Gaussian X_i–we'll clarify.