We would like to thank all the reviewers for taking the time to understand our work, and for their helpful comments!

To Reviewers #1 and #2: We thank both for raising the important question of what the contributions of this work are beyond merely combining CLM and JO. Recall that the primary technical innovation of our work was to show that we can estimate skewness with an SDP and filter the data *without needing to round the SDP*. We clarify below that filtering without rounding is not just for aesthetic considerations or runtime savings: rather, it is an essential workaround to a fundamental complexity-theoretic barrier that arises in our setting but in neither CLM nor JO in isolation. Indeed, as we explain below, with this barrier in mind it should actually be quite surprising that our algorithm works at all.

Reminder of JO's approach Recall the key subroutine in JO was to estimate "skewness" of the weighted dataset, i.e. to estimate $\max_{v\in\{0,1\}^n} |v^\top Av|$ for some data-dependent matrix A. And the seminal paper of Alon-Naor on approximating the cut-norm gives an SDP *and a rounding scheme* that together produce v whose objective value is within a constant factor of optimal. We emphasize that having access to a rounded solution, that is, an actual *bitstring* v as opposed to a psd matrix, is essential for JO to be able to follow the conventional filtering analysis.

A hardness of approximation barrier As mentioned in our technical overview, instead of optimizing $|v^\top Av|$ over $v \in \{0,1\}^n$, we need to optimize $|v^\top Av|$ over $v \in \mathcal{V}_\ell^n$. This is a far more challenging optimization problem and closely connected to things like sparse PCA. Indeed, if $A' := T^\top AT$ where T is the matrix which maps v to $(0, v_2 - v_1, v_3 - v_2, ..., v_n - v_{n-1})$, then this problem becomes that of optimizing $w^\top A'w$ over ℓ-sparse vectors $w \in \{0, \pm 1\}^n$ whose nonzero entries are alternating in sign. Modulo the alternating sign condition, this is at least as hard as densest ℓ-subgraph. Notably, Manurangsi (STOC '17) showed that under the Exponential Time Hypothesis, it is even impossible to efficiently achieve any sub-poly(n) factor approximation for this problem.

Our workaround Existing works on filtering for robust statistics all require filtering along univariate projections of the data, but in the unavoidable absence of a rounding scheme in our setting, the only approach one can try is to just "project" along the solution to our SDP relaxation of $\max_{v\in\mathcal{V}_\ell^n} |v^\top Av|$ at each step, and this is the approach we take.

In light of the above barrier, the fact that our algorithm works should be quite striking: the hardness of approximation result suggests that the integrality gap of our SDP relaxation should be horrible. In particular, the norm $\|\cdot\|_{\mathcal{K}}$ induced by this relaxation might be very distorted relative to the $\mathcal{A}_\ell$ norm we actually care about.

The saving grace is that "regularity" (in the sense of Definition 3.2 in the body) still holds with respect to this distorted norm, which, together with the careful design of the SDP (e.g. so that we still have estimates like Corollary 2.5 in the supplement) allows us to deduce "soundness" (in the sense of Lemma 3.5 in the body) with respect to this norm.

In fact, not being able to round introduces complications even in the "easiest" parts of the traditional analysis of the filtering framework. For instance, even the statement that a large enough subset of the uncorrupted samples will satisfy "ϵ-goodness", which in JO followed from standard tail bounds for sub-exponential random variables, requires proving new concentration inequalities (we explain this in Appendix C). The proof of soundness becomes similarly delicate.

Unrelatedly, another innovation over CLM is that, to achieve $O(\ell^2)$ sample complexity, we needed to significantly tighten their metric entropy bound: even after implementing the above workaround, CLM's bound would have gotten some impractical polynomial dependence on ℓ. As alluded to in Remark 2.4 of the body, in order to get the guarantees of Lemma 3.4 in the body, it was necessary to modify the convex relaxation they considered.

To Reviewer #2:

1. "E.g., the algorithm (in particular, the SDP) is not fully specified in the main body." The full SDP is given in Definition 2.3 in the body, and the references to specific constraints in section 2 are references to the constraints in Definition 2.3. The full algorithm is then given in Algorithm 1 and 2.
2. "But the notion of "structured distribution" in terms of piecewise polynomial approximations considered here is not as convincing." The notion of piecewise polynomials (or splines) we work with is a standard one that has been studied extensively in statistics, see the references in Section 1.2 of the main body. It strictly subsumes classes of "structured" univariate distributions like monotone, multi-modal, concave, convex, log-concave, Gaussian, Poisson, binomial, monotone hazard rate, Besov, etc., and arbitrary mixtures of such distributions.
3. The SDP-based algorithm doesn't seem practical (it takes several minutes for $n = 128$). For such small n, it is not clear that the sample complexity savings in terms of n are significant. Note that it is still possible to see the expected factor of n/ℓ in sample complexity savings. For reference, in JO (see e.g. their Figure 1), for alphabet size 100 and corruption fraction 0.4, they need to draw $100/(0.4)^2 = 625$ batches. In contrast, for $\ell = 10$ and $n = 128$, we only draw $\ell/(0.4)^2 \approx 63$ batches. That said, we completely agree that getting a more practical non-SDP-based algorithm that can work for much larger n is an important direction for future work.