We thank all four reviewers for the careful reading and detailed feedback.

R1, *"Relation to previous work/bounds"*. The algorithms in [4, 5, 6] require to touch all N points before the support $x_1, ..., x_{n+1}$ of $\hat{\mu}$ is found; in contrast, Algorithm 2 can finish much earlier by smart sampling; e.g. for $d = 2$, Algorithm 2 terminates after at most 2 steps (Theorem 4), which is a big reduction in the big data regime when $N \gg 2$; for $d > 2$, the analysis is more subtle since pathological behaviour can occur with small probability, but generically the same behaviour applies. We will add a table with runtimes to further highlight this point, which we believe to be a main contribution of this work. The complexity bounds of Algorithm 1 give here only very conservative bounds for Algorithm 2; see also the discussion starting in line 152.

R1, *"Approximation?"*. Thank you for the suggestion! Firstly we should note that this would exclude important applications, e.g. if $f_1, \dots, f_n$ are monomials up to given degree, then $\hat{\mu}$ integrates every function exactly that is in the linear span of $f_1, \dots, f_n$, i.e. every polynomial up to given degree. This is for example essential to derive error bounds for cubature, see [4, 5]. However, for other applications an approximation would suffice and we have thought about this previously but, put simply, we do not know how one could calculate an approximation with significantly less computational cost. The bottleneck remains in finding the support of $\hat{\mu}$. We believe approximate solutions are an interesting open research question; in fact, already making precise what is meant by a "good approximation" allows for many possibilities since one can choose many different metrics for measures to quantify when μ is close to $\hat{\mu}$ as well as many different ways how an excess number $k > n+1$ of supporting atoms $\{x_1, \dots, x_k\}$ should be penalized.

R1, *"choice of n-points random"*. In Algorithm 1, the points are simply drawn uniformly at random but not in Algorithm 2; for the latter only the first selection is random, and subsequent choices aim to maximize the volume of the (negative) cone which is one of the main reasons for the performance of Algorithm 2. The intersection between the cone and the negative cone is always empty (except for the origin) in both algorithms.

R1, *"Approximate Least Squares"*. We included all experiments done in [6] since it was a NeurIPS paper, thus provides a natural and competitive benchmark. But yes, the measure reduction is in this application only advantageous if exact solutions are required, which is often but not always the case, and we will add a sentence to note this.

R1, *"Complementary"*. In Appendix E we show how [4, 5, 6] can be combined with Algorithm 2. The hybrid algorithm has a high chance of requiring very few steps to find the support, but for pathological cases it can rely on the deterministic procedure of [4, 5, 6].

R1, *"Exposition/Pictures"*. R2, *"proofs in main text"*. Yes, we will add pictures and the extra page also allows to move proofs to the main text and write more about the intuition. Figure 6 shows the situation in dimension 2 and we will expand on this further.

R2, *"Reset strategy"*. The effects of the reset strategy can only be noticed when N and n are close, and n is relatively high. This is the reason why the lines are indistinguishable in the regime of Figure 1. However, the reset strategy always guarantees the convergence of Algorithm 2 (in addition to being optimal in the sense discussed in line 188ff); e.g. without it, a pathological case of cycling through a "sequence" of cone bases could occur, which is the reason why we always apply it, even when N and n are not close or n is relatively low.

R2, *"Explicit optimization"*. Thank you for the suggestions and references! We were not aware of [A, B] and we will include both. [B] implicitly aims to maximize a volume, similar to what we try to achieve (see step 7-9 in Algorithm 2) although we are directly motivated by polygon geometry going back at least to [*Sommerville, 1927, "The relations connecting the angle-sums and volume of a polytope in space of n dimensions"*]. Ideally, we would like to maximize directly the volume but this is a longstanding open question in discrete geometry, see e.g. [*Bollobas, 1997, "Volume estimates and rapid mixing"*], or [15] for an overview; the proposed "angle maximization" in Algorithm 2 serves as a computationally efficient proxy to this problem (although it is not perfect proxy; see pathological cases discussed in line 150ff and how to deal with them in line 250ff). [A] has a similar motivation, but works with gradients in continuous space so it is not obvious to us how to apply it for the subset selection directly, however, combined with a continuous relaxation we believe that this would be an interesting research question.

R3, *"basic techniques/applications"*. Yes, we combine fundamental results from algebra, discrete geometry, and probability to derive a new algorithm. The approach of geometric sampling by exploiting cone geometry is very different from previous approaches [4, 5, 6] to this classic problem; independent of this, even robustness (Proposition 3) was not established for any of the previous algorithms [4, 5, 6]. As for applications, in addition to the ones done in Section 4, and the related work cited in the introduction, we can refer to [5] to approximate integrals, while in [4] the reduction procedure yields numerical methods for Stochastic Differential Equations and Partial Differential Equations. The design and analysis of an effective measure reduction algorithm is the focus of this paper.

R4, *"figures/cone/log-log scale"*. Thank you for the kind words and remarks! Yes, it should be a negative cone; we will change the figures as suggested. In line 230 we refer to Figure 2 (bottom) where the scale should already be log-log.