Dear Reviewers, we thank you for appreciating our work and the constructive feedback, and address your concerns below.

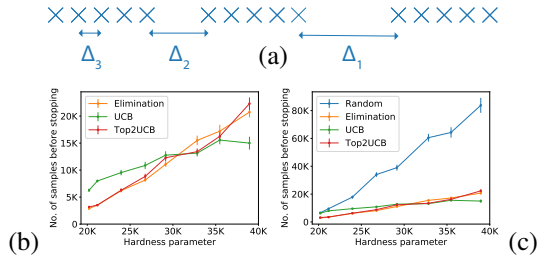

Figure 1: Stopping time experiments

Experiments: (Rev. 1,2,3) We showed empirical mistake probability plots because in reality one cannot always wait for an algorithm to run until termination (similar concerns and plots are shown in Bubeck 2009, Jamieson 2013). We ran stopping time experiments and show our results in Fig. 1b,c, where we plot the empirical stopping time against the theoretical sample complexity (Thm. 2) for different arm configurations. We choose the arm configuration in Fig. 1a containing three unique gaps - a small gap Δ_3 and two large gaps $\Delta_2 < \Delta_1 = \Delta_{\max}$. The hardness parameter is changed by increasing Δ_2 and bringing it closer to Δ_1. We see a linear relationship in Fig. 1b which suggests that the sample complexity expression in Thm. 2 is correct up to constants. In Fig. 1c we include random sampling and see that our adaptive algorithms require up to 5x fewer samples when run until completion. Fig. 1b suggests that UCB may outperform Top2UCB (Rev. 1). Fig. 1c shows that the adaptive algorithms are robust to the hardness of the problem and always outperform random sampling, and the gains increase with hardness (Rev. 2). We used a lower bound based stopping condition for Random, Elimination, Top2UCB, and set $c = 5$ in the UCB stopping condition (value of c chosen empirically similar to Jamieson 2013).

Motivation, Applications and Extensions: The MaxGap bandit can be used to *efficiently* find the "good" set, where good is defined using the max-margin criterion instead of being pre-specified as in the top-k best-arm problem. This is equivalent to clustering the arms into 2 sets, and just like for any clustering algorithm, one can construct adversarial distributions where the clustering returned by the algorithm is different from a desired clustering. The extensions identified to address these issues such as the PAC framework (Rev. 2,3), balanced or constrained clustering (Rev. 3), are challenging, and great avenues for future work. For example, the PAC framework would help in stopping early if two large gaps are within ϵ of each other. In best-arm identification, the PAC problem is solved by constructing a lower bound on the *mean* of *every* arm and stopping as soon as the difference between the highest upper bound and the lowest lower bound of the active set is less than ϵ. In the MaxGap problem, it is difficult to construct a lower bound on the *gap* (as opposed to mean) of *every* arm due to interaction effects; we can only construct a single lower bound on the maximum gap, and this makes the PAC extension non-trivial. We would like to highlight though that with ϵ-close large gaps, our algorithms correctly focus on the large gaps (Fig. 7), and hence stopping early and clustering according to the empirical means will yield a reasonable clustering. We can similarly discuss the complications with other extensions. As pointed out by Rev. 2, we believe this work can act as a solid foundation for this and other extensions such as constrained clustering, clustering with more than 2 clusters (Rev. 1), clustering without specifying number of clusters, and adaptive multi-dimensional clustering.

Naive Application of Best-arm Analysis: (Rev. 3) While the algorithms we propose are conceptually similar to existing bandit algorithms, the analysis of MaxGapUCB is far from a trivial application of the UCB analysis. On a high-level, in best-arm identification, the number of samples of a sub-optimal arm i is bounded by observing that

$$\text{Arm } i \text{ is pulled} \Rightarrow \mu_i + 2c_{T_i(t)} \geq \hat{\mu}_i + c_{T_i(t)} \geq \hat{\mu}_{(1)} + c_{T_{(1)}(t)} \geq \mu_{(1)} \Rightarrow 2c_{T_i(t)} \geq \mu_{(1)} - \mu_i = \Delta_i. \quad (1)$$

The last inequality *directly* bounds the number of samples $T_i(t)$ of a sub-optimal arm i. In MaxGapUCB, the gap upper bound is obtained using the confidence intervals of two arms, and the fact that a sub-optimal gap (i, j) has the highest gap-UCB implies that

$$(\mu_j + 2c_{T_j(t)}) - (\mu_i - 2c_{T_i(t)}) \geq (\hat{\mu}_j + c_{T_j(t)}) - (\hat{\mu}_i - 2c_{T_i(t)}) \geq \Delta_{\max} \Rightarrow 2(c_{T_j(t)} + c_{T_i(t)}) \geq \Delta_{\max} - \Delta_{ij}. \quad (2)$$

Thus unlike the reasoning in (1), the number of samples from arm i is coupled to the number of samples from arm j, and $T_i(t) \to \infty$ if j is not sampled enough. We show in our analysis that this cannot happen in MaxGapUCB. Furthermore, any arm i is coupled with multiple other arms since the ordering of the arms is unknown, and may have to be sampled even if its own gap is small - a phenomenon absent in best-arm analysis because of the independence of the arm means. In our proof, we account for all samples of an arm by defining states the arm can belong to (called levels), and arguing about the confidence intervals of the arms in unison. Please refer to the outline of the proof in Fig. 8 and the complete analysis of MaxGapUCB on pages 16-22.

Modern Analysis and Lower Bound: (Rev. 3) Kaufmann's Track-and-Stop algorithm estimates the optimal proportion of samples needed from an arm based on the instance-dependent lower bound and uses it to guide their sampling strategy. In the lower bound, they obtain the "closest" alternate bandit model having a different best-arm by changing the means of just two arms (see proof of Lemma 3 in their paper). However that characterization of the "closest" alternate model does not hold in the MaxGap bandit problem, and obtaining an instance-dependent lower bound and an asymptotic expected sample complexity bound would require new lower bounding techniques.