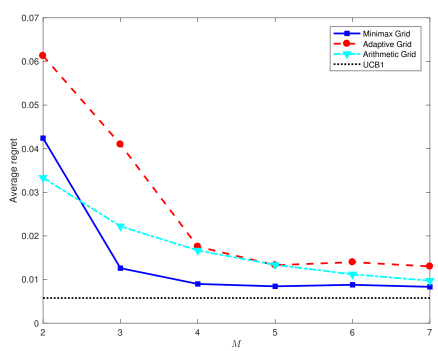
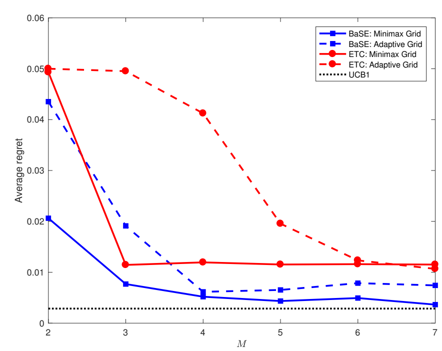
We are grateful to the reviewers for the comments. Below we respond to each and every point of the reviewers (where we refer to the same set of references in the submission):

Reviewer 1: We thank the reviewer for the positive comments. Our corollaries follow from the fact that $T^{\frac{1}{2-2^{1-M}}} = T^{\frac{1}{2-2^{1-\log_2 \log T}}} = \sqrt{T} \cdot T^{\frac{1}{2(\log T-1)}} = \sqrt{T} \cdot e^{\frac{\log T}{2\log T - 1}} = \Theta(\sqrt{T})$ for the minimax regret when $M = \log_2 \log T$ and $T^{\frac{1}{M}} = T^{\frac{1}{\log T}} = e = \Theta(1)$ for the adaptive regret when $M = \log T$. Hence, under the conditions of Corollary 1, the optimal regrets $\Theta(\sqrt{KT})$ or $\Theta(K \log T)$ are attained within logarithmic factors. The lower bound parts are also similar. Moreover, although the data-driven grid is *no weaker* than the static grid by definition, our result shows that it is also *no stronger* than the static grid for the batched bandit problem in the sense that a static grid suffices to essentially achieve the optimal regrets. Hence, one of our contributions is to show that data-driven grids do not improve much over the static ones, which is the focus of Theorem 3. In the final version we will make these points clearer and more explicit.

We appreciate the reviewer's great suggestions of adding experiments. We have done the following: first, we numerically investigate the regret dependence on parameters (T, K, M, Δ) and the choices of different grids. For example, Figure (a) plots the average regrets of BaSE under different grids as a function of M, as well as the regrets of the centralized algorithm UCB1 without batch constraints [ACBF02]. Here we take $T = 5 \times 10^4$, $K = 3$, $\gamma = 1$, standard normal arms with mean $(0.6, 0.5, 0.5)$, and arithmetic grid stands for the grid with equal spacing. We observe from Figure (a) that a very small number of batches (e.g., $M = 4$) are sufficient for BaSE to roughly achieve the centralized performance under the minimax grid. Second, since policies for batched bandits with $K > 2$ arms are missing before our work, we only compare our BaSE policies with the ETC policies in [PRCS16] for two-armed bandits. Under the same setting above, Figure (b) shows that BaSE achieves lower regrets than ETC. Complete experiments will be in the final version.



(a) Comparison of grids used in BaSE.



(b) Comparison of BaSE and ETC.

Reviewer 2: We thank the reviewer for the positive comments. As suggested, in the final version we will change "adaptive regret" into the "problem-dependent regret", "adaptive grid" into "geometric grid", and "data-driven grid" into "adaptive grid". We will also fix the error in the reference [BPR13] and improve the algorithm pesudocode.

The correct rates for the data-driven grid model are great questions. We believe that the additional M^{-2} factor in Theorem 3 is an artifact in our proof, and we conjecture that the same lower bounds in Theorem 2 still hold. We would also like to emphasize that the M^2 multiplicative gap is at most poly-logarithmic in T in our batched bandit problem, for we may apply known centralized lower bounds (e.g., $\Omega(\sqrt{KT})$ for the minimax regret in the fully online case) whenever $M \gg \log T$. However, we believe that resolving this question will be helpful for other problems with limited rounds of adaptivity, and in the final version we will explicitly state it as an open problem.

Reviewer 3: We thank the reviewer for the positive comments. The adversarial setting with batch constraints is a great question, where we believe that proper definitions of the adversary and adversarial regret would be important. It is not hard to show that when the adversary can choose the rewards in $[0, 1]$ arbitrarily and the cumulative reward is compared with the optimal fixed arm in hindsight, the regret will be at least $\Omega(T/M)$ if $2M \leq K$. Consequently, for small M the regret is $\Omega(T)$, as opposed to our current results in the stochastic setting. Hence, to obtain illuminating results we need to restrict the power of the adversary or change the definition of the regret, which would be interesting future directions.

The analysis of any pre-specified grids is also a great question. We conjecture that our BaSE algorithm is nearly optimal for any fixed grids, and the minimax regret is $\tilde{\Theta}(\sum_{i=1}^{M}(t_i - t_{i-1})\sqrt{K/t_{i-1}})$. Intuitively, in the i-th batch $(t_{i-1}, t_i]$, all remaining arms in the BaSE algorithm have been pulled at least t_{i-1}/K times and thus have suboptimality gap $\tilde{O}(\sqrt{K/t_{i-1}})$ with high probability, then the upper bound follows. The lower bound can be established using Lemma 2, which may be off an M^{-1} factor after choosing Δ carefully. We will also fix the error in the reference [BPR13].