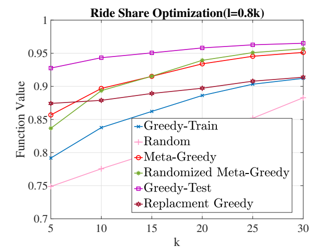
We thank the reviewers for their careful consideration and constructive feedback. Below, please find our responses.

General comments. **GC1:** In order to highlight the novelty of our theoretical results, we'd like to emphasize that the meta-learning objective considered in this paper in *not* submodular and hence providing constant-factor approximations is indeed novel and non-trivial. In the paper, we have provided two different methods: (1) The Meta-Greedy algorithm (Alg. 3) which combines two carefully-chosen deterministic orderings for greedy-selection and leads to a solution that is at least 0.53-optimal (which is indeed not a common result). The proof of Theorem 1 (and Propositions 1,2) introduces new techniques to carefully analyze the interplay between the inner and outer maximization problems in the meta-learning objective. (2) We show that by selecting elements according to a *properly-designed* randomized procedure we obtain in expectation an $1-1/e-o(1)$-optimal solution, where the $o(1)$ term vanishes when l and $k-1$ grow (see Theorem 2 and the probabilistic analysis therein). This result is also novel and non-trivial, since $1-1/e$ is only common for submodular and monotone problems, while our problem is not submodular. Further, besides all the theoretical contributions, another main novelty of this paper is to introduce the first discrete meta-learning framework.

GC2: We'd like to emphasize that we have compared our proposed scheme with one of the main schemes for two-stage submodular optimization (i.e. the Replacement-Greedy method). The comparison and the plot was provided in the supplementary materials, however, in the revised version we will include it in the main body. As illustrated in the figure on the right, our proposed methods lead to a better user-specific solution (on average) compared to the two-stage algorithm (Replacement-Greedy). For the details of this experiment please check the supplementary material of the submitted paper.



Reviewer #1. Q1: The cost of creating m different S_i's. **A1:** Thanks for this careful comment. Please note that in Algorithm 1 these sets can be computed in parallel (as their selection process are independent) to improve the run-time of the algorithm to kn. We'll highlight this point in the final submission. Given that m tasks are involved, any good algorithm would have a complexity dependent on m, however, parallelization could significantly reduce the run-time. **Q2:** Clarity in Alg. 1 and bringing Remark 1 up. **A2:** Thanks for your suggestion: We'll modify Alg. 1 to clarify that m sets S_i are constructed for all possible $i = 1, \ldots, m$. We'll also clarify that Meta-Greedy outputs $m+1$ sets, i.e., S_{tr} and $\{S_i\}_{i=1}^m$, and state Remark 1 earlier after Alg. 1.

Reviewer #2. Q3: Importance and novelty of the theoretical results. **A3:** Please read our general comment (GC1) above. **Q4:** The point of Broader Impact. **A4:** Regarding the broader impact section, we appreciate your comment and will revise this section by better highlighting the fact that our proposed scheme leads to personalized solutions for different users instead of providing a single solution that might one average work well but could perform poorly for some users (e.g. Netflix users). **Q5:** Better use of space **A5:** Following your suggestions, we will include the counter example to submodularity in the main body and will also provide a hint to the proof for Theorems 1,2. These changes will be possible by shortening the introduction and using the extra page (9th page) for the final submission. Thank you.

Reviewer #3. Q6: Same guarantee of Alg. 3 for Algs. 1,2? **A6:** Thank you for this great and insightful question. Alg. 3 performs *strictly* better than each of Alg. 1 and Alg. 2. Indeed, one can provide simple examples where Alg. 1 or Alg. 2 (either of them, not both) obtains a solution which is exactly 1/2-optimal. However, the best solution among Alg. 1 and Alg. 2 will be at least 0.53-optimal. Also, the best performing algorithm among Alg. 1 and Alg. 2 could change for different problems, and this is why we use the maximum of them as the output of our Meta-Greedy method (Alg. 3). We will include some simple examples to better highlight this point. **Q7:** Two-stage approaches and personalization. **A7:** The reviewer is absolutely right that two-stage also allows for personalization. We'll highlight this point and revise the wording of the sentence that the reviewer has mentioned. The main difference is in the two stage approach personalized items could be selected only from a smaller ground set (which is tailored to the training tasks), while in our approach those elements can be selected from the original ground set. Due to this difference our approach could possibly lead to a better personalized solution as illustrated in the figure above (and general comment GC2). **Q8:** Limited computation at test time and room for personalization. **A8:** As the reviewer has correctly pointed out, the proposed approach will be highly beneficial when the number of elements added at test time (i.e., $k-l$) is much smaller than the size of solution set (i.e., k). Please note that this is aligned with the main point of meta-learning which is to find a solution at training time that can be quickly adapted at test time (with very low computation) to the new ask. Indeed, this assumption is realistic, as in most cases, the solutions for different users have several common elements, and only differ on a small number of elements. We hope that the examples in the paper and the experiments could convey this point. **Q9:** Idea: Swapping at test time. **A9:** The reviewer's idea to swap the elements at test time instead of adding elements is very interesting, and it could be a future direction to explore. We'll include it in our concluding remarks.

Reviewer #4. Thanks for your insightful suggestions and careful reading. **Q10:** Comparison with relevant work. **A10:** The closest framework to our setting is the two-stage submodular optimization framework. We have actually compared our proposed Meta-Learning approach with a state-of-the-art two-stage scheme. The comparison (and plot) is provided in the supplementary materials. Please see the general comment (GC2) above for more details. **Q11:** Typo: l not appearing in equation (1). **A11:** Thanks for catching this typo. It should be P instead of l.