Supplementary Material: Incremental Task Learning with Incremental Rank Updates

Rakib Hyder[1], Ken Shao[1], Boyu Hou[1], Panos Markopoulos[2], Ashley Prater-Bennette[3], and M. Salman Asif[1]

[1] University of California Riverside
[2] Rochester Institute of Technology
[3] Air Force Research Laboratory

In this document we provide additional discussion on some questions raised by reviewers that we hope will be informative for the readers. We are also grateful to all the reviewers for all of their questions and comments.

A Relationship between frozen and new factors

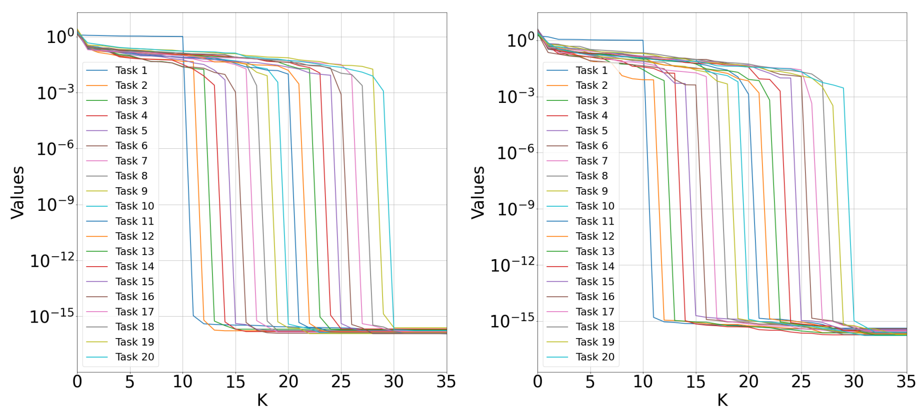
Question: What is the relationship between the newly learned rank-1 weights and the fixed weights learned in the previous task. Are they orthogonal to each other or can the new weights be represented by a linear combination of the old fixed weights?

The newly learned rank-1 factors are not orthogonal to previous factors, but they are linearly independent. To demonstrate this relation, we plot the top singular values for weight matrices in two layers corresponding to 20 tasks in S-CIFAR100 in Fig. 1. The rank of the weight matrices starts at 11 and increases by one for every task. We observed similar trend in other tasks and layers. This suggests that the learned factors are linearly independent of frozen factors.

B Effect of task similarity

Question: The findings suggest that the proposed method benefits from parameter sharing and positive knowledge transfer between tasks, allowing even a low-capacity model to learn comparably well. If the model was trained on a set of controlled task-difficulty benchmarks, the performance/memory metrics would be even more useful. The most important question in this study is how similar the tasks must be for the proposed method to be effective.

Our experiments suggest that a positive knowledge transfer allows low-capacity models to perform well. We performed an experiment by selecting superclasses of CIFAR100 as separate tasks. If all classes in a task become similar (harder classification), the cross-task similarity reduces. We observe ∼60% accuracy for rank-1 ITL and Parallel rank-2. If we select tasks by sampling classes in each task at random, then cross-task similarity increases. We observe ∼65% accuracy for rank-1 ITL.



(a) Layer 1 **(b)** Layer 2

Fig. 1: Top K singular values of weight matrices corresponding to different tasks for S-CIFAR100 with MLP experiments.