Compress then Serve: Serving Thousands of LoRA Adapters with Little Overhead

Rickard Brüel Gabrielsson ¹ Jiacheng Zhu ¹ Onkar Bhardwaj ² Leshem Choshen ¹² Kristjan Greenewald ² Mikhail Yurochkin ² Justin Solomon ¹

Abstract

Fine-tuning large language models (LLMs) with low-rank adaptations (LoRAs) has become common practice, often yielding numerous copies of the same LLM differing only in their LoRA updates. This paradigm presents challenges for systems that serve real-time responses to queries that each involve a different LoRA. Prior works optimize the design of such systems but still require continuous loading and offloading of LoRAs, as it is infeasible to store thousands of LoRAs in GPU memory. To mitigate this issue, we investigate the efficacy of compression when serving LoRAs. We propose a method for the joint compression of LoRAs into a shared basis paired with LoRA-specific scaling matrices. We extend our algorithm to learn clusters of LoRAs that are amenable to joint compression, allowing it to scale gracefully to large LoRA collections. Our experiments with up to 1000 LoRAs¹ demonstrate that compressed LoRAs preserve performance while offering major throughput gains in realistic serving scenarios with over a thousand LoRAs, maintaining 80% of the throughput of serving a single LoRA.

1. Introduction

The myriad uses for foundation models (FMs) have led to a proliferation of specialized models, each fine-tuned to perform a downstream task. To avoid fine-tuning foundation models with billions of parameters, parameter-efficient fine-tuning (PEFT) algorithms were proposed. An especially successful PEFT method is low-rank adaptation (LoRA) (Hu et al., 2021), which learns low-rank additive changes to neural network matrices. Because of the low-rank parameterization, these matrices (called adapter

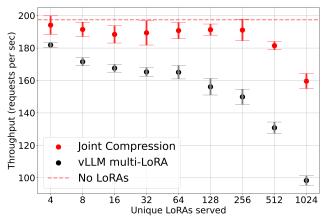


Figure 1: Throughput gains when serving 1000s of compressed LoRAs with vLLM.

weights) contain orders of magnitude fewer parameters than the base model. Still, LoRA can achieve performance on par with full fine-tuning (Hu et al., 2021).

LoRA's popularity has triggered a growing need to serve large collections of LoRA adapters at scale. Proprietary and open-source LLM providers offer fine-tuning services (OpenAI, 2024; TogetherAI, 2024; Predibase, 2024) with user bases likely in the thousands or even hundreds of thousands. As each user wants to use their own finetuned version of the LLM, serving a dedicated fine-tuned LLM per user becomes infeasible. To this end, S-LoRA (Sheng et al., 2023) proposes a system where only the base LLM is placed on an inference server and individual LoRA adapters are switched as needed at inference time. S-LoRA optimizes the system's inner workings via custom CUDA kernels and memory management to increase throughput when serving multiple LoRAs. Multi-LoRA system design has also been adopted in vLLM (Kwon et al., 2023), a stateof-the-art LLM serving engine. Despite optimized system designs, serving LoRAs still has a fundamental limitation: when the number of adapters is large, they need to be constantly loaded and offloaded from GPU memory to accommodate incoming requests, degrading throughput.

The problem of accommodating multiple LoRA adapters is also apparent when placing LLMs on edge devices, where smaller LLMs are fine-tuned for various tasks, and

¹MIT CSAIL ²MIT-IBM Watson AI Lab. Correspondence to: Rickard Brüel Gabrielsson

 cbrg@mit.edu>.

https://huggingface.co/Lots-of-LoRAs

the adapters are swapped depending on the task at hand (Gunter et al., 2024). In this setting, the number of adapters is smaller, e.g., a few dozen (Gunter et al., 2024), but the memory constraints are also more stringent.

In this work, we consider the problem of compressing a collection of LoRAs. We have two key objectives: (1) preserving the performance of the original LoRAs and (2) improving the throughput of serving many LoRAs. We formulate LoRA compression as a reconstruction problem, where the goal is to approximate the original adapters via collections of matrices of a smaller size. We compress Lo-RAs jointly by finding a shared basis and LoRA-specific scaling matrices and propose a joint diagonalization-based algorithm (JD). To improve reconstruction error for large numbers of LoRAs while keeping the number of parameters in check, we propose a clustering approach where each cluster is compressed independently using the joint diagonalization algorithm. Our clustering algorithm is based on alternating between optimizing the cluster assignments and the per-cluster reconstruction error.

Figure 1 showcases the benefits of joint compression. When serving up to 64 unique LoRAs, we use JD without clustering and for 128 or more, we pick the number of clusters to match the performance of compressed and original LoRAs. In each case, the GPU memory footprint of the compressed and original LoRAs is matched for a fair comparison to vLLM's multi-LoRA inference engine. When serving over 1000 LoRAs, compression increases throughput 1.6× and maintains 80% of the throughput of serving the base LLM (or a single LoRA merged into the LLM). §6 presents detailed results.

We summarize our main contributions below:

- We formulate the problem of compressing a collection of LoRAs and propose a joint compression scheme based on joint diagonalization.
- For large numbers of LoRAs, we scale joint compression by proposing a clustering algorithm where each cluster is jointly compressed to minimize reconstruction error.
- We establish theoretical guarantees for the reconstruction error of our compression formulation and relate reconstruction loss to performance empirically.
- We train a collection of more than 1000 high-quality LoRAs for Mistral-7B-Instruct-v0.2 (Jiang et al., 2023a) on 1000 natural instruction tasks (Wang et al., 2022) and demonstrate that our compression techniques preserve the performance of the original LoRAs. We will release over a 1000 LoRAs to facilitate future work as well as the code for our method.
- We incorporate LoRA compression into a state-of-theart LLM serving system and demonstrate that it is possible to serve over 1000 LoRAs across thousands of asynchronous requests with throughput comparable to serving

a single LoRA.

2. Related Work

Parameter-efficient fine-tuning (PEFT) has become prevalent for updating foundation models thanks to the need for efficiency in training and communication (Lialin et al., 2023). Many PEFT methods have been proposed, e.g. (Houlsby et al., 2019; Liu et al., 2022b) and LoRA (Hu et al., 2021) became the standard, partially due to the ease of switching between LoRAs in inference time.

Several works improve LoRA (Liu et al., 2024; Wang et al., 2024), sometimes with algebraic methods like SVD (Meng et al., 2024; Zhang et al., 2023; Jiang et al., 2023b) or by leveraging its statistical properties (Zhu et al., 2024; Zeng & Lee, 2024). Relatively few, however, accelerate inference times. S-LoRA (Sheng et al., 2023) provides an efficient means of switching between LoRAs. Wen & Chaudhuri (2024) adapt training to reduce batch multiplications, accelerating inference. Our method achieves a similar outcome (see Appendix D) without changing the LoRA formulation or requiring that LoRAs be trained in a dedicated way; future improvements to LoRA will also benefit from this aspect of our work (e.g., Meng et al. (2024)).

Punica (Chen et al., 2023) introduces Segmented Gather Matrix-Vector Multiplication (SGMV) to optimize multi-LoRA serving by parallelizing feature-weight multiplications in batches and grouping requests that use the same LoRA. Our approach, by contrast, reduces parameters as a means to serve multiple LoRAs efficiently, providing an orthogonal strategy that can be seamlessly integrated with Punica's methods to enhance performance. In our vLLM experiments, we leveraged the Punica kernel for multi-LoRA implementation, demonstrating the application of our method in conjunction with Punica's optimizations.

Other research proposes alternative PEFT methods that can be more parameter-efficient than LoRA. For example, VeRA (Kopiczko et al., 2024) fine-tunes LLMs by sharing global static parameters while learning local scaling variables; (IA)³ (Liu et al., 2022a) also reduces adapter parameter counts. However, none of these approaches has been as extensively tested or widely-adopted as LoRA. As a result, work that builds on LoRA enjoys a practical advantage due to its broad acceptance in practice.

There are many efforts to compress models (Cheng et al., 2017; Gholami et al., 2022; Sharma et al., 2024; Li et al., 2018). Predominantly, pruning and sparsification methods delete weights (Yadav et al., 2023a), and quantization methods reduce the weights' precision (Dettmers et al., 2024). Some works compress weights to reduce model size but typically require decompression and hence do not save GPU memory (Hershcovitch et al., 2024). Similarly to our

work, a few note increased performance and generalization after compression (Yadav et al., 2023a; Nadjahi et al., 2023; Hershcovitch et al., 2024; Sharma et al., 2024).

Our work also relates to model merging (Choshen et al., 2022; Wortsman et al., 2022; Matena & Raffel, 2021) and mixtures of experts (Muqeeth et al., 2024; Yadav et al., 2024). These methods reuse models trained by others (Choshen et al., 2023; Raffel, 2023), serving them together as one compressed model. Despite this similarity, these methods create a single general model that acts on any input, while ours yields more performant per-task solutions.

3. Rank-Based LoRA Compression

LoRA updates are parameterized by pairs of matrices A, B, whose product BA updates the fixed weight matrices $W_0 \in \mathbb{R}^{d_B \times d_A}$ of a neural network foundation model. Given an input x to a layer, the output of the LoRA-updated model at this layer is $(W_0 + BA)x$.

In formulating our compression algorithms, we consider a collection of given LoRA adapters $\{(A_i, B_i)\}_{i=1}^n$ that we would like to serve. We let r_i refer to the rank of the LoRA adapter-pair (A_i, B_i) , i.e., $B_i \in \mathbb{R}^{d_B \times r_i}$, $A_i \in \mathbb{R}^{r_i \times d_A}$.

While our compression technique has access only to a collection of $\{(A_i, B_i)\}_{i=1}^n$ pairs, in our experiments we will assess the efficacy of compression by comparing how the compressed matrices perform relative to uncompressed LoRAs on typical data. For this reason, although in this section we optimize a Frobenius norm reconstruction error relative to the product B_iA_i , this is a proxy for the nonlinear and complex way that compression errors in the adapters impact transformer performance. Our experiments will thus focus on the performance of the compressed LoRAs against the uncompressed versions on real data in §6.

Our compression methods significantly reduce the overall number of parameters. Reducing parameters theoretically accelerates storage and serving of a collection of LoRAs. This reduction, however, alters the computational dynamics during inference, so parameter reduction alone does not immediately imply faster throughput. In light of the complexities of GPU optimization, we experimentally assess throughput under realistic conditions in §6.4.

3.1. Joint Diagonalization

To scale to many LoRAs, the compressed number of parameters should not scale linearly with n. Hence compressing each LoRA individually (e.g., via SVD as in our experimental baselines) is inherently limited. To address this, we suggest a Joint Diagonalization (JD) method, which optimizes a shared basis onto which we can project the set of n LoRAs. This allows structure to be shared, implicitly

grouping and/or merging the collection of LoRAs.

In this model, each LoRA product B_iA_i is factorized into the form $U\Sigma_iV$, where U and V are shared across all Lo-RAs and Σ_i is specific to each LoRA. In this formulation, every Σ_i shares the same rank r. This allows U and V to be pre-loaded onto the GPU, with Σ_i loaded when necessary for each batch. The matrices Σ_i can be either diagonal or small square matrices, thus significantly reducing the number of LoRA-specific parameters and accelerating multi-LoRA serving.

Objective function. Motivated by the relationship of singular value decomposition to minimizing the Frobenius norm of the reconstruction error, we also propose to minimize the Frobenius norm of the adapter matrix approximation error. Specifically, we use the following objective:

$$\min_{\{\Sigma_i\}_{i=1}^n, U, V} \sum_{i=1}^n \|B_i A_i - U \Sigma_i V^\top\|_{\text{Fro}}^2.$$
 (1)

Note this problem is *not* solved by a single matrix SVD, since U and V are shared among all terms but the Σ_i 's are not. Using the Frobenius norm has the added benefit of making the objective convex in each argument separately, suggesting the possibility of efficient optimization. This objective function is underdetermined, however, so we consider two constrained regimes below.

Full Σ_i approximation. The first method we call JD-Full. Without loss of generality, U and V can be constrained to be orthogonal, so long as Σ_i remains an unconstrained full matrix. JD-Full adopts this restriction to make the optimization better posed, but note it does not restrict the expressiveness of the objective equation 1. This setting yields the following optimization problem:

$$JD-Full_r(\{B_i A_i\}_{i=1}^n) = \underset{\{\Sigma_i\}_{i=1}^n}{\operatorname{argmin}} \sum_{i=1}^n \|B_i A_i - U\Sigma_i V^\top\|_{\operatorname{Fro}}^2$$
$$U^\top U = V^\top V = I_r$$

$$(JD-Full) \qquad (2)$$

An efficient alternating algorithm to optimize this objective function can be found in Appendix A.

Diagonal Σ_i **approximation.** As an alternative, we can leave U, V unconstrained (other than to have r columns) and instead constrain the matrices Σ_i to be diagonal (but not necessarily positive). This formulation yields the following optimization problem:

$$\operatorname{JD-Diag}_r(\{B_i A_i\}_{i=1}^n) = \underset{\{\Sigma_i\}_{i=1}^n, U, V}{\operatorname{argmin}} \sum_{i=1}^n \|B_i A_i - U \operatorname{diag}(\Sigma_i) V^\top\|_{\operatorname{Fro}}^2$$

$$(\operatorname{JD-Diag}) \quad (3)$$

Appendix A provides an efficient alternating least squares algorithm for this objective. This diagonal version has per-LoRA parameter savings when compared to JD-Full, since the diagonal Σ_i only needs r parameters instead of r^2 .

3.2. Clustering

As the number of LoRAs n grows and becomes more diverse, the rank r needed for Joint Diagonalization to achieve good performance will tend to increase. This increases the size of each Σ_i that needs to be stored, especially for JD-Full which will require $O(nr^2)$ storage for these matrices. If the necessary r grows proportionally to n, then this storage will eventually become the bottleneck.

To resolve this limitation with very large n, we propose to group the n LoRAs into k clusters C_j . Each cluster is given its own rank r for JD compression, and the clusters are chosen such that the overall reconstruction error is minimized. Specifically, the overall objective is

$$\min_{\{\{C_j\},U_j,V_j\},\{\Sigma_i\}} \sum_j \sum_{i \in C_j} ||B_i A_i - U_j \Sigma_i V_j||_F^2,$$

optimized by alternating between cluster assignments and the JD of each cluster; Appendix A.3 provides details. Typically, the goal with large n is to have k grow with n as r becomes fixed. Comparing k rank-r JD-Full clusters to a rank-kr JD-Full single cluster compression, the clustered approach requires $O(dkr + nr^2)$ parameters, while the single-cluster approach requires $O(dkr + nk^2r^2)$ parameters due to the increased sizes of the Σ_i s. While these two approaches have the same rank, they may have different reconstruction abilities. Empirically, we find that multiple clusters significantly aid performance for $n \ge 100$.

4. Theoretical Analysis

In this section, we seek to better understand the role of the joint diagonalization method in §3.1 and how it motivates the clustering approach. We focus on the full- Σ_i case with orthogonal U, V matrices. Note that, for the same r, the r-JD-Diag has at least as large reconstruction error as r-JD-Full since it imposes an additional constraint on the Σ_i .

Firstly, note that perfect reconstruction can be achieved if and only if r is large enough, since there exist U, V such that all the B_i , A_i are in the spans of U, V resp. if and only if $r \geq \tilde{r}$:

Proposition 1. Suppose rank $(B_i A_i) = r_i$ for all i, and let

$$\tilde{r} = \max \left\{ \operatorname{rank}([A_1, \dots, A_n]), \operatorname{rank}([B_1^\top \dots, B_n^\top]) \right\}.$$

Note $\max_i r_i \leq \tilde{r} \leq \sum_{i=1}^n r_i$. Then JD-Full (equation 2) with $r = \tilde{r}$ compresses losslessly (perfect reconstruction), while $r < \tilde{r}$ will give nonzero reconstruction error.

Due to training noise, \tilde{r} will equal $\sum_{i=1}^{n} r_i$ almost always. This implies that in most realistic settings, the joint diagonalization approach is a lossy reconstruction.

This reconstruction loss can be significant, as the following theorem shows (proved in Appendix B):

Theorem 1. Consider n LoRAs $\{A_i, B_i\}_{i=1}^n$ with $r, n \leq d^2$, and form the matrix $L = [\operatorname{vec}(B_1A_1) \cdots \operatorname{vec}(B_nA_n)]$. Let σ_j be the singular values of L, sorted from largest to smallest, and let $\bar{\sigma}_j$ be the singular values of $\sum_{i=1}^n B_i A_i$. Then, using JD-Full (equation 2),

$$\sum_{j=1}^{r} \bar{\sigma}_{j}^{2} \leq \sum_{i=1}^{n} \|\Sigma_{i}\|_{\operatorname{Fro}}^{2} = \sum_{i=1}^{n} \|U\Sigma_{i}V^{\top}\|_{\operatorname{Fro}}^{2} \leq \sum_{j=1}^{\min(r^{2}, n)} \sigma_{j}^{2},$$

implying the sum of squared Frobenius norms of the reconstructed LoRAs satisfies

$$\begin{split} & \frac{\sum_{i=1}^{n} \|U\Sigma_{i}V^{\top}\|_{\operatorname{Fro}}^{2}}{\sum_{i=1}^{n} \|B_{i}A_{i}\|_{\operatorname{Fro}}^{2}} \leq \frac{\sum_{j=1}^{\min(r^{2},n)} \sigma_{j}^{2}}{\sum_{j=1}^{n} \sigma_{j}^{2}} \leq 1, \text{ and} \\ & \frac{\sum_{i=1}^{n} \|U\Sigma_{i}V^{\top} - B_{i}A_{i}\|_{\operatorname{Fro}}^{2}}{\sum_{i=1}^{n} \|B_{i}A_{i}\|_{\operatorname{Fro}}^{2}} \geq 1 - \frac{\sum_{j=1}^{\min(r^{2},n)} \sigma_{j}^{2}}{\sum_{j=1}^{n} \sigma_{j}^{2}}. \end{split}$$

In other words, reconstruction error is unavoidable if L's singular values are not concentrated in the top r^2 entries.

Remark 1 (Lower bound and merging). The lower bound $\sum_{j=1}^{r} \bar{\sigma}_{j}^{2}$ could be achieved by setting all the Σ_{i} equal, i.e., using a fully merged model instead of only merging the subspaces U, V and allowing Σ_{i} to vary with i.

Remark 2 (Upper bound and grouping). The upper bound is smallest when the LoRAs are relatively clustered, i.e., when groups of vectors $vec(B_iA_i)$ are similar. This situation raises the magnitude of the largest singular values of L, raising the upper bound in the proposition. As the LoRAs are $d \times d$ matrices that can be thought of as points in \mathbb{R}^{d^2} , for typical values of d well into the hundreds, it is likely that unrelated LoRAs will be unclustered, i.e., they will have relatively low inner products with each other.

For orthogonal LoRAs, the singular values of L are the norms of the LoRAs, suggesting the following corollary:²

²A result for isotropic Gaussian LoRAs could be obtained via the quantiles of the Marchenko-Pastur Law.

Corollary 1. Suppose (e.g., due to normalization) that the inputs to the joint diagonalization algorithm all have unit Frobenius norm, i.e., $||B_iA_i||_{Fro} = 1$. Moreover, assume that the LoRAs are all orthogonal in the sense $\operatorname{tr}((B_iA_i)(B_jA_j)^\top) = 0$ for $i \neq j$. Then, using the JD-Full method equation 2, we have $1 \leq \sum_{i=1}^n ||\Sigma_i||_{Fro}^2 \leq \min(r^2, n)$, implying that the sum of squared Frobenius norms of the reconstructed LoRAs satisfies

$$1 - \frac{1}{n} \ge \frac{\sum_{i=1}^{n} \|U\Sigma_{i}V^{\top} - B_{i}A_{i}\|_{Fro}^{2}}{\sum_{i=1}^{n} \|B_{i}A_{i}\|_{Fro}^{2}} \le 1 - \min\left(\frac{r^{2}}{n}, 1\right).$$

This implies that for the common setting where $r^2 \ll n$, the reconstructed LoRAs will be significantly smaller than the original LoRAs, with significant reconstruction error.

Our analysis illustrates the tradeoffs of joint diagonalization. If the LoRAs are similar or well-clustered, reconstruction error will be low. On the other hand, if the LoRAs are random and orthogonal, reconstruction error will be high.

Since the loss space of transformers is highly complex, increasing reconstruction error does not necessarily degrade LLM performance. Interestingly, Figure 3 below shows that while large reconstruction error rapidly decreases performance, moderate (but still relatively large, at around 60%) reconstruction error does not damage performance and may even slightly outperform the zero-error setting. At the same reconstruction error, clustering outperforms non-clustering. This motivates our focus on minimizing reconstruction error, while also suggesting that our approach achieves something deeper than compression. Specifically, joint diagonalization finds subspaces that are shared among many LoRAs when r is large and *merges* subspaces when r is small. When r is particularly small, this tendency towards averaging all or some of the LoRAs connects to merging LoRAs, whose empirical success (Shah et al., 2023; Huang et al., 2024) could explain the procedure's success despite the nonlinearity of transformers.

Appendix H.11 explores this idea further, comparing reconstruction of real-world LoRAs to reconstruction of randomly sampled LoRAs. The reconstruction error is generally large, but significantly lower than the reconstruction error for random noise, indicating that a major shared component between the LoRAs is successfully retained.

That said, as the number of LoRAs grows, the shared component may not be significant enough to maintain sufficiently low reconstruction error with low rank r. This motivates the introduction of *clustering* in §3.2, since clustering seeks to find groups of LoRAs that are similar and better compressible by joint diagonalization. In particular, if the number of clusters k grows with n, the reconstruction error may no longer degrade with n even when r is fixed.

In the extreme case where k = n, each LoRA is com-

pressed independently. By the Eckart-Young Theorem, JD applied to a single LoRA reduces to an SVD, replacing each rank- r_i LoRA adapter B_iA_i with a reduced rank- r_i approximation, where typically $r < \frac{1}{n} \sum_{i=1}^n r_i$:

$$SVD_r(B_iA_i) = U_i\Sigma_iV_i^{\top}, \quad \forall i = 1, \dots, n.$$
 (4)

As $\Sigma_i V_i^{\top}$ can be saved as a single matrix, this approach has $rn(d_A+d_B)$ parameters. We refer to this k=n method as r-SVD and find that it underperforms our other methods while slightly outperforming the baseline uncompressed LoRAs. This result parallels Jiang et al. (2023b)'s observation that lowering LoRA ranks is beneficial for multi-task learning and model merging.

5. Training & Performance Evaluation

5.1. Training

We trained LoRA adapters on 1000 natural instruction tasks (Wang et al., 2022) using Mistral-7B-Instruct-v0.2 (Jiang et al., 2023a) as the base. We set all LoRA adapter ranks to 16 (i.e., $\forall i, r_i = 16$), except for those in our ablation study (Appendix H.1), where we vary the LoRA rank.

We selected 10 diverse tasks (Table 2 in Appendix C) manually for consistent evaluation across experiments and randomly sampled an additional 990 tasks, resulting in a total of 1000 tasks (Table 3). The tasks went through a robust reviewing protocol to ensure high quality and diversity. Each task data was divided into training, validation, and test sets.

Hyperparameters, such as early stopping, were tuned using the validation sets. Table 1, Appendix C shows that on the test sets, LoRA consistently outperformed the base model in terms of Rouge scores and loss metrics.

5.2. Evaluation

We evaluated multiple metrics for the natural instruction tasks, including cross-entropy loss, Rouge-1, Rouge-L (Lin, 2004), exact match, and *agreement* between uncompressed and compressed LoRA. Here, *agreement* measures the exact match in task-generations between the uncompressed LoRA model and the compressed LoRA model, rather than comparing to ground truth data. While detailed results and discussions for all metrics are provided in Appendix H, our primary focus in the main text is on Rouge-L. We find that all metrics correlate, but Rouge-L correlates most strongly with downstream utility. This finding aligns with prior work (Wang et al., 2022), which demonstrates that Rouge-L correlates well with classification accuracy.

While cross-entropy is used for optimization during training, identical generation outputs across models can yield different cross-entropy losses. Exact match is too rigid and

does not account for the variability in task responses. Similarly, agreement does not capture the inexactness associated with most of our tasks, nor does it account for the performance gains or losses of the compressed LoRAs. Arguably, practitioners are primarily concerned with task performance in the settings for which the LoRA was designed, rather than exact generational agreement between models.

Joint diagonalization optimizes reconstruction error measured by the Frobenius norm, bounded by our theoretical analysis in §4. We empirically study the relation between the reconstruction error and downstream Rouge-L performance in Section 6.2.

Instead of listing absolute performance, we compute the performance difference between the base model and the LoRA model for each task via the ratio

 $Performance \ relative \ to \ LoRA := \frac{method\text{-performance}}{LoRA\text{-performance}}$

for the method in question, highlighting relative improvement wrt the uncompressed LoRAs.

6. Experiments

6.1. Task Performance

For each method, we vary the number n of compressed LoRAs and the compression rank r. We run each experiment three times with different random seeds and report the mean and standard deviation. See Table 7 for results evaluated on the same ten manually-selected tasks (Table 2) across settings. Every compressed collection of LoRAs contains these 10 tasks (i.e., in-distribution tasks), and each collection contains the smaller collections as subsets.

We normalize each LoRA adapter to have a Frobenius norm of one prior to running joint diagonalization. This normalization enhances performance and reduces the variance in reconstruction error. We restore the original norms of the LoRA adapters before reconstruction and testing.

Figure 2 illustrates the Rouge-L scores of the compressed LoRAs divided by the Rouge-L scores of the uncompressed LoRAs. JD variants often increase generalization and outperform the original LoRA. Notably, our JD methods approach the compression efficacy of a single LoRA, and with clustering, this aggressive reduction in size also maintains performance in larger collections. Appendix H includes tables of additional relative and absolute metrics.

For efficiency, we limited the JD methods to ten iterations instead of full convergence. While the alternating algorithm quickly reaches an approximate minimizer, squeezing out the last few digits of precision takes many more iterations with limited to no performance gain. Appendix H.12 also evaluates an alternative iterative algorithm that con-

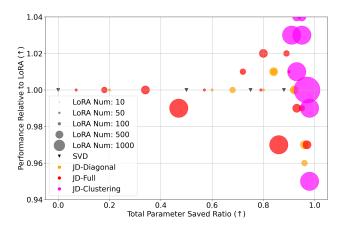


Figure 2: Performance after compression. We compare the performance of compressed LoRAs relative to uncompressed ones, with higher values on both axes reflecting better performance. The Total Parameter Saved Ratio depicts the number of parameters saved for a system with a large number n of different LoRAs. It is computed as: $r_{total} := 1 - \frac{\text{num. parameters after compression}}{\text{num. parameters before compression}}.$

verges more rapidly once U, V are close to a minimizer, with minimal performance differences.

6.2. Performance and Reconstruction Error

Figure 3 relates reconstruction error and performance. The y-axis measures the mean performance improvement of Rouge-L relative to uncompressed LoRA, and the x-axis quantifies the mean relative reconstruction error between the compressed reconstruction of the product BA and the original product BA. Although performance and reconstruction error relate non-linearly, we see a decreasing, somewhat exponential trend. Notably, minimizing reconstruction error does not yield optimal performance, indicating that mild lossy reconstruction may enhance generalization. Interestingly, under the clustering approach, compared to non-clustering, even more aggressive lossy reconstruction can outperform less lossy reconstruction, suggesting that reconstruction error is even less critical for performance in the clustering scenario.

To select hyperparameters (compression rank and number of clusters) for the clustering experiments, we first assessed reconstruction error on a single LoRA module over a range of settings (see Appendix G). These preliminary experiments enabled efficient selection of cluster counts and rank values for compressing all LoRA modules.

6.3. Benefits of Compression

Compressing LoRAs reduces their parameter counts, thus lowering their overall memory footprint. While this offers

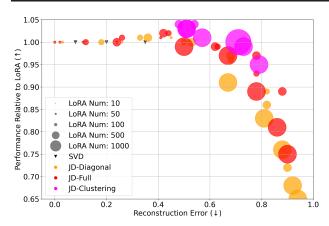


Figure 3: Reconstruction error vs. performance.

many benefits, in both training and inference scenarios parameters are often transferred between different memory hierarchies (e.g., from CPU to GPU), and these transfers usually scale linearly with the amount of data moved. At the same time, compressed LoRAs alter the forward-pass formulation (unless merged with the base model weights). As shown in Figure 5 in the Appendix E, although compression greatly reduces memory usage and transfer time, it does not affect the forward-pass latency. Of the various ways to leverage these improvements, this work focuses on optimizing inference for multiple LoRAs using vLLM.

6.4. Throughput of Serving Compressed LoRAs

The previous sections demonstrate how to select an appropriate joint compression setting guided by the reconstruction error, such that the performance of the original LoRAs is preserved. Naturally, the rank and/or the number of clusters for the compression needs to increase as we compress larger LoRA collections to match LoRA performance.

Figure 4 studies how throughput with various compression settings compares to the vLLM multi-LoRA throughput with the matched GPU memory footprint. Specifically, for each number of unique LoRAs served and each compression setting, we compute the corresponding number of Lo-RAs to be placed on the GPU during serving and report the ratio of the two throughputs. For example, when serving 64 unique LoRAs and using rank 64 JD-Full compression, we report the ratio of throughputs of rank 64 JD-Full and vLLM multi-LoRA with 6 LoRAs allowed on the GPU at a time (see Appendix F for details). As the number of unique LoRAs increases, vLLM multi-LoRA throughput degrades as it needs to schedule the requests and load and offload the adapters. We note that vLLM multi-LoRA already employs advanced optimizations, such as efficient scheduling and non-blocking CPU-GPU communication when swapping LoRAs as well as techniques introduced in S-LoRA (Sheng et al., 2023; Kwon et al., 2023), but system opti-

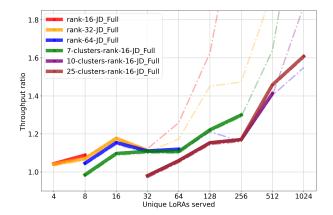


Figure 4: Throughput ratio when serving varying numbers of LoRAs with vLLM. Highlighted settings preserve at least 99% of the uncompressed LoRA performance.

mization alone is insufficient to mitigate throughput degradation when serving many LoRAs.

Figure 4 shows that across LoRA collection sizes our compression techniques improve the throughput of vLLM multi-LoRA. Additionally, we highlight regions for each compression setting where compression is sufficiently moderate to achieve 99%+ of LoRA performance, according to the results in §6.2. Compression with a larger rank or too many clusters does not improve baseline throughput when serving a smaller number of LoRAs and should not be used in such cases. For example, rank 16 JD-Full improves baseline throughput with 4 and 8 LoRAs, but will underperform with more LoRAs, while 25 clusters rank 15 JD-Full does not improve throughput with 32 or fewer LoRAs, but when serving 1000+ LoRAs it improves the throughput significantly while maintaining the performance. Overall, an appropriate joint compression setting improves vLLM multi-LoRA throughput and preserves performance for LoRA collections of any size between 4 and 1024, as in Figure 1. Appendix F provides compression settings for each collection size.

vLLM extensively uses custom CUDA kernels. To accommodate our compression techniques, we minimally adjusted the vLLM code to generate additional kernels needed by the compressed LoRAs and used the Punica (Chen et al., 2023) kernel to further accelerate matrix multiplication. Pseudocode is given in §F.4 to show how we use the batch multiplication kernel. There likely is room for improvement to optimize the newly added kernels.

Additional details. In this experiment, we considered a varying number of rank-16 LoRAs, using a dataset of Shakespeare sonnets as inputs³ arriving *asynchronously*.

³https://www.kaggle.com/
datasets/shivamshinde123/
william-shakespeares-sonnet/data

We measured throughput, i.e., the number of requests served per second when generating ten tokens per request. The base was Mistral 7B Instruct; we simulated random LoRAs and assigned inputs to LoRAs at random. Experiments were conducted on H100 80GB GPU capped at 40% memory consumption to reflect situations where a service provider might want to serve many LoRAs from cheaper hardware with lower memory than higher-end GPUs. This setting applies to the scenario where the LLM is large compared to the size of GPU and yet a provider may want to serve many LoRAs efficiently using one device.

6.5. Recommendations

JD-Full is generally preferred over JD-Diag, although for smaller numbers of LoRAs (less than 100), the performance difference is negligible. While JD-Full alone is effective up to 100 LoRAs, incorporating clustering at scales of 500-1000 LoRAs significantly enhances performance.

We recommend the following procedure for hyperparameter selection. For ≤ 100 LoRAs, JD-Full can be used without substantial degradation, using a rank \approx (number of LoRAs/2) + 7. Beyond 100 LoRAs, clustering becomes increasingly critical. A robust method for any number of LoRAs up to 1000 uses JD-Full with clustering. Specifically, select a LoRA module from the middle of the network, apply a compression rank of 16, and experiment with an exponentially increasing number of clusters. Compute the reconstruction error for each setting on this module across all LoRAs—a computationally efficient process. Choose the minimal number of clusters that achieves a reconstruction loss below 0.6, and then use these settings across LoRA modules. Figure 6 in the Appendix illustrates this procedure applied to 500 LoRAs.

Tuning hyperparameters as discussed above using reconstruction loss as a validation metric is convenient since it can be done efficiently on CPU without expensive LLM evaluation. As our experiments demonstrate, compression settings that achieve below 0.6 reconstruction loss reliably preserve 99% or more of the LoRA performance, sometimes even outperforming the original LoRAs.

For inference, this procedure is executed as a preprocessing step before deploying our inference server. As new Lo-RAs are submitted, they are initially served uncompressed. A background CPU job can periodically re-run the compression algorithm and update the served LoRA parameters with the compressed versions.

7. Discussion

This study introduces approaches to LoRA compression, addressing significant challenges emerging as customization of foundation models such as LLMs and diffusion

models becomes increasingly popular. Our contributions include theoretical formulations, empirical validation, and practical implementations that enhance the understanding and application of LLMs in scalable environments.

Our findings have several implications. Our theoretical bounds on reconstruction error not only increase confidence in the use of compressed models but also lay a groundwork for future explorations. Demonstrating that our compression techniques can preserve up to 100% of the original LoRAs' performance highlights the effectiveness of our methods. Furthermore, integrating LoRA compression into state-of-the-art LLM serving systems demonstrates potential for resource optimization, with throughput for thousands of LoRAs nearing that of a single LoRA.

Our promising results suggest several future research directions. First, further compression may be possible via quantization, since joint-diagonalization and quantization are independent compression strategies. Second, when scaling to hundreds of thousands of LoRAs, joint compression, while effective, will be insufficient to fit all LoRAs onto the GPU, thus requiring a procedure to schedule the requests. Clustering offers opportunities for efficient scheduling that incorporates the cluster assignments of LoRAs corresponding to the incoming requests.

Privacy presents another research direction, particularly regarding the possibility of information leakage during joint compression. As a preliminary study, Appendix H.2 investigates whether a base model with an adapter A for task T_A , after being jointly compressed alongside an adapter B for task T_B , inadvertently improves on T_B . Such an outcome would indicate that adapter A acquired information from adapter B. Our ablation study shows no performance gains on T_B , suggesting that the compressed adapter A remains independent and does not leak—or gain—information from adapter B. A more detailed investigation of the privacy properties of joint compression is an interesting next step.

In conclusion, our research advances LLM deployment by providing robust, scalable, and efficient compression. The ability of compressed LoRAs to maintain high performance while saving resources opens avenues for the broad application and adoption of LLMs across various industries. We encourage the community to build upon our findings and shared LoRAs to further enhance these technologies.

Impact Statement

This paper presents work whose goal is to advance machine learning. There are no societal consequences of our work that we feel must be specifically highlighted here.

References

- Chen, L., Ye, Z., Wu, Y., Zhuo, D., Ceze, L., and Krishnamurthy, A. Punica: Multi-tenant lora serving, 2023. URL https://arxiv.org/abs/2310.18547.
- Cheng, Y., Wang, D., Zhou, P., and Zhang, T. A survey of model compression and acceleration for deep neural networks. *arXiv* preprint arXiv:1710.09282, 2017.
- Choshen, L., Venezian, E., Slonim, N., and Katz, Y. Fusing finetuned models for better pretraining. *ArXiv*, abs/2204.03044, 2022.
- Choshen, L., Venezian, E., Don-Yehiya, S., Slonim, N., and Katz, Y. Where to start? analyzing the potential value of intermediate models. In Bouamor, H., Pino, J., and Bali, K. (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 1446–1470, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023. emnlp-main.90. URL https://aclanthology.org/2023.emnlp-main.90.
- Dettmers, T., Pagnoni, A., Holtzman, A., and Zettlemoyer, L. Qlora: Efficient finetuning of quantized llms. Advances in Neural Information Processing Systems, 36, 2024.
- Gholami, A., Kim, S., Dong, Z., Yao, Z., Mahoney, M. W., and Keutzer, K. A survey of quantization methods for efficient neural network inference. In *Low-Power Computer Vision*, pp. 291–326. Chapman and Hall/CRC, 2022.
- Gunter, T., Wang, Z., Wang, C., Pang, R., Narayanan, A., Zhang, A., Zhang, B., Chen, C., Chiu, C.-C., Qiu, D., et al. Apple intelligence foundation language models. *arXiv* preprint arXiv:2407.21075, 2024.
- Hershcovitch, M., Choshen, L., Wood, A., Enmouri, I., Chin, P., Sundararaman, S., and Harnik, D. Lossless and near-lossless compression for foundation models. arXiv preprint arXiv:2404.15198, 2024.
- Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., and Gelly, S. Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pp. 2790–2799. PMLR, 2019.
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685, 2021.
- Huang, C., Liu, Q., Lin, B. Y., Pang, T., Du, C., and Lin, M. Lorahub: Efficient cross-task generalization via dynamic lora composition, 2024.

- Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. d. l., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., et al. Mistral 7b. arXiv preprint arXiv:2310.06825, 2023a.
- Jiang, W., Lin, B., Shi, H., Zhang, Y., and Kwok, J. T. Byom: Building your own multi-task model for free. *arXiv preprint arXiv:2310.01886*, 2023b.
- Kopiczko, D. J., Blankevoort, T., and Asano, Y. M. VeRA: Vector-based random matrix adaptation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=NjNfLdxr3A.
- Kwon, W., Li, Z., Zhuang, S., Sheng, Y., Zheng, L., Yu, C. H., Gonzalez, J., Zhang, H., and Stoica, I. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th Symposium on Operating Systems Principles*, pp. 611–626, 2023.
- Li, C., Farkhoor, H., Liu, R., and Yosinski, J. Measuring the intrinsic dimension of objective landscapes, 2018.
- Lialin, V., Deshpande, V., and Rumshisky, A. Scaling down to scale up: A guide to parameter-efficient fine-tuning. *arXiv* preprint arXiv:2303.15647, 2023.
- Lin, C.-Y. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL https://aclanthology.org/W04-1013.
- Liu, H., Tam, D., Muqeeth, M., Mohta, J., Huang, T., Bansal, M., and Raffel, C. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning, 2022a. URL https://arxiv.org/abs/2205.05638.
- Liu, H., Tam, D., Muqeeth, M., Mohta, J., Huang, T., Bansal, M., and Raffel, C. A. Few-shot parameterefficient fine-tuning is better and cheaper than in-context learning. *Advances in Neural Information Processing Systems*, 35:1950–1965, 2022b.
- Liu, S.-Y., Wang, C.-Y., Yin, H., Molchanov, P., Wang, Y.-C. F., Cheng, K.-T., and Chen, M.-H. Dora: Weightdecomposed low-rank adaptation, 2024.
- Matena, M. and Raffel, C. Merging models with fisher-weighted averaging. *arXiv preprint arXiv:2111.09832*, 2021.
- Meng, F., Wang, Z., and Zhang, M. Pissa: Principal singular values and singular vectors adaptation of large language models. *arXiv* preprint arXiv:2404.02948, 2024.

- Muqeeth, M., Liu, H., Liu, Y., and Raffel, C. Learning to route among specialized experts for zero-shot generalization. *arXiv preprint arXiv:2402.05859*, 2024.
- Nadjahi, K., Greenewald, K., Gabrielsson, R. B., and Solomon, J. Slicing mutual information generalization bounds for neural networks. In *ICML 2023 Workshop Neural Compression: From Information Theory to Applications*, 2023. URL https://openreview.net/forum?id=cblcwK3SZi.
- OpenAI. Openai fine-tuning api. https://platform.openai.com/docs/guides/fine-tuning, 2024.
- Predibase. Multi-lora inference server that scales to 1000s of fine-tuned llms. https://loraexchange.ai, 2024.
- Raffel, C. Building machine learning models like open source software. *Communications of the ACM*, 66(2): 38–40, 2023.
- Shah, V., Ruiz, N., Cole, F., Lu, E., Lazebnik, S., Li, Y., and Jampani, V. Ziplora: Any subject in any style by effectively merging loras. *arXiv preprint arXiv:2311.13600*, 2023.
- Sharma, P., Ash, J. T., and Misra, D. The truth is in there: Improving reasoning in language models with layer-selective rank reduction. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ozX92bu8VA.
- Sheng, Y., Cao, S., Li, D., Hooper, C., Lee, N., Yang, S., Chou, C., Zhu, B., Zheng, L., Keutzer, K., Gonzalez, J. E., and Stoica, I. S-lora: Serving thousands of concurrent lora adapters, 2023.
- Together AI. Together fine-tuning. https://www.together.ai/products#fine-tuning, 2024.
- Wang, S., Xue, B., Ye, J., Jiang, J., Chen, L., Kong, L., and Wu, C. Prolora: Partial rotation empowers more parameter-efficient lora. *ArXiv*, abs/2402.16902, 2024. URL https://api.semanticscholar.org/CorpusID:268032580.
- Wang, Y., Mishra, S., Alipoormolabashi, P., Kordi, Y., Mirzaei, A., Arunkumar, A., Ashok, A., Dhanasekaran, A. S., Naik, A., Stap, D., et al. Super-naturalinstructions:

- Generalization via declarative instructions on 1600+ nlp tasks. *arXiv preprint arXiv:2204.07705*, 2022.
- Wen, Y. and Chaudhuri, S. Batched low-rank adaptation of foundation models, 2024.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. Huggingface's transformers: State-of-the-art natural language processing, 2020.
- Wortsman, M., Ilharco, G., Gadre, S. Y., Roelofs, R., Gontijo-Lopes, R., Morcos, A. S., Namkoong, H., Farhadi, A., Carmon, Y., Kornblith, S., and Schmidt, L. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International Conference on Machine Learning*, 2022.
- Yadav, P., Choshen, L., Raffel, C., and Bansal, M. Compeft: Compression for communicating parameter efficient updates via sparsification and quantization. arXiv preprint arXiv:2311.13171, 2023a.
- Yadav, P., Tam, D., Choshen, L., Raffel, C., and Bansal, M. TIES-merging: Resolving interference when merging models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023b. URL https://openreview.net/forum?id=xtaX3WyCjl.
- Yadav, P., Raffel, C., Muqeeth, M., Caccia, L., Liu, H., Chen, T., Bansal, M., Choshen, L., and Sordoni, A. A survey on model moerging: Recycling and routing among specialized experts for collaborative learning. arXiv preprint arXiv:2408.07057, 2024.
- Zeng, Y. and Lee, K. The expressive power of low-rank adaptation, 2024.
- Zhang, Q., Chen, M., Bukharin, A., He, P., Cheng, Y., Chen, W., and Zhao, T. Adaptive budget allocation for parameter-efficient fine-tuning. In *The Eleventh International Conference on Learning Representations*, 2023.
- Zhu, J., Greenewald, K., Nadjahi, K., Borde, H. S. d. O., Gabrielsson, R. B., Choshen, L., Ghassemi, M., Yurochkin, M., and Solomon, J. Asymmetry in low-rank adapters of foundation models. arXiv preprint arXiv:2402.16842, 2024.

A. Joint Diagonalization Algorithms

A.1. Alternating Methods

Our goal is to derive algorithms that optimize equation 1. Common to both methods, we expand the objective functional:

$$\sum_{i} \|B_{i}A_{i} - U\Sigma_{i}V^{\top}\|_{\operatorname{Fro}}^{2} = \sum_{i} \operatorname{tr}((B_{i}A_{i} - U\Sigma_{i}V^{\top})(B_{i}A_{i} - U\Sigma_{i}V^{\top})^{\top}) \text{ by definition}$$

$$= \sum_{i} \left[\operatorname{tr}(B_{i}A_{i}A_{i}^{\top}B_{i}^{\top}) - 2\operatorname{tr}(B_{i}A_{i}V\Sigma_{i}^{\top}U^{\top}) + \operatorname{tr}(U\Sigma_{i}V^{\top}V\Sigma_{i}^{\top}U^{\top}) \right]$$

$$= \operatorname{const.} - 2\sum_{i} \operatorname{tr}(B_{i}A_{i}V\Sigma_{i}^{\top}U^{\top}) + \sum_{i} \|U\Sigma_{i}V^{\top}\|_{\operatorname{Fro}}^{2}.$$

$$(5)$$

Using this expansion, we now consider the two settings discussed in §3.1.

Case 1: Non-diagonal Σ_i , orthogonal U, V. Setting the derivative of equation 5 with respect to Σ_i to zero, we find

$$\Sigma_i = \Sigma_i^*(U, V) = U^{\top} B_i A_i V. \tag{6}$$

We simplify our objective function after plugging in this expression:

$$\sum_{i} \|B_{i}A_{i} - U\Sigma_{i}V^{\top}\|_{\operatorname{Fro}}^{2} + \operatorname{const.} = \sum_{i} \left[\|\Sigma_{i}\|_{\operatorname{Fro}}^{2} - 2\operatorname{tr}(B_{i}A_{i}V\Sigma_{i}^{\top}U^{\top}) \right] \text{ from equation 5}$$

$$= \sum_{i} \left[\operatorname{tr}(U^{\top}B_{i}A_{i}VV^{\top}A_{i}^{\top}B_{i}^{\top}U) - 2\operatorname{tr}(B_{i}A_{i}VV^{\top}A_{i}^{\top}B_{i}^{\top}UU^{\top}) \right] \text{ from equation 6}$$

$$= -\sum_{i} \operatorname{tr}(B_{i}A_{i}VV^{\top}A_{i}^{\top}B_{i}^{\top}UU^{\top}).$$

Substituting equation 6, we find

$$U_{opt}, V_{opt} = \arg\max_{\substack{U^{\top}U = I\\VV^{\top} = I}} \sum_{i=1}^{n} \|U^{\top}B_{i}A_{i}V\|_{Fro}^{2} = \arg\max_{\substack{U^{\top}U = I\\VV^{\top} = I}} \sum_{i=1}^{n} \|\Sigma_{i}^{*}(U, V)\|_{Fro}^{2}.$$
(7)

Note that

$$\sum_{i=1}^{n} \|U^{\top} B_i A_i V\|_{\text{Fro}}^2 = \operatorname{tr}\left(\left(\sum_{i=1}^{n} B_i A_i V V^{\top} A_i^{\top} B_i^{\top}\right) U U^{\top}\right)$$
$$= \operatorname{tr}\left(\left(\sum_{i=1}^{n} B_i^{\top} A_i^{\top} U U^{\top} A_i B_i\right) V V^{\top}\right),$$

by the identity $||A||_{Fro}^2 = \operatorname{tr}(A^{\top}A)$. Hence, we optimize equation 7 by alternating between U and V:

- U iteration: Define $M := \sum_i B_i A_i V V^\top A_i^\top B_i^\top$. Parenthesizing this expression properly requires only O((m+n)r) storage/computation time. With this definition, we maximize $\operatorname{tr}(MUU^\top)$ over U satisfying $U^\top U = I$. Since M is positive semidefinite, the optimum is to take U to be the r eigenvectors of M with largest eigenvalue, equivalent to an SVD problem.
- V iteration: Define $N := \sum_i A_i^\top B_i^\top U U^\top B_i A_i$. Similarly to the previous step, we take V to contain the r eigenvectors of N with largest eigenvalue, again solvable using an SVD.

This method decreases the objective in each step.

Case 2: Diagonal Σ_i . If constrain Σ_i to be diagonal, we interpret our objective function equation 1 as a "triple least

squares" problem. We compute gradients:

$$\nabla_{U} \sum_{i} \|B_{i}A_{i} - U\Sigma_{i}V^{\top}\|_{\operatorname{Fro}}^{2} = 2\sum_{i} (U\Sigma_{i}V^{\top} - B_{i}A_{i})V\Sigma_{i}^{\top}$$

$$\nabla_{V} \sum_{i} \|B_{i}A_{i} - U\Sigma_{i}V^{\top}\|_{\operatorname{Fro}}^{2} = 2\sum_{i} (V\Sigma_{i}^{\top}U^{\top} - A_{i}^{\top}B_{i}^{\top})U\Sigma_{i}$$

$$\nabla_{\Sigma_{i}} \sum_{i} \|B_{i}A_{i} - U\Sigma_{i}V^{\top}\|_{\operatorname{Fro}}^{2} = 2U^{\top}(U\Sigma_{i}V^{\top} - B_{i}A_{i})V$$

These expressions suggest efficient $r \times r$ linear systems to solve for U, V:

$$U = \left(\sum_{i} B_{i} A_{i} V \Sigma_{i}^{\top}\right) \left(\sum_{i} \Sigma_{i} V^{\top} V \Sigma_{i}^{\top}\right)^{-1}$$
$$V = \left(\sum_{i} A_{i}^{\top} B_{i}^{\top} U \Sigma_{i}\right) \left(\sum_{i} \Sigma_{i}^{\top} U^{\top} U \Sigma_{i}\right)^{-1}.$$

For Σ_i , we extract the diagonal from our gradient above:

$$\operatorname{diag}(U^{\top}U\Sigma_{i}V^{\top}V)_{j} = (U^{\top}U\Sigma_{i}V^{\top}V)_{jj}$$

$$= \sum_{m} (U^{\top}U)_{jm}\Sigma_{imm}(V^{\top}V)_{mj}$$

$$= (U^{\top}U \circ V^{\top}V)\operatorname{diag}(\Sigma_{i})$$

$$\operatorname{diag}(U^{\top}B_{i}A_{i}V)_{j} = \sum_{m} (U^{\top}B_{i})_{jm}(A_{i}V)_{mj}$$

$$= \sum_{m} (U^{\top}B_{i})_{jm}(V^{\top}A_{i}^{\top})_{jm}$$

$$= (U^{\top}B_{i} \circ V^{\top}A_{i}^{\top})\mathbf{1}$$

$$\Longrightarrow \operatorname{diag}(\Sigma_{i}) = (U^{\top}U \circ V^{\top}V)^{-1}(U^{\top}B_{i} \circ V^{\top}A_{i}^{\top})\mathbf{1}$$

Here o denotes the Hadamard product.

Combining these expressions, we use a simple coordinate descent algorithm cycling between the following three steps:

- 1. Solve for U
- 2. Solve for V
- 3. Solve for the Σ_i 's
- 4. Optionally, normalize so $\sum_{i} \|\Sigma_{i}\|_{\text{Fro}}^{2} = 1$

A.2. Additional Eigenvalue Iteration Algorithm

For the first case in §A.1, we introduce an alternative algorithm that eschews the use of SVD. This alternative is optimized for GPU execution, enabling tractable runs to convergence.

To derive this algorithm, we employ Lagrange multipliers to formulate the derived objective from equation 7:

$$U_{opt}, V_{opt} = \arg\max_{\substack{U^{\top}U = I\\VV^{\top} = I}} \sum_{i=1}^{n} \|U^{\top}B_{i}A_{i}V\|_{Fro}^{2},$$
(8)

yielding the expression

$$\Lambda = -\frac{1}{2} \|U^{\top} B_i A_i V\|_{\text{Fro}}^2 - \frac{1}{2} \text{tr}(X^{\top} (I - U^{\top} U)) - \frac{1}{2} \text{tr}(Y^{\top} (I - V^{\top} V)). \tag{9}$$

Taking the derivatives gives

$$\nabla_U \Lambda = -\sum_i B_i (A_i V) (V^\top A_i^\top) (B_i^\top U) + UX \tag{10}$$

$$\nabla_V \Lambda = -\sum_i A_i^\top (B_i^\top U) (U^\top B_i) (A_i V) + V Y \tag{11}$$

Setting these derivatives to zero shows

$$\sum_{i} B_i(A_i V)(V^{\top} A_i^{\top})(B_i^{\top} U) = UX$$
(12)

$$\sum_{i} A_i^{\top}(B_i^{\top}U)(U^{\top}B_i)(A_iV) = VY.$$
(13)

Here, one can show that the Lagrange multiplier matrices X and Y are diagonal and nonnegative, since the problem reduces to an eigenvalue problem when either U or V is fixed; this is essentially the argument behind the alternating algorithm in Appendix A. Hence, taking inspiration from classical eigenvalue iteration, we use the following updates to improve our estimates of U and V:

$$U_0^{(k+1)} \leftarrow \sum_i B_i(A_i V^{(k)}) ((V^{(k)})^\top A_i^\top) (B_i^\top U^{(k)})$$
(14)

$$V_0^{(k+1)} \leftarrow \sum_i A_i^{\top} (B_i^{\top} U^{(k)}) ((U^{(k)})^{\top} B_i) (A_i V^{(k)})$$
(15)

$$U^{(k+1)} \leftarrow \text{orthogonalize}(U_0^{(k+1)}) \tag{16}$$

$$V^{(k+1)} \leftarrow \text{orthogonalize}(V_0^{(k+1)}) \tag{17}$$

Here, the function orthogonalize orthogonalizes the columns of a matrix, e.g. by using the Q part of the reduced-size QR factorization. Although we lack a formal convergence proof, in practice we find that this method reliably reaches a local optimum of our problem.

By executing matrix operations in the specified sequence, these computations can be rapidly performed on GPUs. Note the expressions above are parenthesized to avoid constructing a large matrix product as an intermediate computation.

A.3. Clustering algorithm

Initialization: We run joint diagonalization with a single U, V then perform k-means with k clusters on the space of Σ_i 's. This gives us our first clusters and we can use random initialization U_j, V_j for each cluster but the Σ_i can be maintained as initialization.

Step 1: Using the alternating JD algorithms from earlier in this section, we optimize the problem $\min_{U_i, V_i, \Sigma_i} \sum_{i \in C_i} ||B_i A_i - U_j \Sigma_i V_j^\top||_F^2$ for each j independently.

Step 2: New cluster assignment for $i : \min_j \min_{\Sigma_i} ||B_i A_i - U_j \Sigma_i V_j^{\top}||_F^2$. If any assignment changes we go to Step 1, else we have converged.

B. Proof of Theorem 1

Proof. For the lower bound, note that by Jensen's inequality,

$$\sum_{i=1}^{n} \|U^{\top} B_i A_i V\|_{\text{Fro}}^2 \ge \left\| U^{\top} \sum_{i=1}^{n} B_i A_i V \right\|_{\text{Fro}}^2,$$

for any U, V. Hence,

$$\sup_{U,V \in St(k,d)} \sum_{i=1}^{n} \|U^{\top} B_i A_i V\|_{Fro}^2 \ge \sup_{U,V \in St(k,d)} \left\| U^{\top} \sum_{i=1}^{n} B_i A_i V \right\|_{Fro}^2.$$
 (18)

By the definition of singular value decomposition, the right hand side of equation 18 is maximized with U, V being the top r singular vectors of $\sum_{i=1}^{n} B_i A_i$, yielding $\|U^{\top} \sum_{i=1}^{n} B_i A_i V\|_{\text{Fro}}^2 = \sum_{i=1}^{r} \bar{\sigma}_i^2$. Recalling that $\Sigma_i = U^{\top} B_i A_i V$ yields the lower bound.

For the upper bound, recall that $\Sigma_i = U^{\top} B_i A_i V$. Rearranging,

$$\operatorname{vec}(\Sigma_i) = (V^{\top} \otimes U^{\top})\operatorname{vec}(B_i A_i).$$

Define

$$\bar{\Sigma} := [\operatorname{vec}(\Sigma_1), \dots, \operatorname{vec}(\Sigma_n)].$$

By our previous simplification,

$$\bar{\Sigma} = (V^{\top} \otimes U^{\top})L.$$

Now

$$\sum_{i=1}^{n} \|\Sigma_i\|_{\operatorname{Fro}}^2 = \|\bar{\Sigma}\|_{\operatorname{Fro}}^2 = \operatorname{tr}\left(((V \otimes U)(V \otimes U)^{\top})(LL^{\top})\right)$$

Since U, V are orthogonal and size $d \times r$, the top r^2 eigenvalues of the symmetric matrix $(V \otimes U)(V \otimes U)^{\top}$ will be equal to 1, and the rest will equal 0. The eigenvalues of the symmetric matrix LL^{\top} will be equal to the squared singular values of L. We can then apply the Von Neumann trace inequality to obtain the upper bound.

The last statement follows from the Pythagorean theorem and the fact that the Σ_i is a projection of B_iA_i to the U,V subspace.

Note that we have only used the fact that the matrix $(V \otimes U)$ has singular values equal to 1; we have not used the fact that it has Kronecker product structure. On the other hand, each vector $\text{vec}(B_iA_i)$ is a sum of r_i Kronecker products and cannot be expressed as a Kronecker product. As a result, while the upper bound in the Von Neumann trace inequality is achieved if the eigenvectors of the two matrices align, the Kronecker product structure is a severe constraint and the upper bound we have provided is generous.

C. Training LoRAs

We trained LoRA adapters on 500 natural instruction tasks (Wang et al., 2022) using Mistral-7B-Instruct-v0.2 (Jiang et al., 2023a) as the base model. All LoRA adapters were configured with a rank of 16, i.e., $\forall i, r_i = 16$. We selected 10 diverse tasks manually for consistent evaluation across experiments and randomly sampled an additional 490 tasks, resulting in a total of 500 tasks. These tasks were exclusively in English (both input and output), ensuring higher quality and thorough review (Wang et al., 2022). Each task dataset was divided into training, validation, and test sets (80-10-10). Hyperparameters, such as early stopping, were tuned using the validation sets; that is, we train for five epochs and take the best-performing epoch-checkpoint per validation loss. Evaluation on the test sets demonstrated that LoRA consistently outperformed the base model in terms of both Rouge scores and loss metrics (see Table 1).

In Table 1, we compare metrics between base model and LoRA finetuning.

Table 1: Comparison of metrics before and after LoRA training across 1000 tasks.

Metric	Base Model	LoRA
Loss	4.14 ± 3.07	0.56 ± 0.58
Exact Match	1.81 ± 6.56	51.38 ± 40.90
Rouge-1	21.70 ± 19.22	68.88 ± 29.73
Rouge-L	20.62 ± 18.21	67.80 ± 30.15

In Table 3 we include all 1000 tasks that were used.

We use Huggingface (Wolf et al., 2020) in our implementation. For the base model, we use quantization with configuration:

Table 2: Main Evaluation Tasks

Task Number	Name	Type	Domain
task280	stereoset_classification_stereotype_type	classification	stereoset
task190	snli_classification	snli	image captions
task391	causal_relationship	commonsense	cause and effect
task290	tellmewhy_question_answerability	answerability	story
task1391	winogrande_easy_answer_generation	commonsense	social and physical
task1342	amazon_us_reviews_title	title generation	amazon reviews
task442	com_qa_paraphrase_question_generation	question generation	wikipedia
task620	ohsumed_medical_subject_headings_answer_generation	keyword tagging	scientific
task1598	nyc_long_text_generation	data to text	restaurants
task039	qasc_find_overlapping_words	overlap extraction	natural science

```
BitsAndBytesConfig(
    load_in_4bit=True,
    bnb_4bit_use_double_quant=True,
    bnb_4bit_quant_type="nf4",
    bnb_4bit_compute_dtype=torch.bfloat16,
)
```

and LoRA configuration:

```
LoraConfig(
    r=16,
    lora_alpha=32,
    target_modules=["q_proj", "k_proj", "v_proj"],
    lora_dropout=0.05,
    bias="none",
    task_type="CAUSAL_LM",
    init_lora_weights=init_lora_weights,
)
```

D. Avoiding Batched Matrix Multiplication (BMM)

Fast LoRA (Wen & Chaudhuri, 2024) aims to alleviate the batched matrix multiplication (BMM) bottleneck when serving many LoRAs. They propose an adapter parameterization that replaces addition with elementwise multiplication, avoiding BMM and improving LoRA throughput at lower ranks. Our JD LoRA formulation also circumvents or heavily reduces the impact of BMM as discussed below, and both individual and joint compression methods can be applied to Fast LoRAs.

In the envisioned deployment scenario, a service provider hosts a large collection of LoRAs. Upon receiving a request, each user specifies both the input data and the desired LoRA identifier. The provider then processes the base model augmented with the specified LoRA for each user's data. As a provider is batching a collection of requests for GPU parallelization, they can expect to frequently have more than one unique LoRA identifier per batch.

Traditionally, a specific LoRA is integrated into the base model by transforming $W_0 \to W_0 + B_i A_i$. Serving multiple LoRAs conventionally would necessitate maintaining and executing a separate copy of the base model for each LoRA, bringing substantial computational overhead. Alternatively, the computation for $W_0 x$ and $B_i A_i x$ can be performed independently and subsequently merged. This strategy necessitates only a single instance of $W_0 x$ computation and storage of LoRA-specific parameters rather than the entire base model.

Consider the batch processing of BAx, where boldface indicates that B_i , A_i are stacked into tensors of dimensions ($b \times b$)

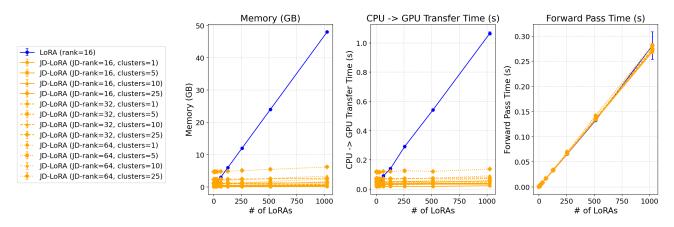


Figure 5: Memory load, transfer time, and forward-pass performance of LoRA and JD-LoRA.

 $m \times r$) and $(b \times r \times n)$ respectively, with batched data **x** shaped $(b \times l \times n)$:

$$\begin{aligned} \mathbf{A}\mathbf{x} &\leftrightarrow (b \times r \times n) \times (b \times l \times n) \rightarrow (b \times l \times r) \text{ bmm} \\ \mathbf{B}(\mathbf{A}\mathbf{x}) &\leftrightarrow (b \times m \times r) \times (b \times l \times r) \rightarrow (b \times l \times m) \text{ bmm}. \end{aligned}$$

Here, "bmm" denotes batched matrix multiplication, a known bottleneck in both throughput and latency. Consider the corresponding operations for our joint compression scheme, $U\Sigma V^{\top}x$:

$$V^{\top}\mathbf{x} \leftrightarrow (\tilde{r} \times n) \times (b \times l \times n) \rightarrow (b \times l \times \tilde{r}) \text{ broadcasted}$$

$$\mathbf{\Sigma}(V^{\top}\mathbf{x}) \leftrightarrow (b \times \tilde{r}) \times (b \times l \times \tilde{r}) \rightarrow (b \times l \times \tilde{r}) \text{ broadcasted}$$

$$U(\mathbf{\Sigma}V^{\top}\mathbf{x}) \leftrightarrow (m \times \tilde{r}) \times (b \times l \times \tilde{r}) \rightarrow (b \times l \times m) \text{ broadcasted}$$

In our optimized setup, batched matrix multiplications can be completely circumvented if the Σ_i matrices are diagonal. If not, given that $\tilde{r} \ll m, n$, any required batched matrix multiplication remains computationally inexpensive.

E. Simple Timing Experiments

In Figure 5, we present a set of simple experiments comparing the memory load, transfer time, and forward-pass performance of LoRA and JD-LoRA. These experiments were conducted across multiple clusters and various rank configurations. For memory usage, we measured all 96 LoRA modules in the Mistral model; however, for transfer time and forward-pass performance, we tested only a single LoRA module. We also implemented F-LoRA (Wen & Chaudhuri, 2024), but contrary to their reported results, we were unable to achieve faster forward-pass performance than standard LoRA.

F. GPU Memory Usage Computation for JD Compression.

The GPU memory consumption is primarily influenced by the number of parameters that need to be stored and processed during inference. In this section, we introduce the detail of how we compute the GPU consumption of our method, and how we find the number of vLLM multi-LoRA that share the same GPU utilization.

- D: Hidden dimension size (e.g., D = 4098).
- r: Rank of the shared basis matrices for compression (e.g., r = 16, 32, 64).
- N: Maximum number of LoRA modules being served simultaneously (max_lora_num).
- c: Number of clusters in our clustering method (e.g., c = 7, 10, 25).

In Figure 1, we use different JD-compression settings for serving different number of unique LoRAs. Specifically:

• Serving 4 unique LoRAs:

Ours: rank 16 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 2.

• Serving 8 unique LoRAs:

Ours: rank 16 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 2.

• Serving 16 unique LoRAs:

Ours: rank 32 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 3.

• Serving 32 unique LoRAs:

Ours: rank 64 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 5.

• Serving 64 unique LoRAs:

Ours: rank 64 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 6.

• Serving 128 unique LoRAs:

Ours: 7 clusters, rank 16 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 8.

• Serving 256 unique LoRAs:

Ours: 10 clusters, rank 16 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 10.

• Serving 512 unique LoRAs:

Ours: 25 clusters, rank 16 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 26.

• Serving 1024 unique LoRAs:

Ours: 25 clusters, rank 16 JD-Full.

vLLM multiLoRA baseline: max-gpu-lora = 28

F.1. Baseline GPU Memory Usage

The baseline for our comparison is the standard LoRA method with a rank of 16. The total parameter count for the baseline is given by:

Params_{baseline} =
$$D \times 2 \times 16$$
.

This accounts for the parameters in the LoRA-adapted layers, where the factor of 2 represents the weights and biases.

F.2. GPU Memory Usage for JD Full Method

For the Joint Decomposition (JD) Full method without clustering, the total parameter count is:

Params_{ID Full} =
$$D \times 2 \times r + N \times r^2$$
.

- $D \times 2 \times r$: Parameters for the base model adapted with rank-r LoRA.
- $N \times r^2$: Additional parameters introduced by each of the N LoRA modules, each of size $r \times r$.

The GPU memory usage ratio relative to the baseline is:

$$\mbox{GPU Usage Ratio}_{\mbox{\scriptsize JD_Full}} = \frac{\mbox{\scriptsize Params}_{\mbox{\scriptsize JD_Full}}}{\mbox{\scriptsize Params}_{\mbox{\scriptsize baseline}}} = \frac{D \times 2 \times r + N \times r^2}{D \times 2 \times 16}.$$

F.3. GPU Memory Usage for Clustering Method

When employing clustering, the parameter count changes due to the addition of cluster-specific parameters:

Params_{Clustering} =
$$D \times 2 \times r \times c + N \times (r^2 + 1)$$
.

• $D \times 2 \times r \times c$: Parameters for the base model adapted with rank-r LoRA across c clusters.

• $N \times (r^2 + 1)$: Additional parameters for each LoRA module and cluster assignments.

The GPU memory usage ratio is:

$$\text{GPU Usage Ratio}_{\text{Clustering}} = \frac{\text{Params}_{\text{Clustering}}}{\text{Params}_{\text{baseline}}} = \frac{D \times 2 \times r \times c + N \times (r^2 + 1)}{D \times 2 \times 16}.$$

F.4. Punica

In our vLLM experiments, we specifically used the Punica kernel for implementing multi-LoRA, applying our approach in conjunction with Punica's capabilities. Our custom function, add_lora_slice_with_sigma, implements the following key steps:

- 1. Initialize Buffers: Creates temporary storage for intermediate calculations if not already provided.
- 2. Apply Matrix A: Transforms x using matrix A, storing the result in buffer.
- 3. Apply Matrix Sigma: Further transforms buffer using Sigma, storing the result in buffer_sigma.
- 4. Apply Matrix B and Update y: Finally, transforms buffer_sigma using B, applies scaling, and updates a slice of y in place.

Below is the pseudocode for add_lora_slice_with_sigma, illustrating the integration:

Listing 1: Pseudocode for 'add_lora_slice_with_sigma'

```
Function add lora slice with sigma(y, x, wa t all, wb t all, wsigma t all, indices,
      layer_idx, scale, y_offset, y_slice_size, buffer=None):
        Initialize buffers if not provided
      if buffer is None:
          buffer = create_tensor(shape=(x.size(0), R), dtype=float32)
4
          buffer_sigma = create_tensor(shape=(buffer.size(0), R), dtype=float32)
       # Step 1: Apply matrix A
      dispatch_bgmv_low_level(buffer, x, wa_t_all, indices, layer_idx, scale=1.0)
       # Step 2: Apply matrix Sigma
      dispatch_bgmv_low_level(buffer_sigma, buffer, wsigma_t_all, indices, layer_idx, scale
10
       # Step 3: Apply matrix B and update y slice
      dispatch_bgmv_low_level(y, buffer_sigma, wb_t_all, indices, layer_idx, scale, y_offset
11
          , y_slice_size)
  End Function
```

G. Selecting Number of Clusters

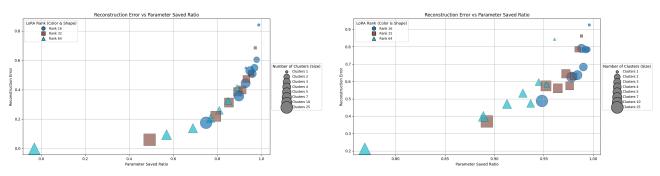
To identify optimal hyperparameters for the clusters compression method, we analyzed the relationship between reconstruction error and the parameter saved ratio for a single LoRA module, as shown in Figure 6. By comparing the results across different numbers of Low-Rank Adaptation (LoRA) configurations (100 and 500, depicted in subfigures 6a and 6b), we were able to observe the trade-off between model size reduction and reconstruction accuracy. Based on these findings, we selected the rank and number-of-clusters hyperparameters that effectively balance these two objectives. The chosen settings were then used to conduct full-scale experiments.

H. Additional Results

This section elaborates on the results that underpin the figures presented in the main text and showcases a consistent correlation across various evaluation metrics. Additionally, we assess the significance of achieving convergence and the performance of compression on new unseen LoRA models.

H.1. LoRAs of different ranks

In Table 4, we report the performance of our compression method on LoRAs with ranks uniformly sampled between 16 and 64 (mean rank of 43). In Table 5, we present compression results for LoRAs of rank 43, matching the average rank from Table 4. Because these same-rank LoRAs have an identical parameter count, they also exhibit identical parameter-saving



- (a) Recon. Error vs Parameter Saved Ratio for 100 LoRAs
- (b) Recon. Error vs Parameter Saved Ratio for 500 LoRAs

Figure 6: Comparison of reconstruction error against the parameter saved ratio for different numbers of LoRA configurations for a single LoRA module. The left subplot shows results for 100 LoRAs, while the right subplot displays results for 500 LoRAs. These plots illustrate the trade-off between reconstruction accuracy and compression efficiency, providing insights into optimal parameter settings for compression.

ratios. While performance declines slightly for LoRAs with varying ranks, our compression method still preserves over 99% of the original performance.

H.2. Privacy Ablation

We investigate whether jointly compressing certain tasks results in improved performance on tasks within that same compressed group, compared to tasks outside of the group. In other words, we examine whether information about which tasks were compressed together could be inferred from subsequent performance differences. Table 6 presents the results.

In the Compressed Together setting, tasks (task1391, task190, task280, task290, and task391) are compressed jointly with five other tasks. We then evaluate their cross-task performance within this group. In the Compressed Separately setting, the same set of tasks is each compressed alongside nine other tasks that are not part of the original group. We again evaluate the cross-task performance using the same sets of tasks, allowing us to compare and assess any differences in performance attributable to joint versus separate compression.

Table 6: Privacy Ablation: Performance on tasks compressed together vs. performance on tasks compressed separately

Source Task	Task Combination	Test Loss	Exact Match	Rouge1	RougeL
	Com	pressed Toge	ether		
	task1391 on task190	2.320	29	29.00	29.00
	task1391 on task280	1.445	10	12.00	12.00
task1391	task1391 on task290	2.662	0	0.00	0.00
	task1391 on task391	1.560	4	7.00	7.00
	Average	1.997	10.8	12.00	12.00
	task190 on task1391	1.392	0	0.00	0.00
	task190 on task280	1.647	2	2.00	2.00
task190	task190 on task290	1.838	0	0.00	0.00
task190	task190 on task391	2.622	2	2.00	2.00
	Average	1.875	1.0	1.00	1.00
	task280 on task1391	2.259	14	16.47	16.47
	task280 on task190	2.922	20	20.50	20.50
task280	task280 on task290	0.729	33	43.07	43.07
	task280 on task391	2.318	44	62.52	62.52
	Average	2.057	27.8	35.64	35.64
	task290 on task1391	0.867	36	46.58	46.58
	task290 on task190	1.553	36	36.00	36.00
task290	task290 on task280	1.036	43	43.40	43.40
	task290 on task391	0.461	59	86.33	86.33
	Average	0.979	43.5	53.08	53.08
	task391 on task1391	0.502	65	65.75	65.75

task391

Table 6: Privacy Ablation: Performance on tasks compressed together vs. performance on tasks compressed separately (continued)

Source Task	Task Combination	Test Loss	Exact Match	Rouge1	RougeL								
	task391 on task190	1.421	31	31.00	31.00								
	task391 on task280	0.417	69	69.00	69.00								
	task391 on task290	0.265	71	90.33	90.33								
	Average	0.651	59.0	64.02	64.02								
	Comp	ressed Sepa	rately										
	task1391 on task190 2.347 30 30.00 30.1												
	task1391 on task280	1.543	11	13.58	13.58								
task1391	task1391 on task290	2.477	0	0.00	0.00								
	task1391 on task391	1.253	10	18.00	18.00								
	Average	1.905	12.8	15.40	15.40								
	task190 on task1391	1.388	0	0.00	0.00								
	task190 on task280	1.603	2	2.00	2.00								
task190	task190 on task290	1.771	0	0.00	0.00								
	task190 on task391	2.326	4	4.00	4.00								
	Average	1.772	1.5	1.50	1.50								
	task280 on task1391	3.111	2	5.13	5.13								
	task280 on task190	2.711	19	19.00	19.00								
task280	task280 on task290	0.948	16	22.09	22.09								
	task280 on task391	2.449	39	53.11	53.11								
	Average	2.305	19.0	24.83	24.83								
	task290 on task1391	0.848	41	47.58	47.58								
	task290 on task190	1.355	38	38.00	38.00								
task290	task290 on task280	1.050	41	41.00	41.00								
	task290 on task391	0.463	59	86.33	86.33								
	Average	0.929	44.8	53.23	53.23								
	task391 on task1391	0.428	68	68.74	68.74								
	task391 on task190	1.507	31	31.00	31.00								
task391	task391 on task280	0.368	70	70.00	70.00								
	task391 on task290	0.269	73	91.00	91.00								
	Average	0.643	60.5	65.18	65.18								
	J												

H.3. Relative Rouge-L Performance and Compression Rate

Table 7 presents comprehensive results from the experiments underlying Figure 2 for each evaluation task. Additionally, we incorporate results using the Ties-merging benchmark (Yadav et al., 2023b), which consolidates all LoRA-adapters into a single adapter of identical configuration and parameter count; this integration significantly compromises performance.

H.4. Absolute Rouge-L Performance and Compression Rate

Table 8 provides the full results behind Table 7, but with Rouge-L scores instead of relative performance compared to LoRA.

H.5. Relative Rouge-1 Performance and Compression Rate

Table 9 provides full results for relative performance of Rouge-1, which shows the same trends as the results for relative performance of Rouge-L (Table 7).

H.6. Absolute Rouge-1 Performance and Compression Rate

Table 10 provides full results for absolute performance of Rouge-1, which shows the same trends as the results for absolute performance of Rouge-L (Table 8).

H.7. Relative Exact-Match Performance and Compression Rate

Table 11 provides full results for relative performance of exact-match, which shows the same trends as the results for relative performance of Rouge-L (Table 7).

H.8. Loss and Compression Rate

Table 12 provides full results for test loss (cross-entropy), which shows the same trends as the results for relative performance of Rouge-L (Table 7).

H.9. Agreement and Compression Rate

Table 13 provides full results for *agreement*, which shows the same trends as the results for relative performance of Rouge-L (Table 7). Note that *agreement* measures the exact match in task generations between the uncompressed LoRA model and the compressed LoRA model, rather than comparing to the task's ground truth data. The comparison is very strict and requires an exact match between the generations of the two models (LoRA and the compressed LoRA), comparing each sample one at a time.

H.10. Reconstruction Error and Compression Rate

Table 14 provides the full results of the experiments behind Figure 3 for every evaluation task.

H.11. Reconstruction Error: Trained vs. Random

Table 15 provides the reconstruction error on random (untrained) LoRA matrices. Comparing with Table 14, we find that reconstruction error is consistently higher on random (untrained LoRA) matrices than on trained LoRA matrices. This demonstrates that after training, LoRAs have a shared structure that JD exploits.

H.12. Convergence

Table 16 presents outcomes where the JD-Full algorithm is executed until convergence. Our convergence criterion is defined as follows:

$$\max\left(\|U_{t+1} - U_t U_t^\top U_{t+1}\|_{\operatorname{Fro}}/\|U_{t+1}\|_{\operatorname{Fro}}, \|V_{t+1} - V_t V_t^\top V_{t+1}\|_{\operatorname{Fro}}/\|V_{t+1}\|_{\operatorname{Fro}}\right) < \tau \tag{19}$$

where the tolerance threshold τ is set to 0.001. Due to the slow per-iteration computation times of the primary JD-Full algorithm, which quickly reaches an approximate optimum but then has a long tail of convergence for final digits of precision, we devised an alternative eigenvalue iteration algorithm (Appendix A.2) optimized for GPU acceleration. Our analysis indicates that adherence to this convergence criterion does not significantly alter the results.

H.13. Out-of-distribution Performance (LoRA-hub)

For completeness, we incorporate results using the protocol of LoRA-hub (Huang et al., 2024). That is, 100 LoRA-adapters are sampled, independent of the evaluation task, representing a measure of out-of-distribution performance. This also means that each result on a task is averaged across all 100 LoRA-adapters (as there is no *a priori* LoRA-to-task mapping). These results were obtained without normalizing the LoRA-adapters before applying the JD algorithms, a step we later identified as beneficial. We present performance comparison in Table 18. Table 17 presents the average agreement between uncompressed and compressed LoRA across 10 evaluation tasks. Results per task for JD-diagonal and JD-full are shown in Table 19 and Table 20, respectively.

From these tables, we find that the JD algorithms successfully maintain performance in this out-of-distribution context.

Table 3: List of all 1000 Tasks

Task ID	Description	Task ID	Description	Task ID	Description		
task100	concentrate all cleanurs from index to judge concentrate all cleanurs from index to judge all the cleanurs of the graphs of the material to the graphs of the material to the graphs of the material to the cleanurs of the cl	task1009 task1024		task101 task104	review and conceitoms all elements from index to j review and conceitoms all elements from index to j ple translation food pale ple translation food ple ple translation	task102 task1047	consensages sonitate generation pib translation conflicts design pib translation conflicts design pib translation conflicts design pib translation conflicts translation
task 1048 task 1062 task 1082	pib translation telugu english pib translation marathi bengali pib translation marathi hindi serro of medicits	task105 task107 task1063	ph translution beaught kinds place to the property of the prop	task1053 task1079	pib translation hindi urdu pib translation english gujarati pib translation english marahi musummodorium musummodorium marahi	mak102 mak1047 mak1057 mak1058 mak1058 mak1050 mak1090 mak11010 mak1115 mak1115 mak1125 mak1125 mak1125 mak1125 mak1126 mak1135 mak1126 mak1135 mak1146 mak1146 mak115 mak	pio translation coglisis rategos pio translatione englisis studio pio translatione englisis studio con constante rates to constante rates to franchistica en el
task1094 task110 task1106	array or presences ted translation on pt logic floxt semience generation ted translation ar it	task1095 task1101 task1109	cracks mensioned array ted translation ja gl ted translation en st ted translation en st	task1097 task1102 task111	introperational per la proposition de la constitución de la constituci	task1098 task1105 task1110	test translation en pr test translation ja lis test translation ar gl test translation be gl
task1111 task1116 task1120 task113	ted translation he it alt id is translation alt ja fil answer generation count from your of letter	task1113 task1117 task1122 task1130	ted translation he fa alt is ind answer generation alt klem ja translation alt klem ja translation	task1114 task1118 task1127 task1132	ted translation he pt alt is fill translation alt is th translation	task1115 task112 task1128 task1135	alt ja id translation and simple perfected identification alt this translation the state of the
trak (1088) trak (1094) trak (110) trak (110) trak (110) trak (110) trak (1116) trak (1116) trak (1120) trak (113) trak (113) trak (1142) trak (1147) trak (1147) trak (1150)	near recommensesse me classification near ar commensesse me classification country currency	task/10095 task/10095 task/1100 task/1100 task/1100 task/1100 task/1110 task/1117 task/1117 task/1120 task/1130 task/1144 task/1146 task/1204 task/1204 task/1204 task/1204 task/1204	and transduction by far and the process of the proc	tank109 tank1097 tank1097 tank1102 tank111 tank1114 tank1114 tank11132 tank1140 tank1145 tank1145 tank1145 tank1149 tank1152 tank1158 tank1159 tank1197 tank1197 tank1197	nor pl communicate me classification nor jap communicate me classification item check edible	task1141 task1146 task115	two matter can. The control of the
task1156 task1156 task1167 task1186	delete max mm bard analogical reasoning tools penn treebank course pos tagging me hungo classification	task1157 task1168 task1188	xwap max run burd analogical reasoning rocens for containers berown course pos tagging count max froq char	task1152 task1158 task117 task1189	bard analogical reasoning cansalors bard analogical reasoning manipulating items up i manulation on de check char in string	task1154 task116 task118 task119	hard analogical reasoning trived. cons2-bens constituents or examining an analogue of the constituent of the
task1190 task1194 task1199	add integer to list kth largest element atomic classification xattr	task1191 task1196 task1200	food vog noerveg atomic classification oeffect atomic classification seffect	task1192 task1197 task1201	food flavor profile* atomic classification oreact atomic classification xintent	task1193 task1198 task1202	food course classification atomic classification owner atomic classification word
task1203 task1207 task1210 task1214	storric classification madesmof	task1208 task1211 task1215	atomic classification sreason	task1205 task1209 task1212 task1216	atomic classification happeoperty	task1206 task121 task1213 task1217	zest text mountainen stomier charifestion desires
task1214 task1219 task1224 task1228 task1233 task1238	atoretic classification xwant tool translation en es tool translation en es tool translation en es tool translation en er tool translation en er tool translation en he tool tool tool tool tool tool tool too	task1215 task1225 task1225 task1225 task1235 task1239 task1242 task1247 task1254 task1266 task1266	domes (Casudacino Basadoveut amonic Casudacino apubbod amonic Casudacino apubbod to a compation ja bac to a compation ja bac to a compation be to a compation be to a compation be to a compation of the to a compation of the to a compation of the to a compation of the and translation of the	task1216 task1221 task1226 task123 task1236	atomic classification causes ted translation on the ted translation or on conals seet dictionary ted translation to es	task1217 task1222 task1227 task1232 task1237 task1240	atonic amove generation ted translation in on ted translation or ju ted translation or ju ted translation at ce
task1238 task1241 task1246 task1252	ted translation gl en ted translation gl ar ted translation gl pt ted translation st gl	task1239 task1242 task1247	ted translation gl fa ted translation gl he ted translation it en	task123 task1236 task124 task1244 task1248 task1255	our internation of our constant of dictionary constant are dictionary constant pair averages constant pair averages test translations of plated translations if a cold translation in pt.	task1240 task1245 task125 task1257	to characteristic et y ju to characteristic et y to characteristic e
task1258 task1263 task127	ted translation pl es	task126 task1264 task1270	ned translation it fa scan refrestred test generations command action all soft translation of pt scan refrestred test generation command action short strengs informativeness chamistation test chamistation test chamistation soft translation soft translati	tusk1266	ted translation fa pa	task 1262 task 1267 task 1272 task 1278 task 1281	
task1273 task1279 task1283 task1288	too transcence pt ra was long text generation action command all ted translation fapt ted translation pt the distribution pt librage quality classification who may re-manufacture to re-manufac	task1274 task128 task1284 task1289	ted translation pt on scan wheetered test generation command action wheet hrugo informativeness classification tree classification	task1276 task1280 task1285 task129	ted translation pt ex ted translation pt it kpa keypoint matching scan lone text remeation action command short	task1278 task1281 task1286 task1292	ted translation for the ted translation for the properties assume that the properties assume that properties as properties
task1288 task1293 task1309 task1312 task1316 task132	the measurement of the control of th	task1294 task131 task1313	while parameter varietisms with get a grane and successful long war he get a grane and successful long underly valling stade. The successful long to the successful long open religion to to translation open religion to to translation open religion to the translation open religion to the translation open religion to the successful long to the	task130 task1310 task1314 task1318	kpa keypoint mulching scale long text generation action command action long arrantereview rating chaotification country abstraction country text contract country text contract country text contract country contract contr	mak1296 mak1292 mak1308 mak1311 mak1315 mak1319 mak1322 mak1336 mak1338 mak1338 mak1342 mak1350 mak1370 mak1370 mak1370 mak1370 mak1370 mak1370	amazontroview rating classification find canno array
task132 task1323 task1327	dais text modification open subtitles hi en translation qa zre answer generation from question	task1320 task1324 task1328	country domain tld open subtiles to en translation quare relation generation from question	task1321 task1325 task1330	country continent quare quartien generation on subject relation open subtitles on the translation	task1322 task1326 task1331	country by basined profits go are question generated from moreover go are question generated from moreover go are question generated from moreover go are question generated from the good good good good good good good goo
task 1323 task 1327 task 1327 task 1339 task 1339 task 1346 task 1353 task 1360 task 1371 task 1378 task 1381	cneck stap year prixian equity evaluation corpus text completion glue cola grammatical correctness classification hind encorp translation on hi	task1340 task1347 task1354	eneck variating state distringsysty , mor test correpression compression glue ste-b similarity classification sent comp classification	task1341 task1351 task1355	country continent par re quotifus guaratimes on subject relation per requesting guaratimes on subject relation prices are guilty evaluation corpus guarde classifier points are guilty evaluation corpus and corpus summeratation and corpus summeratation and corpus summeratation and corpus summeratation and consideration subjects are consideration and consideration flustery and an experimental consideration flustery and its retainment	task13342 task1352 task1359	pertuta equity evaluation ecopia settament exacutier attazion in erivolvas tille hind encorp intanalation in en mattere sense annote generation
task1390 task1371 task1378 task1381	namer sense multiple choice qu generation newscomm translation quarel correct answer generation quarel incorrect orbion reneration	task1364 task1373 task1379 task1382	hans answer generation newscerm translation quarel incorect answer generation quarel incorect answer assert	task137 task1375 task138 task1383	detoxilying-time classification toxicity new-cornen translation detoxifying-time classification flaency quared write incornect answer	task1370 task1377 task1380 task1384	nerwiccerni classificazion nerwiccerni translazion quared correct optimi generation deal orno dislore classificazioni
task1385 task139 task1394	anli r1 entailment detectifying-less classification topicality meta were task classification	task1386 task1390 task1397	anli r2 entailment was fixed conference curepa cede tra fr on translation	task1387 task1391 task1398	ardi r3 entailment winogrande cosy answer generation obspa question generation	task1389 task1393 task1399	hellaswag completion sapergluc copa text completion obea assures generation
task1404 task141 task141	date conversion odd-man-out classification category odd-man-out classification no category	task1405 task1412 task1420	ova incorrect answer generation find median web questions question answering matthing general	task1406 task1418 task1421	origin seminence generation ich semalaket element bless somantic relation classification mathpa other	task1409 task1419 task1422	cince valuati quis immanyyy dati text generation mathag gain mathag physics
task1423 task1427 task1431 task1435	mathqa geometry country region in world head qu answer generation ro sis matallel language translation to to en	task1424 task1428 task1432 task144	mathqu probability country surface area head qu language translation en to ex subissa succision nanvecimi	task1425 task1429 task1433 task1443	country iso numeric evalution semantic relation classification head on largeage translation es to en string to number	task1426 task143 task1434 task1444	country independence year odd-man-out classification generate calegory head qu-classification round roomer of two
task1445 task1449 task1453	closest integers disease entity extraction be5cdr dataset person entity extraction btc corpus	tmk1446 tmk145 tmk146	fasthest integers als argument similarity death penalty als argument similarity gan control	task1447 task1451 task147	drug extraction ade drug dose extraction afs argument similarity gay marriage	task1448 task1452 task1479	disease entity extraction nebi dataset locazion entity extraction bic corpus organization entity extraction bic corpus
trakt 199 trakt 1994 trakt 400 trakt 400 trakt 404 trakt 422 trakt 422 trakt 423 trakt 433 trakt 443 trakt 444 trakt 443 trakt 443 trakt 444 trakt	chemical extraction chempost dataset organism substance extraction anem dataset bengali personal hate speech binary classification	task1484 task1488 task1491	gene extraction lineaeus dataset sarcasmidetection healtine classification bengali political hate speech binary classification	mak1271 mak1271 mak1273 mak1285 mak1285 mak1380 mak1380 mak1381	and it of conductors; and it is conductors;	maki 1993 maki 1999 maki 1403 maki 1409 maki 1419 maki 1422 maki 1426 maki 143 maki 1434 maki 1444 maki 1445 maki 1452 maki 1453 maki 1453 maki 1453	Coll extraction mera dataset als argament quality death perulty bength alse speech classification
task1502 task1506 task151	accent arug evens classification hatexplain classification celebrity minimal deb span toraga find location easy clean	task1496 task1503 task1507 task1510	agend with centred uniford work and of the control	task1498 task1504 task1508 task1517		task1505 task1509	m n pyrman glastity gast control received in the control received in the control received received relation classification condition anisotypus limit asserter generation
task1519 task153 task1538	qa sel question generation tomqu find location bard clean malayalam offenseval dravidian classification conflictor prayer man extra	tmk152 tmk1533 tmk1541	evaluation selutions extractions tomage find lecution many motion daily disable format classification was pleasy-classification place answer generation and text monitoriation place answer generation and text monitoriation	task1520 task1534 task1542	quart answer generation daily dialog question classification every ith element from starting wars mayor more requires the	task1529 task1537 task1544	sciral(1,1 classification tural difference) devident classification cont/1002 manuel entity recognition network generation
task1545 task1551 task1560 task1566	special content of the procession of the content of	mal 279 1 mal 27	jileg answer generation and text modification propura question generation	task1517 task1530 task1534 task1542 task1549 task1549 task1554 task1558 task1574 task158 task158	larent calendarian yot el laware gamenta yot el laware gamenta wenyin helimat inne natuta helimat laware helimat laware helimat laware helimat laware perpara katindarian	task1519 task1529 task1537 task1544 task155 task156 task1565 task1575 task1575 task1580	summer out amount prosection where the control of t
task1566 task1572 task1576 task1581 task1585 task1593	samson sensorsy anazon reviews multi english language classification eque-perturbed answer generation rest® bytectrom contraion	task1573 task1577 task1582 task159	samsum classification amazon reviews multi japanese language classification bless hyperayan generation check freeasers of words in sentence roir	task1574 task158 task1583 task1590 task1595	arrazzon reviews multi language identification count frequency of words bless recrospen classification dislocency but recreation	task 1575 task 1580 task 1584 task 1592 task 1596	arrazon reviews multi sertiment classification capas-perturbed quotions generation evalution merconym classification evalution merconym classification value arrowers toosic classification
task1593 task1598 task1601 task1605	yahoo answers topics classification nyc long text generation webquestions answer generation	task1594 task1599 task1602	and text modification propun question permittin propun question permittin auxuro reviews multi japanes laspuage clausification bless hypersup programies bless hypersup programies bless hypersup programs yahro autworn topics question promittion smcalibor chandication webspecies question promittion webspecies question promittion to the text clausification	task1595 task160 task1603 task1607	event2mind text generation 1 replace letter in a sentence succillow sentence generation	task1596 task1600 task1604 task1608	eventCarind lext generation 2 smealther sentence generation other text Constitution
task1609 task1617		task161 task1618 task1631	ethos text classification count words containing letter ce alligned classify let eng	task1610 task162 task1645	ethen text classification with the control of the c	tusk1622 tusk1622	squais en annover generation ce alligned translate eng sel dirdi qui text modification dataset; card for extalorias independence corpus text classification
task1648 task1655 task1660		task1649 task1656 task1661 task167 task167	opus books en-do translation gonuq answer generation super glue classification states and translation secretics	task1657 task1662	opus books en ex franchitien goong question generation cedr re classification nd nonder bits test prodification	task166 task166	opus beoks ca-en translation clasis yenterice generation trainplecops specifien generation trainplecops specifien generation
task1669 task1677 task1686 task1691 task1706	nikh trandation super glie question generation super glie question generation super glie question generation superide a translation superide a translation superide a translation superide superide superide quel narun translation q	task1689 task1703	mathiga answer selection end amaza translation lopocch textmedification	task1670 task168 task169 task1704	strategyaş spastien decumposition strategyaş sentence generatien iş specch textmodification	task1676 task1685 task1690 task1705 task1712	manyolik translation qela anaza translation popeció classification
task1713 task1721	convide sentence generation civil comments obscenity classification civil comments severtoxicity classification	task1714 task1722	methy is avever edection gold attent translation gold attent translation gold translation con- gold translation con- gold translation con- gold translation con- constant increase generation coli comments threat classification coli comments threat classification coli comments threat classification coli control translation gold t	task169 task1704 task1711 task172 task1723 task1727	polis use generates upl translation on fail civil comments sexually explicit classification waps what is the effect	task1720 task1724	point automation for the property of the prope
task1729 task176 task180 task190	break decompose questions intervention extraction	task1731 task177 task181 task192	quartz question anovering para-emit paraphrasing onference extraction heterotas sentimes rementation	task174 task178	up de translation on ja quartz question answering thyric generation tentingent 40 classification	task175 task179 task184 task196	spli translation on pl participant extraction break generate quantities seriment 410 answer enceration
task190 task201 task207 task211	intervention contaction soli classification multi incutral classification max element fasts logic 2text classification	task192 task202 task208 task219	tonicione continuos precration mali contradiction classification combinations of list procedures of list poculories title answer generation	task195 task205 task209 task209	remove even elements stancedetection classification controsspa passage inappropriate binary	task196 task206 task210 task223	collular conjecture logic Zient structured text generation quartz explanation generation
task024 task246 task250	contemp answer generation dream question generation spl translation en ar	task243 task247 task251	accusation generation sel intersection dream answer generation spl translation on fi	task244 task249 task254	are statements in set union count clements in a count clements in set union enhanced was pronoun discernbiguation spl translation in on	tusk245 tusk025 tusk255	Consisting species government. check presence in set intersection commonqui incorrect answer generation spl translation it en
task257 task263 task269	xpl translation ar en xpl translation pl en cxsg counterfactual story generation committee local classification	task260 task265 task270 task275	spl translation zh en puper reviews language identification cseg counterfactual context generation mbarout aue manufactuage generation	task261 task266 task272 task276	upl translation on en paper environ reviewer perspective classification europeal translation environd user classification	tusk262 tusk267 tusk273 tusk277	and translations in em consciousate and reverse all elements from index i to j currepail classification deferred a vertices convenient strendings
trak211 trak227 trak6024 trak6024 trak6024 trak6240 trak6240 trak6257 trak6263 trak6263 trak6264 trak6274 trak6274 trak6274 trak6274 trak6274 trak6274 trak6276 trak6276 trak6276 trak6276 trak6376 trak6376 trak6376 trak6376 trak6373 trak6373 trak6373 trak6373	mail south of the control of the con	mak219 mak228 mak248 mak248 mak248 mak248 mak251 mak251 mak256 mak279 mak279 mak279 mak285 mak279 mak286 mak286 mak286 mak286 mak286 mak286 mak286 mak316 mak316 mak317 mak317 mak317 mak318 mak318 mak318 mak346 ma	and contraction for the institution and contraction for the institution of the institutio	mik022 mik229 mik224 mik229 mik244 mik229 mik244 mik259 mik254 mik256 mik256 mik266 mi	names over determine control process of the c	mai/223 mai/235 mai/245 mai/245 mai/245 mai/245 mai/252 mai/252 mai/252 mai/252 mai/273 mai/233 mai/293 mai/293 mai/293 mai/293 mai/293 mai/293 mai/294 mai/394 mai/314 mai/344 mai/344 mai/344 mai/344 mai/344 mai/344	strangener programmer god inner translation
task294 task298 task305 task312	storycommoneneme motiv text generation storyclose correct end classification jeopurdy answer generation normal curoparl sv en translation	task295 task299 task306 task315	serincial 2009 task4 commensione neasoning desyclopes senfence generation jeopardy answer generation detable curropart aven language identification	task296 task300 task307 task316	storyclone coerect end classification storyclone coder generation geopastly answer generation final crows-pains classification stereotype	task297 task304 task308 task317	storyclose incorrect end classification instruction flow blead resolution jougnarly arever generation all cross-spains classification storetype type cross-spains classification startedype type
task318 task322 task326 task933	stereoset classification gender jigsaw classification threat jigsaw classification threat	task319 task323 task327	stereoset classification profession jigsaw classification sevanily explicit jigsaw classification toxic jigsaw classification toxic	task320 task324 task328	stemoset classification race jigouw classification disagree jigouw classification insult buttered classification insult	tusk321 tusk325 tusk329	stereoret classification religion jipasu classification religion jipasu classification identify attack jupa classification jup
task335 task340 task344 task035	hateeval classification aggressive en winomt classification gender pro hybriden answer generation	task336 task341 task345	finitered classification aggressive or winoms classification greateration bybridga answer generation	task337 task342 task346	hateeval classification individual en wazent classification profession pro hybridga classification	task034 task343 task347	winogrande question modification object winosmi classification profession and hybridaja incorrect anower generation
task354 task358 task364	winogrance question meaninciation person casino classification negotiation no need casino classification negotiation uv part regard social impact classification	task355 task359 task365	worster consideration geniary acertimatesty peo cation classification negotiation order need symbolic remove content of the content of the symbolic remove colds symbolic remove colds	task356 task362 tssk366	winterm cassincation geneter incrementarity and castno classification negotiation sucff need spolin yexand protept response sub-classification synthetic return primes	task357 task363 tssk367	casion Cassification negotation enal pre- casion Cassification negotation email talk ss2: polarity classification synthetic remove flusts
task368 task372 task376	synthetic even or odd calculation synthetic palindrome numbers reverse order of weeds bools was no mention	task369 task373 task377	synthetic remove odds synthetic round tens place remove words of given length bods continue mention	task370 task374 tssk378	synthetic remove driviable by 3 synthetic post or neg calculation reverse words of given length behiving sensor on the synthetic posts of	task371 task375	syndhesi tentrer laine. Leading for effective and the control of
task380 task384 task388 task391 task399	reviews orace or weens beed by yes no question socializing question classification socyate token classification causal relationship semeral 2018 task! tweet sadness detection mickey it sentence perturbation generation	task381 task385 task389 task393	henines works of poers aregun boulq quantum generation boulq quantum generation korque generate temporal quantion plausable round generation purso paraphrase classification miskey at sentence perturbation generation	task382 task386 task039 task397 task403	que find overlapping words	task383 task387 task390 task398 task406	semencal 2018 taskl irong classification tompac text span selection semencal 2018 taskl tweet joy detection
task408 task413 task417	mickey ex sentence perturbation generation	tak399 tak490 tak400 tak414 tak424 tak424 tak424 tak438	pairs paraparane Cantillandino mickey at sentence perturbation generation mickey at sentence perturbation generation hinderglish corpora is on translation	task410 task415 task425 task043	words, commonwearts inferenced words on commonwearts inferenced words on the commonweart inferenced was sufficient to the commonwearts of the comm	task411 task416 task426	mikkey y i mikinete pisatuloides generation mikkey y i sentence perturbation generation mikkey de sentence perturbation generation hinderglish coppera ha-en classification
task428 task431 task437 task440	senteral inversion senteral object count alt en ja answer generation one mi marallel corrus nu-en classification	task429 task434 task438 task441	senteral tense alt en hi answer generation eng guj parallel corpus en gu translation eng guj parallel corpus un-en lametune identification en gu parallel corpus un-en lametune identification	task435	executial terms answering incomplete questions alt on ja translation eng gaj parallel coepus gu on translation com on namebrase onession encation	task430 task436 task044 task446	mekor de serances peturbalento praneration hundergish corporate ben exclusivations hundergish corporate ben exclusivations als ja or translation constant turns industifying casential words
trak-440 trak-448 trak-452 trak-456 trak046 trak047	open paracrawl on il translation open paracrawl on ig translation matrix intention classification matrix intention crassifien beginne	tmk441 tmk449 tmk453 tmk457 tmk460 tmk471	opia paracrawi (g en translation wa ag asswer generation matrex conditional classification	tmk442 tmk045 tmk454 tmk458 tmk461 tmk472	miscelláneous sentence pumphrassing swag incorrect answer generation matres negation classification	task446 task450 task455 task459	opus paracrawl so en translation swag contest, peneration matres static classification
task477 task490	semenda internion dal na ja asserva postentiam ora pia pradici oraza spe-net lassificación ora pia pradici oraza spe-net lassificación ora pia pradici oraza spe-net lassificación materia interiora classificación materia interiora classificación materia interiora classificación materia interiora classificación materia interiora construir general operation materia produción position or dil alphabetacia demana in los ageneral sianas construirán materia materi	task483 task491	makey is electrone production generation makey as universe production generation to the control of the contro	task472 task488 task492 task496	spirity ejection journame centra; all allybridged elements from list in order activa; all allybridged elements from list in order activa; all allybridged elements in list separately activa; all ally activates activates activates activates and activates	task459 task462 task462 task475 task489 task493 task497	contain them to destribe queented works only per personnel on the traditions makes the traditions makes the traditions makes to the contained makes to the contained personnel on the contained personnel on the contained personnel on the contained personnel on the contained makes queen personnel makes queen per
task499 task506 task513	extract and add all numbers from list extract and add all numbers from list position of all alphabetical elements in list argument stance classification	task050 task507 task514	position of all numerical elements in list	task504 task509 task515	count all alphabetical elements in list could ed all alphabetical and numerical elements in list separately senteral odd word out	task505 task512 task516	
task517	position of all alphabetical elements in list argument stance classification emo classify emotion of dialogue curoparl en en translation alt translation in en alt translation in em alt translation on the disconfiner (availleration)	task/95 task/507 task/514 task/518 task/531 task/537 task/543 task/560	argument consequence classification emo different dialogue emotions exerceparl ex en translation alt translations the em alt translations the em alt translation en emit.	traik-504 traik-509 traik-519 traik-515 traik-521 traik-521 traik-539 traik-539 traik-547 traik-561 traik-574 traik-574 traik-579 traik-583 traik-590 traik-594 traik-590 traik-594 traik-691 traik-691 traik-691 traik-691	trivia question classification europarl en-ex-classification alt translation ens en alt translation ensk en	mai-596 mai-512 mai-512 mai-512 mai-512 mai-512 mai-513 mai-513 mai-523 mai-523 mai-5241 mai-5263 mai-5263 mai-5263 mai-5264 mai-5284 mai-5284 mai-5291 mai-5291 mai-5296 mai-5294 mai-5296 mai-	find if numbers or alphabets are more in list current or-an language identification alt translation lish on disconfuse extensor generation disconfuse extensor generation
inde.535 inde.545 inde.551 inde.551 inde.568 inde.576 inde.581 inde.581 inde.585 inde.580 inde.590 inde.590 inde.590 inde.600 inde.600 inde.600 inde.600 inde.600 inde.600 inde.600 ind	alt translation on the discofuse classification circa question generation	task560 task565 task573	alt translation en entk circa answer generation air dialogue classification consisting tishen classification	task561 task566 task574	alt translation en by circa classification air dialogue sentence generation serialism classification	task563 task567 task575	discribuse anover gineration circs tal generation circs tal generation and along a classification in challenge classification in challenge classification in the challenge classification in t
task581 task585 task589	sociality question generation preposition classification amazonfood summary text generation	task582 task586 task059	naturalquestion answer generation amazoniood polarity classification ropes story generation	task583 task587 task590	udeps cng course pos tagging amazonfood polarity correction classification amazonfood summary correction classification	task584 task588 task591	udeps eig fine pos tagging amazonfood raing classification seig anteres potentation
task592 task596 task605 task609	discofone chassification cract aquestion generation promotion contract question generation socializing questions generation proposition chassification manuscriptical transmitte postation mencha speniesing generation metcha speniesing generation field the langual common subsequence in you list found to langual common subsequence in you list field the langual common subsequence in you list metch speniesing the language of the location field the langual common subsequence in you list plephoch repeat copy lugic field i clements.	tmk585 tmk573 tmk577 tmk582 tmk586 tmk596 tmk593 tmk600 tmk600 tmk610 tmk617 tmk620 tmk627	all translations much consistent and translations are consistent as a construction of the construction of	task594 task601 task607 task614	all manifolium eith en international either international either art delinging sentence generation art delinging sentence generation art delinging sentence generation authority organization authority organization authority organization authority organization authority organization also generation al	task595 task604 task608 task615	Twice creations described in the control of the con
task616 task062	cota classification bighench repeat copy logic xlwic sentence based on given word sentence generation first i elements	task617 task620 task627 task630	amazoneview category text generation obsumed medical subject headings answer generation white word with some meaning sentence generation dbreefal. It classification	task618 task622 task628	replace alphabets in a list by their position in english alphabet	task619 task625 task629	obsured abstract life generation thick tree of lefts amover generation depedia 14 chaosifectation depedia 14 chaosifectation depedia 14 chaosifectation depedia 14 chaosifectation
task633 task639 task642	dbpedia 14 answer generation	task630 task636 task664 task6643	extract and sort unique alphabets in a list all elements except first i refresed classification	task631 task637 task640 task645	Arms: whole 4 was undersoon assuming dominate granulation of popular 4 incorrect answer generation extract and sort unstque digits in a list could chaosification summersurization and the popular form of the popular form.	task632 task638 task641 task648	rralli wex classification estil classification answer generation
task642 task665 task654 task660 task666 task671	ment were their interance generation coult classification timelerared consideral sentence classification bible face translation miran face translation miranla answer generation astronomy	task643 task651 task657 task662 task667 task672	refrest denotedam greate in translation global were for translation global were for translation global were for translation global were for translation management mana	task645 task652 task659 task664 task667 task673	memorization property of the translation under privacy pells, that propine translation under privacy pells, that propine translation under privacy periodic translation under periodic periodic pells propine under periodic periodic pells propine under periodic periodic pells propine under periodic pells pells under periodic under periodic periodic pells pells under under periodic periodic under periodic perio	task648 task653 task066 task665 task670 task674	
task671 task675 task681 task685	bible f. on translation miran fa on translation miran fa on translation miran fa on translation miran fa on translation anti-type (and prevention) goody well/formed query sentence generation loopy of multipoplation text (classification redelige miranslation of the contraction of the contraction miranslation of the contraction of the contraction miranslation answer generation conceptual physics miranslation answer generation conceptual physics miranslation answer generation of contraction (loop)	task679	numericence bope odi emplish text classification online privacy policy text classification manuful answer moneration college biology	task673 task668 task683 task687	georgie wellformed query classification abdactiventi incorrect answer generation online privacy policy text purpose answer generation munito answer remention college rhomistry	task684 task688	immerior langer (maniferer) clauselation analyting quasing memoriant perpendiant personal personal person
task689 task693 task697	mentalu answer generation college mathematics mentalu answer generation conceptual physics mentalu answer generation formal logic	task686 task669 task694 task698 task701 task706	abductiventi classification manuful answer generation econometrics manuful answer generation global facts	task691 task695 task699 task703 task707	manula answer generation college physics manula answer generation electrical engineering manula answer generation high school biology	task692 task696 task696 task704 task704	munda anewer generation computer security numba anewer generation elementary mathematics adoktorient incorrect classification and the security of the security for the security for the security of the security for the security of the secur
task700 task705 task709 task714 task718	amendo havener generation (contol legal contents) and a control generation (contol legal contents) and a contents of the contol contents of the contents of th	task/01 task/06 task/071 task/15	mmentu answer generation high school computer science mmendu answer generation high school mathematics abductivenil answer generation mmendu answer generation international law	task703 task707 task710 task716 task072	manular answer generation high school geography minuth answer generation high school interseconomics minuth answer generation high school statistics minuth answer generation jurisprudence	task704 task708 task713 task717 task720	shabot only informed catalogists must be a series of the catalogists must be a series processed by the class of prices must be a series prices of the catalogists must be a series of the catalogist
task721 task725	mentalu answer generation machine learning mentalu answer generation medical genetics mentalu answer generation matrition commonseanceja answer reneration	task707 task715 task719 task722 task726 task731	manufu answer generation runtugement manufu answer generation random topic manufu answer generation philosophy manufu answer generation reofessorsal roychol-sev	task723 task727	abductivenii anviver generation ' mmmlu anviver generation mend disputes mmmlu anviver generation pechistery mmmlu anviver generation multic relations	task724 task728	menta asswer ponenzion mirkeing menta asswer ponenzion mendo lecurion menta asswer ponenzion mendo lecurion menta asswer ponenzion predevional accounting menta asswer ponenzion predevional accounting menta asswer ponenzionio necurity studios
task734 task739	mmmlu answer generation sociology lhorstq question generation lhorstq answer generation frequency	task735	manulu answer generation us foreign policy squad1.1 question generation at2 arithmetic questions arithmetic	task736 task740	mining answer generation protect relations mining answer generation virology housing answer generation quantity yelp restaurant review classification	task737 task741 task075	mmnla navwe generation world religious lhoest arewer generation place squadd 1 amore generation
task750 task754 task761 task761 task774 task774	equal financia circuic answering avamp continue-division question answering app review classification aplash explanation to sql	task755 task762 task770	ind longest substring and replace its scretcd lowercase version in both lists emea fr sk translation puresx english text modification	task740 task756 task764 task771 task776 task078	find longert substring and return all unique alphabets in it emea by el classification pawsx kerean text modification	task753 task076 task769 task772	n varge mensens upan nöbli störversteg spilands correcting sejl mistake opd stammarization pawss french dest mosification
task774 task778 task785 task079	pawex german text modification pawex english french translation pawex korean spanish translation consila concat strangs	task775 task779 task787 task790	pawox chanese text modification pawox english spanish translation pawox korean chinese translation pawox fenceh korean translation	task776 task678 task788 task792	pawax japanese text modification all elements except last i pawax kocean japanese translation pawax french german translation	task784 task789 task794	parox enguno luceau translution parox kuozan french translution parox french english translution parox french pignaese translution
IndC785 Indc079 Indc795 Indc079 Indc795 Indc080 Indc080 Indc080 Indc080 Indc080 Indc081 Indc830 Indc830 Indc830 Indc830 Indc830 Indc830 Indc857 Indc869	pawex spanish english translation piqu answer generation pawex german chinese translation	mm (2.45 mm) (2.	mentals was one generation enable to the committee of the	mat/788 mat/792 mat/997 mat/993 mat/997 mat/993 mat/997 mat/993 mat/813 mat/819 mat/813 mat/813 mat/823 mat/824 mat/825 mat/826 mat/826 mat/826 mat/827 mat/837 mat/838 mat/8991 mat/8391	Boood, some flementen quantify and the properties of the properti	maid:7:00 maid:7	winger addition quident automatique of the control
task810 task815 task820 task083	p-week crimese spanish translation pawex japunese french translation proton answer generation babi tl single supporting fact answer generation	task811 task817 task821 task830	purest, enterior german translation purest, japanese german translation protoay apersion generation poleval2019 mt translation	task813 task819 task823 task831	powex japanese enguen translation per sentiment classification pexuan-tigender sentiment analysis giga fron classification	msk814 msk082 msk828 msk832	puras propused descript Internation habit 1 single supporting fact question generation cops commonstense cause effect polyacity for the properties of the polyacity for the absolute fact of the polyacity for the polyacity for the polyacity fact of the polyacity for the polyacity fact of the polyacity fact
task833 task839 task843 task843	poem sentiment classification oft classification financial phrasebank classification synthetic longest malindrome	tmk834 tmk084 tmk844 tmk851	multidataset classification bubi t1 single supporting fact identify relevant fact francial phrasebank classification without enablishy evens	tank835 tank840 tank846 tank857	muthdataset answer generation para pdt en ex translation patencique dissification synthetic multiply odds	task838 task841 task885 task846	cit Clavatezion pura pid de or translation umaturas aktivih arithmetic com ai 2 Catalifazion
task857 task860 task864	inquisitive question generation prost need generation audit singleop question answering	task858 task861 task865	inquisitive span detection proof may answers generation manyes addeds question answering	task859 task862 task866	prost question generation audis multidist question answering maways multidist question answering	task086 task863 task867	translated symbol aribmetic audor malitop question autoverring marsps malitop question autoverring marsps malitop question autoverring
task868 task874 task879 task089	mawps sangroup question answering opus shousansy sr schema guided date8 classification swap words verification	task875 task875 task888 task890	m — operator autous neutratelic emotion classification identify typo verification gend classification	task872 task877 task888 task891	opus xionamoy minumen eng xhosa kdel translation review classification gap coreference resolution	task878 task889 task892	vpm anomary TRABARION MONA CITY folder translations generations classification gpr reverse conference resolution
task899 task902 task908	gap nll the blank construence resolution freebase qu topic generation deceptive opinion spam classification dialogic identify familial relationships	task996 task990 task903 task909	mum uniquage classification equation learner algebra deceptive opinion spum classification dialogue prevalent speakers	task900 task904 task901	mecesse qu lopic question generation froebase qu category classification hate speech offensive classification all elements from index i to j	task901 task905	
task911 task922 task976	ultailige tachtisty names retranscraps event hinds would generation coached cone pref word generation coached cone pref word generation coached cone pref word generation contain commandate into contain commandate into contain the preference of the contained phrispoch are test unto completion contain remove duplicates pht translations word materials	task913 task923 task927 task927	bianet translation event2mind classifier yelp negative to positive style transfer with anno style transfer	task914 task924 task928	all elements from index is 10 busine translation in 10 busine translati	task092 task0925 task925 task936 task956 task963 task967 task969	institute colorisational makes colorisation consideration per disconsideration personal control and colorisation personal control and colorisation personal control and colorisation personal colorisation provided personal licençation in many terms of personal licensideration in many terms of personal licensideration indicates in licensideration licensi
task993 task994 task996 task964 task997	consile calculate mean consile list index subtraction librispeech are text auto completion	task965 task965	conala max absolute value ancora-ca-ner text auto completion libriopeech asr missing word prediction	task934 task955 task962 task966 task982	wiki auto style transfer ancora-ca-ner missing word prediction raletaker fact checking based on given context	task956 task963 task967	lectecede 420 streng pass word check librispeceds aur next word reed.ciento reletakar incorrect fact generation based on given paragraph
task997 task990	pib translation urdu marathi	task998 task991	pib translation english tamil	task982 task995	pib translation tempali english	task996	reverse elements between index i and j pib translation english bengali

Table 4: Results for Different Numbers of Clusters (different LoRA-ranks)

Metric	Uncompressed	1 Cluster	2 Clusters	4 Clusters	8 Clusters
Loss	0.5009	1.6023	0.9986	0.5603	0.5000
Exact match	69.6103	37.3392	60.1071	66.6474	69.6289
ROUGE-1	79.5755	50.3598	71.5644	76.2930	79.0576
ROUGE-L	78.9355	49.7491	71.0193	75.6152	78.3958
Recon. error	0	0.8311	0.6990	0.4846	0.3246
Agreement	1.0	36.6706	62.6699	72.2231	80.5545
Exact match ratio	1.0	0.5724	0.8079	0.9219	0.9541
Relative ROUGE-1	1.0	0.6220	0.8533	0.9406	0.9917
Relative ROUGE-L	1.0	0.6221	0.8550	0.9404	0.9915
Param. saved ratio	1.0	0.99	0.98	0.97	0.94

Table 5: Results for Different Numbers of Clusters (same LoRA-ranks of 43)

Metric	Uncompressed	1 Cluster	2 Clusters	4 Clusters	8 Clusters
Loss	0.5764	0.7767	0.6277	0.5841	0.5659
Exact match	61.7746	53.2000	60.2000	61.5000	61.3000
ROUGE-1	79.9322	74.7695	79.4621	80.6950	80.4298
ROUGE-L	77.7961	72.4257	77.2141	78.3369	78.1883
Recon. error	0	0.7333	0.5594	0.3999	0.2502
Agreement	1.0	0.0000	6.6667	6.6667	6.6667
Exact match ratio	1.0	0.8175	0.9261	0.9618	1.0206
Relative ROUGE-1	1.0	0.9560	1.0153	1.0310	1.0241
Relative ROUGE-L	1.0	0.9501	1.0118	1.0268	1.0231
Param. saved ratio	1.0	0.99	0.98	0.97	0.94

Model Type	Method Type					Ta	sks					. Average	Para. Saved
Woder Type		task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	Average	raia. Saveu
	base lora	$\begin{array}{c} 0.26 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.19 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.11 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c c} 0.11 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.19 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.28 \pm 0.21 \\ 1.00 \pm 0.00 \end{array}$	1.00/1.00 0.00/0.00
TIES	10 50 100 500	$\begin{array}{c} 0.81 \pm 0.00 \\ 0.59 \pm 0.00 \\ 0.55 \pm 0.00 \\ 0.37 \pm 0.00 \end{array}$	$\begin{array}{c} 0.57 \pm 0.02 \\ 0.41 \pm 0.00 \\ 0.40 \pm 0.00 \\ 0.26 \pm 0.00 \end{array}$	$\begin{array}{c} 0.45 \pm 0.04 \\ 0.18 \pm 0.05 \\ 0.20 \pm 0.05 \\ 0.01 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.10 \pm 0.01 \\ 0.03 \pm 0.01 \\ 0.01 \pm 0.02 \\ 0.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.83 \pm 0.01 \\ 0.91 \pm 0.01 \\ 0.88 \pm 0.00 \\ 0.83 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.47 \pm 0.00 \\ 0.31 \pm 0.00 \\ 0.33 \pm 0.00 \\ 0.29 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.69 \pm 0.01 \\ 0.65 \pm 0.00 \\ 0.64 \pm 0.00 \\ 0.57 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.57 \pm 0.00 \\ 0.62 \pm 0.00 \\ 0.57 \pm 0.02 \\ 0.37 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.82 \pm 0.01 \\ 0.32 \pm 0.04 \\ 0.01 \pm 0.00 \\ 0.01 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.85 \pm 0.00 \\ 0.84 \pm 0.00 \\ 0.82 \pm 0.00 \\ 0.43 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.62 \pm 0.23 \\ 0.48 \pm 0.28 \\ 0.44 \pm 0.30 \\ 0.31 \pm 0.26 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 1.00 1.00 / 1.00
SVD	SVD 2 SVD 4 SVD 8 SVD 16	$\begin{array}{c} 0.98 \pm 0.03 \\ 0.99 \pm 0.04 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.07 \pm 0.02 \\ 1.04 \pm 0.01 \\ 1.02 \pm 0.01 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.98 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 0.99 \pm 0.02 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.00 \pm 0.10 \\ 0.99 \pm 0.08 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.00 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.00 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.00 \pm 0.04 \\ 1.00 \pm 0.03 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	0.88 / 0.88 0.75 / 0.75 0.50 / 0.50 0.00 / 0.00
10 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 1.02 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.01 \\ 1.05 \pm 0.01 \\ 1.03 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.00 \pm 0.01 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.99 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.96 \pm 0.00 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.02 \pm 0.02 \\ 0.97 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.13 \pm 0.03 \\ 1.05 \pm 0.03 \\ 0.99 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.99 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.98 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.05 \\ 1.00 \pm 0.03 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	1.00 / 0.90 1.00 / 0.80 1.00 / 0.60 1.00 / 0.20 1.00 / -0.60
10 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 1.02 \pm 0.00 \\ 1.02 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.06 \pm 0.01 \\ 1.04 \pm 0.01 \\ 1.03 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.99 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.98 \pm 0.00 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.02 \\ 0.96 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.07 \pm 0.00 \\ 1.00 \pm 0.02 \\ 0.98 \pm 0.01 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.01 \pm 0.00 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.03 \\ 1.00 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	1.00/0.90 0.99/0.79 0.97/0.57 0.88/0.07 0.50/-1.10
50 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.98 \pm 0.04 \\ 1.00 \pm 0.02 \\ 1.02 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$\begin{array}{c} 0.98 \pm 0.01 \\ 1.02 \pm 0.02 \\ 1.05 \pm 0.02 \\ 1.08 \pm 0.01 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.92 \pm 0.06 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.84 \pm 0.07 \\ 0.96 \pm 0.01 \\ 0.99 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.92 \pm 0.02 \\ 0.95 \pm 0.02 \\ 0.97 \pm 0.00 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.68 \pm 0.05 \\ 0.84 \pm 0.02 \\ 0.99 \pm 0.01 \\ 0.98 \pm 0.01 \\ 0.97 \pm 0.03 \end{array}$	$\begin{array}{c} 0.87 \pm 0.10 \\ 1.00 \pm 0.13 \\ 1.09 \pm 0.03 \\ 1.11 \pm 0.03 \\ 1.01 \pm 0.03 \end{array}$	$\begin{array}{c} 0.88 \pm 0.07 \\ 0.98 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.83 \pm 0.02 \\ 0.88 \pm 0.01 \\ 0.90 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$\begin{array}{c} 0.89 \pm 0.10 \\ 0.96 \pm 0.07 \\ 1.00 \pm 0.05 \\ 1.01 \pm 0.04 \\ 1.00 \pm 0.02 \end{array}$	1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84 1.00 / 0.68
50 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.99 \pm 0.04 \\ 1.02 \pm 0.00 \\ 1.02 \pm 0.01 \\ 1.02 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 1.00 \pm 0.02 \\ 1.06 \pm 0.02 \\ 1.06 \pm 0.01 \\ 1.02 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.96 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.95 \pm 0.02 \\ 0.98 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.94 \pm 0.01 \\ 0.96 \pm 0.00 \\ 0.98 \pm 0.01 \\ 0.98 \pm 0.00 \\ 0.99 \pm 0.00 \end{array}$	$\begin{array}{c} 0.64 \pm 0.10 \\ 0.95 \pm 0.01 \\ 1.03 \pm 0.01 \\ 0.98 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$ \begin{vmatrix} 1.01 \pm 0.15 \\ 1.09 \pm 0.02 \\ 1.11 \pm 0.00 \\ 1.03 \pm 0.04 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.97 \pm 0.02 \\ 1.01 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.87 \pm 0.00 \\ 0.89 \pm 0.01 \\ 0.98 \pm 0.02 \\ 1.00 \pm 0.00 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.93 \pm 0.12 \\ 0.99 \pm 0.05 \\ 1.02 \pm 0.04 \\ 1.01 \pm 0.03 \\ 1.00 \pm 0.01 \end{array}$	1.00/0.98 0.99/0.95 0.97/0.89 0.88/0.72 0.50/0.18
100 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.80 \pm 0.07 \\ 0.95 \pm 0.06 \\ 1.01 \pm 0.03 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.89 \pm 0.06 \\ 0.98 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.02 \pm 0.01 \\ 1.06 \pm 0.00 \end{array}$	$\begin{array}{c} 0.93 \pm 0.03 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.96 \pm 0.01 \\ 0.91 \pm 0.06 \\ 0.98 \pm 0.02 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.50 \pm 0.09 \\ 0.80 \pm 0.14 \\ 0.96 \pm 0.01 \\ 0.99 \pm 0.01 \\ 0.99 \pm 0.00 \end{array}$	$ \begin{array}{c} 0.78 \pm 0.01 \\ 0.89 \pm 0.06 \\ 0.94 \pm 0.01 \\ 0.97 \pm 0.00 \\ 0.98 \pm 0.00 \end{array} $	$\begin{array}{c} 0.28 \pm 0.07 \\ 0.60 \pm 0.10 \\ 0.88 \pm 0.05 \\ 1.00 \pm 0.03 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.52 \pm 0.10 \\ 0.77 \pm 0.26 \\ 1.11 \pm 0.08 \\ 1.11 \pm 0.02 \\ 1.11 \pm 0.03 \end{array}$	$\begin{array}{c} 0.78 \pm 0.03 \\ 0.91 \pm 0.02 \\ 0.96 \pm 0.02 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.01 \end{array}$	$ \begin{array}{c} 0.81 \pm 0.02 \\ 0.83 \pm 0.02 \\ 0.87 \pm 0.03 \\ 0.89 \pm 0.02 \\ 0.98 \pm 0.01 \end{array} $	$\begin{array}{c} 0.72 \pm 0.22 \\ 0.86 \pm 0.14 \\ 0.97 \pm 0.07 \\ 1.00 \pm 0.05 \\ 1.01 \pm 0.04 \end{array}$	1.00 / 0.99 1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84
100 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.95 \pm 0.01 \\ 1.00 \pm 0.02 \\ 1.02 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$\begin{array}{c} 0.97 \pm 0.03 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.02 \\ 1.05 \pm 0.01 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 0.97 \pm 0.03 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.97 \pm 0.03 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.93 \pm 0.01 \\ 0.97 \pm 0.01 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.92 \pm 0.01 \\ 0.95 \pm 0.00 \\ 0.96 \pm 0.00 \\ 0.98 \pm 0.00 \\ 0.99 \pm 0.00 \end{array}$	$\begin{array}{c} 0.64 \pm 0.03 \\ 0.86 \pm 0.03 \\ 0.99 \pm 0.01 \\ 1.03 \pm 0.01 \\ 0.98 \pm 0.00 \end{array}$	$\begin{array}{c} 0.89 \pm 0.16 \\ 1.12 \pm 0.03 \\ 1.09 \pm 0.01 \\ 1.10 \pm 0.01 \\ 1.00 \pm 0.03 \end{array}$	$\begin{array}{c} 0.87 \pm 0.02 \\ 0.96 \pm 0.01 \\ 0.99 \pm 0.02 \\ 1.01 \pm 0.00 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.83 \pm 0.01 \\ 0.87 \pm 0.00 \\ 0.89 \pm 0.00 \\ 0.99 \pm 0.01 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.89 \pm 0.11 \\ 0.97 \pm 0.07 \\ 0.99 \pm 0.05 \\ 1.02 \pm 0.04 \\ 1.00 \pm 0.01 \end{array}$	1.00/0.99 0.99/0.97 0.97/0.93 0.88/0.80 0.50/0.34
100 w/clusters (C)	16 C 5 16 C 7	1.13 ± 0.01 1.12 ± 0.01	1.03 ± 0.01 1.01 ± 0.01	1.00 ± 0.00 1.00 ± 0.00	1.00 ± 0.00 1.00 ± 0.00	0.99 ± 0.01 0.99 ± 0.01	0.96 ± 0.00 0.96 ± 0.01	1.01 ± 0.02 1.02 ± 0.02	1.23 ± 0.02 1.24 ± 0.05	1.05 ± 0.01 1.03 ± 0.01	0.99 ± 0.06 0.99 ± 0.05	1.04 ± 0.08 1.04 ± 0.08	1.00/0.95 1.00/0.93
500 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.57 \pm 0.07 \\ 0.61 \pm 0.12 \\ 0.73 \pm 0.02 \\ 0.84 \pm 0.00 \\ 0.99 \pm 0.03 \end{array}$	$ \begin{vmatrix} 0.55 \pm 0.03 \\ 0.55 \pm 0.08 \\ 0.63 \pm 0.11 \\ 0.92 \pm 0.02 \\ 0.99 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.83 \pm 0.04 \\ 0.83 \pm 0.02 \\ 0.89 \pm 0.04 \\ 0.97 \pm 0.03 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.78 \pm 0.16 \\ 0.84 \pm 0.12 \\ 0.97 \pm 0.00 \\ 0.98 \pm 0.01 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.85 \pm 0.04 \\ 0.91 \pm 0.02 \\ 0.94 \pm 0.00 \\ 0.94 \pm 0.00 \\ 0.96 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.68 \pm 0.07 \\ 0.71 \pm 0.05 \\ 0.83 \pm 0.05 \\ 0.88 \pm 0.02 \\ 0.92 \pm 0.03 \end{vmatrix} $	$\begin{array}{c} 0.24 \pm 0.01 \\ 0.29 \pm 0.05 \\ 0.45 \pm 0.09 \\ 0.60 \pm 0.15 \\ 0.66 \pm 0.06 \end{array}$	$ \begin{vmatrix} 0.43 \pm 0.01 \\ 0.47 \pm 0.08 \\ 0.50 \pm 0.07 \\ 0.53 \pm 0.01 \\ 0.84 \pm 0.14 \end{vmatrix} $	$\begin{array}{c} 0.76 \pm 0.06 \\ 0.79 \pm 0.04 \\ 0.82 \pm 0.02 \\ 0.85 \pm 0.05 \\ 0.92 \pm 0.02 \end{array}$	$ \begin{vmatrix} 0.79 \pm 0.01 \\ 0.79 \pm 0.01 \\ 0.80 \pm 0.02 \\ 0.80 \pm 0.02 \\ 0.84 \pm 0.01 \end{vmatrix} $	$ \begin{array}{c} 0.65 \pm 0.20 \\ 0.68 \pm 0.20 \\ 0.76 \pm 0.18 \\ 0.83 \pm 0.15 \\ 0.91 \pm 0.11 \end{array} $	1.00 / 1.00 1.00 / 1.00 1.00 / 0.99 1.00 / 0.98 1.00 / 0.97
500 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.57 \pm 0.01 \\ 0.79 \pm 0.05 \\ 1.02 \pm 0.00 \\ 1.03 \pm 0.01 \\ 1.03 \pm 0.00 \end{array}$	$\begin{array}{c} 0.43 \pm 0.07 \\ 0.54 \pm 0.04 \\ 0.96 \pm 0.01 \\ 0.97 \pm 0.02 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 0.78 \pm 0.01 \\ 0.93 \pm 0.02 \\ 0.94 \pm 0.01 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.97 \pm 0.00 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.96 \pm 0.00 \\ 0.97 \pm 0.00 \\ 0.96 \pm 0.00 \\ 0.98 \pm 0.00 \\ 0.99 \pm 0.01 \end{array}$	$\begin{array}{c} 0.83 \pm 0.01 \\ 0.90 \pm 0.01 \\ 0.97 \pm 0.01 \\ 0.96 \pm 0.00 \\ 0.97 \pm 0.01 \end{array}$	$\begin{array}{c} 0.64 \pm 0.00 \\ 0.69 \pm 0.01 \\ 0.73 \pm 0.01 \\ 0.87 \pm 0.01 \\ 0.99 \pm 0.02 \end{array}$	$\begin{array}{c} 0.53 \pm 0.03 \\ 0.50 \pm 0.00 \\ 0.54 \pm 0.01 \\ 1.07 \pm 0.02 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 0.83 \pm 0.01 \\ 0.86 \pm 0.02 \\ 0.91 \pm 0.01 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.83 \pm 0.00 \\ 0.83 \pm 0.01 \\ 0.86 \pm 0.00 \\ 0.87 \pm 0.00 \\ 0.87 \pm 0.00 \end{array}$	$\begin{array}{c} 0.75 \pm 0.17 \\ 0.81 \pm 0.16 \\ 0.89 \pm 0.14 \\ 0.97 \pm 0.06 \\ 0.99 \pm 0.05 \end{array}$	1.00/1.00 0.99/0.99 0.97/0.96 0.88/0.86 0.50/0.47
500 w/clusters (C)	16 C 7 16 C 10 16 C 25 64 C 5 64 C 7	1.09 1.10 1.10 1.09 1.12	1.00 1.01 1.00 0.98 1.02	0.99 1.00 1.00 1.00 1.00	1.00 0.99 0.99 1.00 1.00	0.98 0.97 0.99 0.99 1.00	0.95 0.93 0.96 0.96 0.96	0.72 0.70 0.98 0.99 0.99	0.87 1.30 1.31 1.18 1.22	0.98 1.02 1.03 1.04 1.04	0.90 0.88 0.91 0.87 0.93	0.95 0.99 1.03 1.01 1.03	1.00/0.98 1.00/0.98 1.00/0.95 0.97/0.93 0.97/0.91
1000 w/clusters (C)	16 C 25	1.09	0.98	1.00	1.00	0.97	0.96	0.72	1.30	1.05	0.91	1.00	1.00/0.97

Table 7: Relative In-Distribution ROUGE-L scores for various tasks and methods

Model Type	Method Type					Tas	sks					Average	Para. Saved
model Type		task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	Tiverage	
	base lora	$\begin{array}{c} 24.44 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c }\hline 1.60 \pm 0.00 \\ 86.00 \pm 0.00 \\ \end{array}$	$\begin{array}{c} 19.13 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 39.22 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 10.27 \pm 0.00 \\ 94.33 \pm 0.00 \end{array}$	$\begin{array}{c} 35.46 \pm 0.00 \\ 74.88 \pm 0.00 \end{array}$	$ \begin{array}{c c} 7.85 \pm 0.00 \\ 74.40 \pm 0.00 \end{array} $	$\begin{array}{c} 6.22 \pm 0.00 \\ 26.68 \pm 0.00 \end{array}$	$\begin{array}{c c} 17.82 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 38.87 \pm 0.00 \\ 50.32 \pm 0.00 \end{array}$	$\begin{array}{c} 20.24 \pm 13.27 \\ 78.87 \pm 22.56 \end{array}$	1.00/1.00 0.00/0.00
TIES	10 50 100 500	$76.50 \pm 0.00 \\ 55.80 \pm 0.00 \\ 52.43 \pm 0.00 \\ 35.18 \pm 0.00$	$\begin{array}{c} 49.00 \pm 1.73 \\ 35.00 \pm 0.00 \\ 34.00 \pm 0.00 \\ 22.00 \pm 0.00 \end{array}$	$44.33 \pm 4.04 \\ 18.00 \pm 5.20 \\ 19.67 \pm 4.62 \\ 1.00 \pm 0.00$	$\begin{array}{c} 9.80 \pm 0.58 \\ 2.42 \pm 0.50 \\ 1.09 \pm 1.66 \\ 0.00 \pm 0.00 \end{array}$	$78.56 \pm 0.96 \\ 85.78 \pm 0.96 \\ 83.33 \pm 0.00 \\ 78.00 \pm 0.00$	$\begin{array}{c} 35.24 \pm 0.00 \\ 23.03 \pm 0.00 \\ 24.89 \pm 0.00 \\ 21.46 \pm 0.00 \end{array}$	$\begin{array}{c} 51.37 \pm 0.67 \\ 48.03 \pm 0.00 \\ 47.52 \pm 0.00 \\ 42.22 \pm 0.04 \end{array}$	$\begin{array}{c} 15.26 \pm 0.12 \\ 16.50 \pm 0.00 \\ 15.18 \pm 0.42 \\ 9.93 \pm 0.13 \end{array}$	$77.67 \pm 1.15 \\ 30.00 \pm 3.46 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00$	$42.72 \pm 0.01 \\ 42.47 \pm 0.02 \\ 41.19 \pm 0.03 \\ 21.50 \pm 0.03$	$\begin{array}{c} 48.05 \pm 23.61 \\ 35.70 \pm 23.01 \\ 32.03 \pm 24.50 \\ 23.27 \pm 23.64 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 1.00 1.00 / 1.00
SVD	SVD 2 SVD 4 SVD 8 SVD 16	$\begin{array}{c} 93.15 \pm 2.77 \\ 94.01 \pm 3.60 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 92.24 \pm 1.85 \\ 89.21 \pm 0.71 \\ 87.40 \pm 0.59 \\ 86.00 \pm 0.00 \end{array}$	$\begin{array}{c} 99.09 \pm 0.18 \\ 99.05 \pm 0.09 \\ 99.05 \pm 0.09 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 93.44 \pm 0.14 \\ 93.65 \pm 0.03 \\ 93.65 \pm 0.03 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 93.89 \pm 0.35 \\ 94.66 \pm 0.63 \\ 94.36 \pm 0.38 \\ 94.33 \pm 0.00 \end{array}$	$\begin{array}{c} 73.74 \pm 0.51 \\ 74.89 \pm 0.33 \\ 74.58 \pm 0.12 \\ 74.90 \pm 0.03 \end{array}$	$\begin{array}{c} 74.55 \pm 0.98 \\ 73.61 \pm 1.15 \\ 75.07 \pm 0.00 \\ 74.23 \pm 0.18 \end{array}$	$\begin{array}{c} 26.80 \pm 2.79 \\ 26.34 \pm 2.13 \\ 26.71 \pm 0.27 \\ 26.68 \pm 0.00 \end{array}$	$\begin{array}{c} 95.06 \pm 1.35 \\ 93.98 \pm 0.77 \\ 95.51 \pm 1.09 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 50.21 \pm 0.44 \\ 50.47 \pm 0.54 \\ 50.89 \pm 0.07 \\ 50.30 \pm 0.02 \end{array}$	$\begin{array}{c} 79.11 \pm 22.72 \\ 78.90 \pm 22.68 \\ 81.01 \pm 21.74 \\ 78.36 \pm 22.97 \end{array}$	0.88 / 0.88 0.75 / 0.75 0.50 / 0.50 0.00 / 0.00
10 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 96.67 \pm 0.58 \\ 95.67 \pm 0.58 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 87.00 \pm 1.00 \\ 90.00 \pm 1.00 \\ 88.33 \pm 0.58 \\ 86.67 \pm 0.58 \\ 86.00 \pm 0.00 \end{array}$	$\begin{array}{c} 99.00 \pm 0.00 \\ \end{array}$	$\begin{array}{c} 94.00 \pm 0.67 \\ 93.00 \pm 0.33 \\ 93.67 \pm 0.00 \\ 93.67 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 93.11 \pm 0.38 \\ 94.89 \pm 0.51 \\ 94.78 \pm 0.38 \\ 94.33 \pm 0.00 \\ 94.33 \pm 0.00 \end{array}$	$\begin{array}{c} 72.08 \pm 0.06 \\ 73.86 \pm 0.31 \\ 74.61 \pm 0.13 \\ 74.92 \pm 0.13 \\ 74.88 \pm 0.00 \end{array}$	$ 76.26 \pm 1.19 71.92 \pm 0.84 74.97 \pm 0.58 74.96 \pm 0.51 74.40 \pm 0.00 $	$\begin{array}{c} 30.11 \pm 0.79 \\ 27.89 \pm 0.70 \\ 26.35 \pm 0.25 \\ 26.45 \pm 0.23 \\ 26.68 \pm 0.00 \end{array}$	$\begin{array}{c} 94.00 \pm 1.73 \\ 94.67 \pm 0.58 \\ 96.00 \pm 0.00 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 49.30 \pm 0.46 \\ 50.36 \pm 0.26 \\ 50.99 \pm 0.06 \\ 50.21 \pm 0.12 \\ 50.27 \pm 0.02 \end{array}$	79.15 ± 22.18 79.13 ± 22.75 79.37 ± 22.94 79.02 ± 22.84 78.92 ± 22.77	1.00 / 0.90 1.00 / 0.80 1.00 / 0.60 1.00 / 0.20 1.00 / -0.60
10 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 97.00 \pm 0.00 \\ 96.67 \pm 0.58 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 91.00 \pm 1.00 \\ 89.33 \pm 0.58 \\ 88.67 \pm 0.58 \\ 86.67 \pm 0.58 \\ 86.00 \pm 0.00 \end{array}$	$\begin{array}{c} 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 93.56 \pm 0.19 \\ 93.22 \pm 0.19 \\ 93.67 \pm 0.00 \\ 93.67 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 93.56 \pm 0.69 \\ 94.44 \pm 0.19 \\ 94.56 \pm 0.38 \\ 94.33 \pm 0.00 \\ 94.33 \pm 0.00 \end{array}$	$\begin{array}{c} 73.60 \pm 0.36 \\ 74.11 \pm 0.19 \\ 74.56 \pm 0.13 \\ 75.04 \pm 0.03 \\ 74.90 \pm 0.03 \end{array}$	$74.94 \pm 1.25 71.74 \pm 0.59 75.47 \pm 0.58 74.40 \pm 0.00 74.29 \pm 0.19$	$\begin{array}{c} 28.66 \pm 0.03 \\ 26.74 \pm 0.50 \\ 26.26 \pm 0.34 \\ 26.53 \pm 0.13 \\ 26.68 \pm 0.00 \end{array}$	$\begin{array}{c} 96.00 \pm 1.00 \\ 94.67 \pm 0.58 \\ 96.00 \pm 0.00 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 50.15 \pm 0.20 \\ 50.63 \pm 0.24 \\ 50.89 \pm 0.17 \\ 50.36 \pm 0.03 \\ 50.30 \pm 0.03 \end{array}$	$\begin{array}{c} 79.75 \pm 22.72 \\ 79.06 \pm 23.01 \\ 79.41 \pm 22.97 \\ 79.00 \pm 22.81 \\ 78.92 \pm 22.77 \end{array}$	1.00/0.90 0.99/0.79 0.97/0.57 0.88/0.07 0.50/-1.10
50 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 92.76 \pm 3.53 \\ 95.33 \pm 2.08 \\ 97.00 \pm 0.00 \\ 96.33 \pm 0.58 \\ 95.67 \pm 0.58 \end{array}$	$\begin{array}{c} 84.67 \pm 1.15 \\ 87.33 \pm 2.08 \\ 90.33 \pm 1.53 \\ 92.67 \pm 0.58 \\ 88.33 \pm 0.58 \end{array}$	$\begin{array}{c} 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$86.17 \pm 5.81 \\ 92.60 \pm 0.29 \\ 93.78 \pm 0.19 \\ 93.56 \pm 0.19 \\ 93.56 \pm 0.19$	$\begin{array}{c} 79.68 \pm 6.21 \\ 90.32 \pm 1.04 \\ 93.00 \pm 0.58 \\ 93.00 \pm 0.58 \\ 94.67 \pm 0.67 \end{array}$	$ 69.07 \pm 1.54 \\ 71.16 \pm 1.47 \\ 72.37 \pm 0.35 \\ 73.32 \pm 0.24 \\ 74.82 \pm 0.24 $	$\begin{array}{c} 50.65 \pm 3.97 \\ 62.51 \pm 1.64 \\ 73.39 \pm 0.93 \\ 73.03 \pm 1.09 \\ 72.36 \pm 2.07 \end{array}$	$\begin{array}{c} 23.27 \pm 2.60 \\ 26.60 \pm 3.54 \\ 29.06 \pm 0.80 \\ 29.51 \pm 0.93 \\ 26.90 \pm 0.75 \end{array}$	$\begin{array}{c} 83.90 \pm 6.43 \\ 93.33 \pm 1.15 \\ 95.67 \pm 0.58 \\ 95.00 \pm 0.00 \\ 95.33 \pm 0.58 \end{array}$	$\begin{array}{c} 41.86 \pm 0.96 \\ 44.35 \pm 0.41 \\ 45.43 \pm 0.34 \\ 50.16 \pm 0.74 \\ 50.73 \pm 0.46 \end{array}$	$71.10 \pm 23.99 \\ 76.25 \pm 23.81 \\ 78.90 \pm 23.29 \\ 79.56 \pm 22.51 \\ 79.14 \pm 22.90$	1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84 1.00 / 0.68
50 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 94.06 \pm 3.54 \\ 97.00 \pm 0.00 \\ 96.67 \pm 0.58 \\ 97.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 85.67 \pm 1.15 \\ 85.67 \pm 1.53 \\ 91.00 \pm 2.00 \\ 91.00 \pm 1.00 \\ 88.00 \pm 0.00 \end{array}$	$\begin{array}{c} 98.67 \pm 0.58 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 90.35 \pm 1.37 \\ 93.67 \pm 0.00 \\ 93.56 \pm 0.19 \\ 93.33 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 89.90 \pm 1.91 \\ 92.22 \pm 0.69 \\ 93.22 \pm 0.51 \\ 94.11 \pm 0.51 \\ 94.44 \pm 0.19 \end{array}$	$\begin{array}{c} 70.32 \pm 0.66 \\ 71.88 \pm 0.30 \\ 73.16 \pm 0.41 \\ 73.51 \pm 0.23 \\ 74.25 \pm 0.21 \end{array}$	$\begin{array}{c} 47.62 \pm 7.28 \\ 71.01 \pm 1.02 \\ 76.28 \pm 0.51 \\ 73.17 \pm 0.58 \\ 74.97 \pm 0.58 \end{array}$	$\begin{array}{c} 26.88 \pm 3.96 \\ 29.07 \pm 0.65 \\ 29.67 \pm 0.12 \\ 27.53 \pm 1.12 \\ 26.79 \pm 0.09 \end{array}$	$\begin{array}{c} 92.33 \pm 1.53 \\ 95.67 \pm 1.53 \\ 95.33 \pm 0.58 \\ 95.00 \pm 1.00 \\ 96.00 \pm 0.00 \end{array}$	$\begin{array}{c} 43.68 \pm 0.24 \\ 44.97 \pm 0.41 \\ 49.31 \pm 1.00 \\ 50.56 \pm 0.06 \\ 50.86 \pm 0.19 \end{array}$	$\begin{array}{c} 73.95 \pm 24.73 \\ 78.02 \pm 23.18 \\ 79.72 \pm 22.50 \\ 79.42 \pm 22.93 \\ 79.30 \pm 22.82 \end{array}$	1.00/0.98 0.99/0.95 0.97/0.89 0.88/0.72 0.50/0.18
100 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$76.43 \pm 7.07 \\ 90.10 \pm 5.85 \\ 95.56 \pm 2.49 \\ 96.00 \pm 0.00 \\ 95.00 \pm 0.00$	$ \begin{array}{c} 76.67 \pm 4.93 \\ 84.00 \pm 1.00 \\ 86.67 \pm 0.58 \\ 87.33 \pm 1.15 \\ 91.00 \pm 0.00 \\ \end{array} $	$\begin{array}{c} 91.61 \pm 2.75 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ \end{array}$	89.99 ± 1.07 85.52 ± 5.34 92.24 ± 1.68 93.89 ± 0.19 93.56 ± 0.19	$47.55 \pm 8.56 \\ 75.69 \pm 12.75 \\ 90.89 \pm 1.17 \\ 93.00 \pm 0.58 \\ 93.11 \pm 0.19$	$\begin{array}{c} 58.08 \pm 0.72 \\ 66.62 \pm 4.18 \\ 70.35 \pm 0.45 \\ 72.70 \pm 0.30 \\ 73.05 \pm 0.20 \end{array}$	20.77 ± 5.50 44.66 ± 7.26 65.62 ± 4.03 74.34 ± 2.07 74.52 ± 0.95	$\begin{array}{c} 13.90 \pm 2.79 \\ 20.49 \pm 7.07 \\ 29.58 \pm 2.02 \\ 29.66 \pm 0.54 \\ 29.67 \pm 0.67 \end{array}$	$73.93 \pm 3.13 86.67 \pm 1.86 91.67 \pm 2.31 93.67 \pm 0.58 95.33 \pm 0.58$	40.74 ± 0.85 42.01 ± 0.94 43.64 ± 1.36 44.82 ± 0.89 49.42 ± 0.65	$\begin{array}{c} 58.97 \pm 26.83 \\ 69.48 \pm 25.14 \\ 76.52 \pm 23.02 \\ 78.44 \pm 22.87 \\ 79.37 \pm 22.38 \end{array}$	1.00 / 0.99 1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84
100 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 90.70 \pm 1.07 \\ 95.33 \pm 1.53 \\ 97.00 \pm 0.00 \\ 96.33 \pm 0.58 \\ 96.33 \pm 0.58 \end{array}$	83.00 ± 2.65 85.00 ± 1.00 85.67 ± 1.53 90.33 ± 0.58 88.67 ± 0.58	$\begin{array}{c} 96.00 \pm 3.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ \end{array}$	$\begin{array}{c} 91.22 \pm 2.94 \\ 93.50 \pm 0.22 \\ 93.78 \pm 0.19 \\ 93.00 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 87.94 \pm 0.54 \\ 91.44 \pm 0.84 \\ 92.56 \pm 0.19 \\ 93.89 \pm 0.19 \\ 94.89 \pm 0.19 \end{array}$		$ \begin{array}{c} 47.57 \pm 2.54 \\ 63.64 \pm 1.98 \\ 73.29 \pm 0.64 \\ 76.50 \pm 1.01 \\ 72.90 \pm 0.12 \end{array} $	$\begin{array}{c} 23.75 \pm 4.33 \\ 29.82 \pm 0.81 \\ 29.15 \pm 0.24 \\ 29.45 \pm 0.35 \\ 26.77 \pm 0.68 \end{array}$	$\begin{array}{c} 82.33 \pm 2.08 \\ 91.67 \pm 0.58 \\ 94.33 \pm 1.53 \\ 96.00 \pm 0.00 \\ 96.00 \pm 0.00 \end{array}$	$\begin{array}{c} 41.51 \pm 0.67 \\ 43.94 \pm 0.18 \\ 44.97 \pm 0.05 \\ 49.81 \pm 0.34 \\ 50.83 \pm 0.09 \end{array}$	$71.27 \pm 24.23 \\ 76.43 \pm 23.01 \\ 78.18 \pm 23.03 \\ 79.74 \pm 22.47 \\ 79.35 \pm 23.04$	1.00/0.99 0.99/0.97 0.97/0.93 0.88/0.80 0.50/0.34
100 w/clusters (C)	16 C 5 16 C 7	98.33 ± 0.47 97.67 ± 0.47	89.00 ± 0.82 87.00 ± 0.82	99.00 ± 0.00 99.00 ± 0.00	93.25 ± 0.40 93.46 ± 0.29	92.89 ± 0.87 93.11 ± 0.68	72.32 ± 0.36 72.52 ± 0.43	77.08 ± 1.67 77.66 ± 1.30	28.26 ± 0.38 28.51 ± 1.26	96.67 ± 0.47 95.33 ± 0.47	68.30 ± 15.72 68.46 ± 14.94	81.51 ± 20.63 81.27 ± 20.35	1.00/0.95 1.00/0.93
500 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 54.44 \pm 6.87 \\ 58.08 \pm 11.52 \\ 69.21 \pm 2.03 \\ 79.77 \pm 0.37 \\ 93.83 \pm 2.52 \end{array}$	$ \begin{vmatrix} 47.00 \pm 2.83 \\ 47.00 \pm 7.07 \\ 54.50 \pm 9.19 \\ 79.50 \pm 2.12 \\ 85.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 82.21 \pm 3.59 \\ 82.06 \pm 1.69 \\ 88.33 \pm 4.04 \\ 95.89 \pm 2.83 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 73.38 \pm 14.97 \\ 78.62 \pm 11.23 \\ 91.11 \pm 0.38 \\ 91.89 \pm 1.39 \\ 93.78 \pm 0.19 \end{array}$	$80.08 \pm 3.71 \\ 85.57 \pm 1.48 \\ 88.78 \pm 0.38 \\ 88.67 \pm 0.00 \\ 90.56 \pm 0.38$	51.02 ± 5.31 52.98 ± 3.81 62.36 ± 3.52 65.92 ± 1.79 68.95 ± 1.92			72.67 ± 6.03 75.33 ± 4.04 77.67 ± 2.31 81.00 ± 5.00 87.33 ± 2.31	$\begin{array}{c} 39.65 \pm 0.28 \\ 39.78 \pm 0.42 \\ 40.42 \pm 0.98 \\ 40.34 \pm 0.80 \\ 42.15 \pm 0.73 \end{array}$	$\begin{array}{c} 53.16 \pm 24.97 \\ 55.66 \pm 25.48 \\ 62.16 \pm 26.05 \\ 67.82 \pm 26.35 \\ 72.83 \pm 25.93 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 0.99 1.00 / 0.98 1.00 / 0.97
500 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 54.30 \pm 1.13 \\ 75.10 \pm 4.92 \\ 96.94 \pm 0.42 \\ 97.67 \pm 0.58 \\ 98.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 37.00 \pm 5.66 \\ 46.50 \pm 3.54 \\ 82.50 \pm 0.71 \\ 83.50 \pm 2.12 \\ 88.50 \pm 0.71 \end{vmatrix} $	$77.67 \pm 0.58 \\ 91.67 \pm 1.53 \\ 93.33 \pm 0.58 \\ 98.00 \pm 0.00 \\ 99.00 \pm 0.00$	$\begin{array}{c} 91.00 \pm 0.00 \\ 91.56 \pm 0.19 \\ 93.89 \pm 0.69 \\ 93.56 \pm 0.19 \\ 93.78 \pm 0.19 \end{array}$	$\begin{array}{c} 90.56 \pm 0.19 \\ 91.56 \pm 0.38 \\ 90.67 \pm 0.00 \\ 92.00 \pm 0.00 \\ 93.00 \pm 0.88 \end{array}$		47.56 ± 0.29 51.17 ± 0.81 54.63 ± 0.79 65.02 ± 0.81 73.77 ± 1.21	$14.18 \pm 0.67 \\ 13.44 \pm 0.02 \\ 14.49 \pm 0.27 \\ 28.49 \pm 0.55 \\ 27.59 \pm 0.39$	$79.00 \pm 1.00 \\ 81.67 \pm 1.53 \\ 86.33 \pm 0.58 \\ 93.00 \pm 0.00 \\ 95.33 \pm 0.58$	$\begin{array}{c} 41.58 \pm 0.23 \\ 41.92 \pm 0.42 \\ 43.16 \pm 0.08 \\ 43.85 \pm 0.12 \\ 43.81 \pm 0.17 \end{array}$	$60.31 \pm 24.42 \\ 65.84 \pm 25.64 \\ 72.49 \pm 26.64 \\ 76.47 \pm 23.77 \\ 78.18 \pm 24.16$	1.00/1.00 0.99/0.99 0.97/0.96 0.88/0.86 0.50/0.47
500 w/clusters (C)	16 C 7 16 C 10 16 C 25 64 C 5 64 C 7	95.00 96.00 96.00 95.00 98.00	86.00 87.00 86.00 84.00 88.00	98.00 99.00 99.00 99.00 99.00	93.67 93.00 92.71 93.67 94.00	91.67 91.33 93.00 92.67 93.33	71.19 69.93 72.13 72.32 72.18	54.69 53.48 74.59 75.60 75.83	20.03 30.09 30.21 27.17 28.14	90.00 94.00 95.00 96.00 96.00	46.34 44.89 46.66 44.43 47.68	74.66 75.87 78.53 77.99 79.22	1.00/0.98 1.00/0.98 1.00/0.95 0.97/0.93 0.97/0.91
1000 w/clusters (C)	16 C 25	95.00	84.00	99.00	93.67	90.67	72.20	55.04	29.97	97.00	46.86	76.34	1.00/0.97

Table 8: Absolute In-Distribution ROUGE-L scores for various tasks and methods

Model Type	Method Type					Ta	sks					. Average	Para. Saved
Woder Type	Wiemod Type	task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	Average	raia. Saveu
	base lora	$\begin{array}{c} 0.26 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.02 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.19 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.42 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c c} 0.11 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.19 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.29 \pm 0.22 \\ 1.00 \pm 0.00 \end{array}$	1.00/1.00 0.00/0.00
TIES	10 50 100 500	$\begin{array}{c} 0.81 \pm 0.00 \\ 0.59 \pm 0.00 \\ 0.55 \pm 0.00 \\ 0.37 \pm 0.00 \end{array}$	$\begin{array}{c} 0.57 \pm 0.02 \\ 0.41 \pm 0.00 \\ 0.40 \pm 0.00 \\ 0.26 \pm 0.00 \end{array}$	$\begin{array}{c} 0.45 \pm 0.04 \\ 0.18 \pm 0.05 \\ 0.20 \pm 0.05 \\ 0.01 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.10 \pm 0.01 \\ 0.03 \pm 0.01 \\ 0.01 \pm 0.02 \\ 0.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.83 \pm 0.01 \\ 0.91 \pm 0.01 \\ 0.88 \pm 0.00 \\ 0.83 \pm 0.00 \end{array}$	$\begin{array}{c} 0.52 \pm 0.00 \\ 0.34 \pm 0.00 \\ 0.36 \pm 0.00 \\ 0.31 \pm 0.00 \end{array}$	$\begin{array}{c} 0.71 \pm 0.01 \\ 0.67 \pm 0.00 \\ 0.65 \pm 0.00 \\ 0.58 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.58 \pm 0.00 \\ 0.62 \pm 0.00 \\ 0.57 \pm 0.02 \\ 0.37 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.82 \pm 0.01 \\ 0.32 \pm 0.04 \\ 0.01 \pm 0.00 \\ 0.01 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.80 \pm 0.00 \\ 0.78 \pm 0.00 \\ 0.78 \pm 0.00 \\ 0.41 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.62 \pm 0.22 \\ 0.48 \pm 0.27 \\ 0.44 \pm 0.29 \\ 0.32 \pm 0.26 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 1.00 1.00 / 1.00
SVD	SVD 2 SVD 4 SVD 8 SVD 16	$\begin{array}{c} 0.98 \pm 0.03 \\ 0.99 \pm 0.04 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.07 \pm 0.02 \\ 1.04 \pm 0.01 \\ 1.02 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.10 \\ 0.99 \pm 0.08 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.99 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.04 \\ 1.00 \pm 0.03 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	0.88 / 0.88 0.75 / 0.75 0.50 / 0.50 0.00 / 0.00
10 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 1.02 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.01 \\ 1.05 \pm 0.01 \\ 1.03 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.99 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.97 \pm 0.00 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.03 \pm 0.02 \\ 0.97 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.12 \pm 0.03 \\ 1.04 \pm 0.03 \\ 0.99 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.99 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.99 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.04 \\ 1.01 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	1.00 / 0.90 1.00 / 0.80 1.00 / 0.60 1.00 / 0.20 1.00 / -0.60
10 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 1.02 \pm 0.00 \\ 1.02 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.06 \pm 0.01 \\ 1.04 \pm 0.01 \\ 1.03 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.99 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.99 \pm 0.00 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.02 \\ 0.96 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.07 \pm 0.00 \\ 1.00 \pm 0.02 \\ 0.98 \pm 0.01 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.01 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.01 \pm 0.00 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.02 \pm 0.03 \\ 1.00 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	1.00/0.90 0.99/0.79 0.97/0.57 0.88/0.07 0.50/-1.10
50 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.98 \pm 0.04 \\ 1.00 \pm 0.02 \\ 1.02 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$\begin{array}{c} 0.98 \pm 0.01 \\ 1.02 \pm 0.02 \\ 1.05 \pm 0.02 \\ 1.08 \pm 0.01 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.92 \pm 0.06 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.85 \pm 0.06 \\ 0.96 \pm 0.01 \\ 0.99 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.94 \pm 0.02 \\ 0.96 \pm 0.02 \\ 0.97 \pm 0.01 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.69 \pm 0.05 \\ 0.85 \pm 0.02 \\ 0.99 \pm 0.01 \\ 0.98 \pm 0.02 \\ 0.97 \pm 0.03 \end{array}$	$\begin{array}{c} 0.88 \pm 0.10 \\ 1.00 \pm 0.12 \\ 1.09 \pm 0.03 \\ 1.10 \pm 0.03 \\ 1.00 \pm 0.03 \end{array}$	$\begin{array}{c} 0.88 \pm 0.07 \\ 0.98 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.01 \end{array}$	$ \begin{vmatrix} 0.86 \pm 0.01 \\ 0.90 \pm 0.00 \\ 0.94 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.90 \pm 0.10 \\ 0.97 \pm 0.06 \\ 1.01 \pm 0.04 \\ 1.02 \pm 0.04 \\ 1.00 \pm 0.02 \end{array}$	1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84 1.00 / 0.68
50 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.99 \pm 0.04 \\ 1.02 \pm 0.00 \\ 1.02 \pm 0.01 \\ 1.02 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 1.00 \pm 0.02 \\ 1.06 \pm 0.02 \\ 1.06 \pm 0.01 \\ 1.02 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.96 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ \end{vmatrix} $	$\begin{array}{c} 0.95 \pm 0.02 \\ 0.98 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.95 \pm 0.01 \\ 0.97 \pm 0.00 \\ 0.98 \pm 0.00 \\ 0.98 \pm 0.00 \\ 0.99 \pm 0.00 \end{array}$	$\begin{array}{c} 0.65 \pm 0.09 \\ 0.96 \pm 0.01 \\ 1.03 \pm 0.01 \\ 0.98 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$ \begin{vmatrix} 1.01 \pm 0.15 \\ 1.09 \pm 0.03 \\ 1.11 \pm 0.00 \\ 1.03 \pm 0.04 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.97 \pm 0.02 \\ 1.01 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.88 \pm 0.01 \\ 0.93 \pm 0.00 \\ 0.99 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.94 \pm 0.11 \\ 0.99 \pm 0.04 \\ 1.02 \pm 0.04 \\ 1.01 \pm 0.02 \\ 1.00 \pm 0.01 \end{array}$	1.00/0.98 0.99/0.95 0.97/0.89 0.88/0.72 0.50/0.18
100 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.80 \pm 0.07 \\ 0.95 \pm 0.06 \\ 1.01 \pm 0.03 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.89 \pm 0.06 \\ 0.98 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.02 \pm 0.01 \\ 1.06 \pm 0.00 \end{array}$	$\begin{array}{c} 0.93 \pm 0.03 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{array}{c c} 0.96 \pm 0.01 \\ 0.91 \pm 0.06 \\ 0.98 \pm 0.02 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array} $	$\begin{array}{c} 0.51 \pm 0.09 \\ 0.80 \pm 0.13 \\ 0.96 \pm 0.01 \\ 0.99 \pm 0.01 \\ 0.99 \pm 0.00 \end{array}$	$\begin{array}{c} 0.81 \pm 0.02 \\ 0.91 \pm 0.05 \\ 0.95 \pm 0.01 \\ 0.98 \pm 0.00 \\ 0.98 \pm 0.00 \end{array}$	$\begin{array}{c} 0.30 \pm 0.07 \\ 0.62 \pm 0.10 \\ 0.90 \pm 0.05 \\ 1.00 \pm 0.03 \\ 1.00 \pm 0.01 \end{array}$	$ \begin{array}{c} 0.54 \pm 0.11 \\ 0.78 \pm 0.25 \\ 1.11 \pm 0.07 \\ 1.11 \pm 0.02 \\ 1.11 \pm 0.03 \end{array} $	$\begin{array}{c} 0.78 \pm 0.03 \\ 0.91 \pm 0.02 \\ 0.96 \pm 0.02 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.83 \pm 0.02 \\ 0.85 \pm 0.01 \\ 0.88 \pm 0.02 \\ 0.92 \pm 0.00 \\ 0.99 \pm 0.02 \end{array}$	$\begin{array}{c} 0.73 \pm 0.21 \\ 0.87 \pm 0.14 \\ 0.98 \pm 0.07 \\ 1.00 \pm 0.05 \\ 1.01 \pm 0.04 \end{array}$	1.00 / 0.99 1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84
100 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.95 \pm 0.01 \\ 1.00 \pm 0.02 \\ 1.02 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$\begin{array}{c} 0.97 \pm 0.03 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.02 \\ 1.05 \pm 0.01 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 0.97 \pm 0.03 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{array}{c} 0.97 \pm 0.03 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \end{array} $	$\begin{array}{c} 0.93 \pm 0.01 \\ 0.97 \pm 0.01 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.93 \pm 0.01 \\ 0.96 \pm 0.00 \\ 0.97 \pm 0.00 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.66 \pm 0.03 \\ 0.87 \pm 0.03 \\ 0.99 \pm 0.01 \\ 1.03 \pm 0.01 \\ 0.98 \pm 0.00 \end{array}$	$\begin{array}{c} 0.90 \pm 0.16 \\ 1.12 \pm 0.03 \\ 1.10 \pm 0.01 \\ 1.10 \pm 0.01 \\ 1.00 \pm 0.03 \end{array}$	$\begin{array}{c} 0.87 \pm 0.02 \\ 0.96 \pm 0.01 \\ 0.99 \pm 0.02 \\ 1.01 \pm 0.00 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.85 \pm 0.01 \\ 0.89 \pm 0.00 \\ 0.93 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.90 \pm 0.10 \\ 0.98 \pm 0.07 \\ 1.00 \pm 0.04 \\ 1.02 \pm 0.03 \\ 1.00 \pm 0.01 \end{array}$	1.00/0.99 0.99/0.97 0.97/0.93 0.88/0.80 0.50/0.34
100 w/clusters (C)	16 C 5 16 C 7	1.13 ± 0.01 1.12 ± 0.01	1.03 ± 0.01 1.01 ± 0.01	1.00 ± 0.00 1.00 ± 0.00	1.00 ± 0.00 1.00 ± 0.00	0.99 ± 0.01 0.99 ± 0.01	0.97 ± 0.00 0.97 ± 0.00	1.01 ± 0.02 1.02 ± 0.02	1.22 ± 0.02 1.22 ± 0.05	1.05 ± 0.01 1.03 ± 0.01	1.00 ± 0.05 1.01 ± 0.03	1.04 ± 0.07 1.04 ± 0.07	1.00/0.95 1.00/0.93
500 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.57 \pm 0.07 \\ 0.61 \pm 0.12 \\ 0.73 \pm 0.02 \\ 0.84 \pm 0.00 \\ 0.99 \pm 0.03 \end{array}$	$\begin{array}{c} 0.55 \pm 0.03 \\ 0.55 \pm 0.08 \\ 0.63 \pm 0.11 \\ 0.92 \pm 0.02 \\ 0.99 \pm 0.00 \end{array}$	$\begin{array}{c} 0.83 \pm 0.04 \\ 0.83 \pm 0.02 \\ 0.89 \pm 0.04 \\ 0.97 \pm 0.03 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.78 \pm 0.16 \\ 0.84 \pm 0.12 \\ 0.97 \pm 0.00 \\ 0.98 \pm 0.01 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.85 \pm 0.04 \\ 0.91 \pm 0.02 \\ 0.94 \pm 0.00 \\ 0.94 \pm 0.00 \\ 0.96 \pm 0.00 \end{array}$	$\begin{array}{c} 0.73 \pm 0.07 \\ 0.75 \pm 0.05 \\ 0.86 \pm 0.03 \\ 0.90 \pm 0.02 \\ 0.93 \pm 0.02 \end{array}$	$\begin{array}{c} 0.24 \pm 0.02 \\ 0.30 \pm 0.05 \\ 0.46 \pm 0.09 \\ 0.62 \pm 0.14 \\ 0.68 \pm 0.05 \end{array}$	$ \begin{vmatrix} 0.45 \pm 0.01 \\ 0.49 \pm 0.07 \\ 0.51 \pm 0.07 \\ 0.54 \pm 0.01 \\ 0.85 \pm 0.14 \end{vmatrix} $	$\begin{array}{c} 0.76 \pm 0.06 \\ 0.79 \pm 0.04 \\ 0.82 \pm 0.02 \\ 0.85 \pm 0.05 \\ 0.92 \pm 0.02 \end{array}$	$ \begin{array}{c c} 0.81 \pm 0.00 \\ 0.82 \pm 0.01 \\ 0.83 \pm 0.01 \\ 0.83 \pm 0.01 \\ 0.85 \pm 0.00 \end{array} $	$\begin{array}{c} 0.66 \pm 0.20 \\ 0.69 \pm 0.20 \\ 0.77 \pm 0.18 \\ 0.84 \pm 0.15 \\ 0.92 \pm 0.11 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 0.99 1.00 / 0.98 1.00 / 0.97
500 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.57 \pm 0.01 \\ 0.79 \pm 0.05 \\ 1.02 \pm 0.00 \\ 1.03 \pm 0.01 \\ 1.03 \pm 0.00 \end{array}$	$\begin{array}{c} 0.43 \pm 0.07 \\ 0.54 \pm 0.04 \\ 0.96 \pm 0.01 \\ 0.97 \pm 0.02 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 0.78 \pm 0.01 \\ 0.93 \pm 0.02 \\ 0.94 \pm 0.01 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.97 \pm 0.00 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.96 \pm 0.00 \\ 0.97 \pm 0.00 \\ 0.96 \pm 0.00 \\ 0.98 \pm 0.00 \\ 0.99 \pm 0.01 \end{array}$	$\begin{array}{c} 0.86 \pm 0.01 \\ 0.92 \pm 0.00 \\ 0.97 \pm 0.01 \\ 0.97 \pm 0.00 \\ 0.97 \pm 0.00 \end{array}$	$\begin{array}{c} 0.65 \pm 0.00 \\ 0.70 \pm 0.01 \\ 0.74 \pm 0.01 \\ 0.88 \pm 0.01 \\ 1.00 \pm 0.02 \end{array}$	$ \begin{vmatrix} 0.55 \pm 0.02 \\ 0.52 \pm 0.00 \\ 0.55 \pm 0.01 \\ 1.07 \pm 0.02 \\ 1.04 \pm 0.02 \end{vmatrix} $	$\begin{array}{c} 0.83 \pm 0.01 \\ 0.86 \pm 0.02 \\ 0.91 \pm 0.01 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.01 \end{array}$	$ \begin{array}{c c} 0.84 \pm 0.00 \\ 0.85 \pm 0.00 \\ 0.87 \pm 0.00 \\ 0.90 \pm 0.00 \\ 0.93 \pm 0.00 \end{array} $	$\begin{array}{c} 0.76 \pm 0.17 \\ 0.81 \pm 0.16 \\ 0.89 \pm 0.14 \\ 0.98 \pm 0.05 \\ 1.00 \pm 0.03 \end{array}$	1.00/1.00 0.99/0.99 0.97/0.96 0.88/0.86 0.50/0.47
500 w/clusters (C)	16 C 7 16 C 10 16 C 25 64 C 5 64 C 7	1.09 1.10 1.10 1.09 1.12	1.00 1.01 1.00 0.98 1.02	0.99 1.00 1.00 1.00 1.00	1.00 0.99 0.99 1.00 1.00	0.98 0.97 0.99 0.99 1.00	0.96 0.94 0.97 0.97 0.97	0.72 0.72 0.98 0.99 1.00	0.88 1.29 1.30 1.17 1.22	0.98 1.02 1.03 1.04 1.04	0.93 0.92 0.96 0.93 0.99	0.95 1.00 1.03 1.02 1.04	1.00/0.98 1.00/0.98 1.00/0.95 0.97/0.93 0.97/0.91
1000 w/clusters (C)	16 C 25	1.09	0.98	1.00	1.00	0.97	0.97	0.74	1.29	1.05	0.94	1.00	1.00/0.97

Table 9: Relative In-Distribution ROUGE-1 scores for various tasks and methods

M-d-l T	Mathad Tona					Tas	ks					A	D C1
Model Type	Method Type	task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	Average	Para. Saved
	base lora	$\begin{array}{c} 24.44 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.60 \pm 0.00 \\ 86.00 \pm 0.00 \end{array}$	$\begin{array}{c} 19.13 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	39.22 ± 0.00 93.67 ± 0.00	10.42 ± 0.00 94.33 ± 0.00	39.88 ± 0.00 78.43 ± 0.00	8.05 ± 0.00 74.90 ± 0.00	6.96 ± 0.00 26.87 ± 0.00	17.82 ± 0.00 95.00 ± 0.00	55.03 ± 0.00 68.66 ± 0.00	22.43 ± 16.49 81.14 ± 20.67	1.00 / 1.00 0.00/0.00
TIES	10 50 100 500	$76.50 \pm 0.00 \\ 55.80 \pm 0.00 \\ 52.43 \pm 0.00 \\ 35.18 \pm 0.00$	$\begin{array}{c} 49.00 \pm 1.73 \\ 35.00 \pm 0.00 \\ 34.00 \pm 0.00 \\ 22.00 \pm 0.00 \end{array}$	$44.33 \pm 4.04 \\ 18.00 \pm 5.20 \\ 19.67 \pm 4.62 \\ 1.00 \pm 0.00$	$\begin{array}{c} 9.80 \pm 0.58 \\ 2.42 \pm 0.50 \\ 1.09 \pm 1.66 \\ 0.00 \pm 0.00 \end{array}$	$\begin{array}{c} 78.56 \pm 0.96 \\ 85.78 \pm 0.96 \\ 83.33 \pm 0.00 \\ 78.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 40.44 \pm 0.00 \\ 26.75 \pm 0.00 \\ 28.57 \pm 0.00 \\ 24.32 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 53.10 \pm 0.67 \\ 49.96 \pm 0.00 \\ 48.89 \pm 0.00 \\ 43.80 \pm 0.04 \end{array}$	$ \begin{vmatrix} 15.48 \pm 0.12 \\ 16.73 \pm 0.00 \\ 15.18 \pm 0.42 \\ 9.96 \pm 0.13 \end{vmatrix} $	$\begin{array}{c} 77.67 \pm 1.15 \\ 30.00 \pm 3.46 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 54.89 \pm 0.06 \\ 53.87 \pm 0.02 \\ 53.44 \pm 0.02 \\ 27.90 \pm 0.03 \end{array}$	$\begin{array}{c} 49.98 \pm 23.33 \\ 37.43 \pm 23.49 \\ 33.76 \pm 25.22 \\ 24.40 \pm 23.79 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 1.00 1.00 / 1.00
SVD	SVD 2 SVD 4 SVD 8 SVD 16	$\begin{array}{c} 93.15 \pm 2.77 \\ 94.01 \pm 3.60 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 92.24 \pm 1.85 \\ 89.21 \pm 0.71 \\ 87.40 \pm 0.59 \\ 86.00 \pm 0.00 \end{array}$	$\begin{array}{c} 99.09 \pm 0.18 \\ 99.05 \pm 0.09 \\ 99.05 \pm 0.09 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 93.44 \pm 0.14 \\ 93.65 \pm 0.03 \\ 93.65 \pm 0.03 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 93.89 \pm 0.35 \\ 94.66 \pm 0.63 \\ 94.36 \pm 0.38 \\ 94.33 \pm 0.00 \end{array}$	$ \begin{vmatrix} 77.33 \pm 0.29 \\ 78.42 \pm 0.23 \\ 78.21 \pm 0.03 \\ 78.44 \pm 0.03 \end{vmatrix} $	$\begin{array}{c} 75.40 \pm 1.01 \\ 74.09 \pm 1.12 \\ 75.57 \pm 0.00 \\ 74.73 \pm 0.18 \end{array}$	$\begin{array}{c} 26.90 \pm 2.68 \\ 26.47 \pm 2.06 \\ 26.88 \pm 0.27 \\ 26.87 \pm 0.00 \end{array}$	$\begin{array}{c} 95.06 \pm 1.35 \\ 93.98 \pm 0.77 \\ 95.51 \pm 1.09 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 67.71 \pm 0.49 \\ 69.37 \pm 0.21 \\ 69.33 \pm 0.08 \\ 68.62 \pm 0.04 \end{array}$	$\begin{array}{c} 81.33 \pm 20.85 \\ 81.22 \pm 20.80 \\ 83.02 \pm 19.87 \\ 80.76 \pm 21.05 \end{array}$	0.88 / 0.88 0.75 / 0.75 0.50 / 0.50 0.00 / 0.00
10 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 96.67 \pm 0.58 \\ 95.67 \pm 0.58 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 87.00 \pm 1.00 \\ 90.00 \pm 1.00 \\ 88.33 \pm 0.58 \\ 86.67 \pm 0.58 \\ 86.00 \pm 0.00 \end{array}$	$\begin{array}{c} 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 94.00 \pm 0.67 \\ 93.00 \pm 0.33 \\ 93.67 \pm 0.00 \\ 93.67 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 93.11 \pm 0.38 \\ 94.89 \pm 0.51 \\ 94.78 \pm 0.38 \\ 94.33 \pm 0.00 \\ 94.33 \pm 0.00 \end{array}$	$ \begin{array}{c} 76.08 \pm 0.17 \\ 77.46 \pm 0.24 \\ 78.28 \pm 0.07 \\ 78.45 \pm 0.16 \\ 78.43 \pm 0.00 \end{array} $	$\begin{array}{c} 77.26 \pm 1.47 \\ 72.53 \pm 1.00 \\ 75.47 \pm 0.58 \\ 75.46 \pm 0.51 \\ 74.90 \pm 0.00 \end{array}$	$\begin{array}{c} 30.15 \pm 0.72 \\ 27.98 \pm 0.71 \\ 26.53 \pm 0.25 \\ 26.64 \pm 0.23 \\ 26.87 \pm 0.00 \end{array}$	$\begin{array}{c} 94.00 \pm 1.73 \\ 94.67 \pm 0.58 \\ 96.00 \pm 0.00 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 68.25 \pm 0.18 \\ 69.16 \pm 0.41 \\ 69.36 \pm 0.05 \\ 68.70 \pm 0.14 \\ 68.59 \pm 0.03 \end{array}$	$\begin{array}{c} 81.55 \pm 20.03 \\ 81.44 \pm 20.80 \\ 81.64 \pm 21.06 \\ 81.29 \pm 20.92 \\ 81.18 \pm 20.86 \end{array}$	1.00 / 0.90 1.00 / 0.80 1.00 / 0.60 1.00 / 0.20 1.00 / -0.60
10 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 97.00 \pm 0.00 \\ 96.67 \pm 0.58 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 91.00 \pm 1.00 \\ 89.33 \pm 0.58 \\ 88.67 \pm 0.58 \\ 86.67 \pm 0.58 \\ 86.00 \pm 0.00 \end{array}$	$\begin{array}{c} 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 93.56 \pm 0.19 \\ 93.22 \pm 0.19 \\ 93.67 \pm 0.00 \\ 93.67 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 93.56 \pm 0.69 \\ 94.44 \pm 0.19 \\ 94.56 \pm 0.38 \\ 94.33 \pm 0.00 \\ 94.33 \pm 0.00 \end{array}$	$77.64 \pm 0.25 \\ 77.84 \pm 0.21 \\ 78.19 \pm 0.08 \\ 78.46 \pm 0.03 \\ 78.44 \pm 0.03$	$\begin{array}{c} 75.78 \pm 1.25 \\ 72.24 \pm 0.59 \\ 75.97 \pm 0.58 \\ 74.90 \pm 0.00 \\ 74.79 \pm 0.19 \end{array}$	$\begin{array}{c} 28.71 \pm 0.09 \\ 26.84 \pm 0.50 \\ 26.43 \pm 0.34 \\ 26.72 \pm 0.13 \\ 26.87 \pm 0.00 \end{array}$	$\begin{array}{c} 96.00 \pm 1.00 \\ 94.67 \pm 0.58 \\ 96.00 \pm 0.00 \\ 95.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 68.69 \pm 0.08 \\ 69.55 \pm 0.08 \\ 69.38 \pm 0.11 \\ 68.65 \pm 0.03 \\ 68.64 \pm 0.03 \end{array}$	$\begin{array}{c} 82.09 \pm 20.68 \\ 81.38 \pm 21.11 \\ 81.69 \pm 21.07 \\ 81.24 \pm 20.91 \\ 81.17 \pm 20.86 \end{array}$	1.00/0.90 0.99/0.79 0.97/0.57 0.88/0.07 0.50/-1.10
50 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 92.76 \pm 3.53 \\ 95.33 \pm 2.08 \\ 97.00 \pm 0.00 \\ 96.33 \pm 0.58 \\ 95.67 \pm 0.58 \end{array}$	$\begin{array}{c} 84.67 \pm 1.15 \\ 87.33 \pm 2.08 \\ 90.33 \pm 1.53 \\ 92.67 \pm 0.58 \\ 88.33 \pm 0.58 \end{array}$	$\begin{array}{c} 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 86.17 \pm 5.81 \\ 92.60 \pm 0.29 \\ 93.78 \pm 0.19 \\ 93.56 \pm 0.19 \\ 93.56 \pm 0.19 \end{array}$	$\begin{array}{c} 79.83 \pm 6.08 \\ 90.35 \pm 1.00 \\ 93.00 \pm 0.58 \\ 93.00 \pm 0.58 \\ 94.67 \pm 0.67 \end{array}$	$\begin{array}{c} 73.55 \pm 1.39 \\ 75.43 \pm 1.33 \\ 76.27 \pm 0.49 \\ 77.24 \pm 0.19 \\ 78.45 \pm 0.14 \end{array}$	$ \begin{array}{c} 51.72 \pm 3.78 \\ 63.84 \pm 1.64 \\ 74.39 \pm 0.90 \\ 73.76 \pm 1.25 \\ 72.86 \pm 2.07 \end{array} $	$\begin{array}{c} 23.75 \pm 2.66 \\ 26.97 \pm 3.21 \\ 29.28 \pm 0.81 \\ 29.58 \pm 0.93 \\ 27.00 \pm 0.77 \end{array}$	$\begin{array}{c} 83.90 \pm 6.43 \\ 93.33 \pm 1.15 \\ 95.67 \pm 0.58 \\ 95.00 \pm 0.00 \\ 95.33 \pm 0.58 \end{array}$	$\begin{array}{c} 59.05 \pm 0.94 \\ 61.94 \pm 0.32 \\ 64.84 \pm 0.27 \\ 69.04 \pm 0.54 \\ 69.61 \pm 0.18 \end{array}$	$\begin{array}{c} 73.44 \pm 22.08 \\ 78.61 \pm 21.60 \\ 81.36 \pm 20.83 \\ 81.92 \pm 20.44 \\ 81.45 \pm 21.00 \end{array}$	1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84 1.00 / 0.68
50 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 94.06 \pm 3.54 \\ 97.00 \pm 0.00 \\ 96.67 \pm 0.58 \\ 97.00 \pm 0.00 \\ 95.00 \pm 0.00 \end{array}$	$\begin{array}{c} 85.67 \pm 1.15 \\ 85.67 \pm 1.53 \\ 91.00 \pm 2.00 \\ 91.00 \pm 1.00 \\ 88.00 \pm 0.00 \end{array}$	$\begin{array}{c} 98.67 \pm 0.58 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 90.35 \pm 1.37 \\ 93.67 \pm 0.00 \\ 93.56 \pm 0.19 \\ 93.33 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 89.97 \pm 1.78 \\ 92.22 \pm 0.69 \\ 93.22 \pm 0.51 \\ 94.11 \pm 0.51 \\ 94.44 \pm 0.19 \end{array}$	$ \begin{array}{c} 74.46 \pm 0.58 \\ 75.86 \pm 0.22 \\ 77.17 \pm 0.38 \\ 77.23 \pm 0.17 \\ 77.97 \pm 0.24 \end{array} $	$\begin{array}{c} 49.03 \pm 7.07 \\ 71.68 \pm 0.65 \\ 77.11 \pm 0.51 \\ 73.67 \pm 0.58 \\ 75.47 \pm 0.58 \end{array}$	$\begin{array}{c} 27.14 \pm 3.94 \\ 29.26 \pm 0.70 \\ 29.75 \pm 0.03 \\ 27.62 \pm 1.12 \\ 26.96 \pm 0.09 \end{array}$	$\begin{array}{c} 92.33 \pm 1.53 \\ 95.67 \pm 1.53 \\ 95.33 \pm 0.58 \\ 95.00 \pm 1.00 \\ 96.00 \pm 0.00 \end{array}$		$\begin{array}{c} 76.19 \pm 22.80 \\ 80.39 \pm 20.81 \\ 82.09 \pm 20.33 \\ 81.74 \pm 20.97 \\ 81.58 \pm 20.92 \end{array}$	1.00/0.98 0.99/0.95 0.97/0.89 0.88/0.72 0.50/0.18
100 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$76.43 \pm 7.07 \\ 90.10 \pm 5.85 \\ 95.56 \pm 2.49 \\ 96.00 \pm 0.00 \\ 95.00 \pm 0.00$	$76.67 \pm 4.93 \\ 84.00 \pm 1.00 \\ 86.67 \pm 0.58 \\ 87.33 \pm 1.15 \\ 91.00 \pm 0.00$	$\begin{array}{c} 91.61 \pm 2.75 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 89.99 \pm 1.07 \\ 85.52 \pm 5.34 \\ 92.24 \pm 1.68 \\ 93.89 \pm 0.19 \\ 93.56 \pm 0.19 \end{array}$	$\begin{array}{c} 47.89 \pm 8.62 \\ 75.88 \pm 12.57 \\ 90.89 \pm 1.17 \\ 93.00 \pm 0.58 \\ 93.11 \pm 0.19 \end{array}$	$ \begin{vmatrix} 63.17 \pm 1.31 \\ 71.15 \pm 3.61 \\ 74.57 \pm 0.50 \\ 76.68 \pm 0.18 \\ 76.93 \pm 0.23 \end{vmatrix} $	22.23 ± 5.27 46.10 ± 7.39 67.07 ± 3.81 74.84 ± 2.23 75.13 ± 0.84	$ \begin{vmatrix} 14.46 \pm 2.89 \\ 21.04 \pm 6.76 \\ 29.78 \pm 1.92 \\ 29.79 \pm 0.50 \\ 29.75 \pm 0.73 \end{vmatrix} $	$73.93 \pm 3.13 \\ 86.67 \pm 1.86 \\ 91.67 \pm 2.31 \\ 93.67 \pm 0.58 \\ 95.33 \pm 0.58$	$\begin{array}{c} 57.17 \pm 1.05 \\ 58.64 \pm 1.02 \\ 60.28 \pm 1.51 \\ 63.49 \pm 0.34 \\ 67.89 \pm 1.34 \end{array}$		1.00 / 0.99 1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84
100 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 90.70 \pm 1.07 \\ 95.33 \pm 1.53 \\ 97.00 \pm 0.00 \\ 96.33 \pm 0.58 \\ 96.33 \pm 0.58 \end{array}$	83.00 ± 2.65 85.00 ± 1.00 85.67 ± 1.53 90.33 ± 0.58 88.67 ± 0.58	$\begin{array}{c} 96.00 \pm 3.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \\ \end{array}$	$\begin{array}{c} 91.22 \pm 2.94 \\ 93.50 \pm 0.22 \\ 93.78 \pm 0.19 \\ 93.00 \pm 0.00 \\ 93.67 \pm 0.00 \end{array}$	$\begin{array}{c} 87.94 \pm 0.54 \\ 91.44 \pm 0.84 \\ 92.56 \pm 0.19 \\ 93.89 \pm 0.19 \\ 94.89 \pm 0.19 \end{array}$	$\begin{array}{c c} 73.07 \pm 0.93 \\ 75.00 \pm 0.19 \\ 76.01 \pm 0.13 \\ 77.04 \pm 0.30 \\ 78.16 \pm 0.18 \end{array}$	49.41 ± 2.04 65.09 ± 2.23 73.96 ± 0.89 77.33 ± 1.01 73.40 ± 0.12	$ \begin{array}{c c} 24.17 \pm 4.22 \\ 30.20 \pm 0.81 \\ 29.46 \pm 0.21 \\ 29.49 \pm 0.35 \\ 26.86 \pm 0.68 \end{array} $	$\begin{array}{c} 82.33 \pm 2.08 \\ 91.67 \pm 0.58 \\ 94.33 \pm 1.53 \\ 96.00 \pm 0.00 \\ 96.00 \pm 0.00 \end{array}$	$\begin{array}{c} 58.18 \pm 0.44 \\ 60.92 \pm 0.26 \\ 64.07 \pm 0.37 \\ 68.76 \pm 0.25 \\ 69.47 \pm 0.23 \end{array}$	$73.60 \pm 22.23 \\ 78.72 \pm 20.72 \\ 80.58 \pm 20.59 \\ 82.12 \pm 20.35 \\ 81.64 \pm 21.15$	1.00/0.99 0.99/0.97 0.97/0.93 0.88/0.80 0.50/0.34
100 w/clusters (C)	16 C 5 16 C 7	$98.33 \pm 0.47 \\ 97.67 \pm 0.47$	89.00 ± 0.82 87.00 ± 0.82	99.00 ± 0.00 99.00 ± 0.00	93.25 ± 0.40 93.46 ± 0.29	92.89 ± 0.87 93.11 ± 0.68	76.33 ± 0.28 76.55 ± 0.29	78.24 ± 2.26 79.03 ± 2.05	28.45 ± 0.40 28.62 ± 1.29	96.67 ± 0.47 95.33 ± 0.47	75.93 ± 8.31 76.55 ± 6.91	82.81 ± 20.02 82.63 ± 19.76	1.00/0.95 1.00/0.93
500 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 54.44 \pm 6.87 \\ 58.08 \pm 11.52 \\ 69.21 \pm 2.03 \\ 79.77 \pm 0.37 \\ 93.83 \pm 2.52 \end{array}$	$\begin{array}{c} 47.00 \pm 2.83 \\ 47.00 \pm 7.07 \\ 54.50 \pm 9.19 \\ 79.50 \pm 2.12 \\ 85.00 \pm 0.00 \end{array}$	$\begin{array}{c} 82.21 \pm 3.59 \\ 82.06 \pm 1.69 \\ 88.33 \pm 4.04 \\ 95.89 \pm 2.83 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 73.38 \pm 14.97 \\ 78.62 \pm 11.23 \\ 91.11 \pm 0.38 \\ 91.89 \pm 1.39 \\ 93.78 \pm 0.19 \end{array}$	$ 80.13 \pm 3.68 \\ 85.57 \pm 1.48 \\ 88.78 \pm 0.38 \\ 88.67 \pm 0.00 \\ 90.56 \pm 0.38 $	57.42 ± 5.29 59.19 ± 3.70 67.71 ± 2.59 70.27 ± 1.73 73.25 ± 1.86	$ \begin{array}{c c} \hline 18.33 \pm 1.33 \\ 22.76 \pm 3.95 \\ 34.79 \pm 6.86 \\ 46.64 \pm 10.58 \\ 51.14 \pm 3.86 \\ \hline \end{array} $	12.19 ± 0.30 13.15 ± 1.94 13.80 ± 1.95 14.63 ± 0.25 22.93 ± 3.86	$ 72.67 \pm 6.03 75.33 \pm 4.04 77.67 \pm 2.31 81.00 \pm 5.00 87.33 \pm 2.31 $	$\begin{array}{c} 55.79 \pm 0.20 \\ 56.07 \pm 0.52 \\ 56.78 \pm 0.73 \\ 56.88 \pm 0.55 \\ 58.48 \pm 0.20 \end{array}$	$\begin{array}{c} 55.64 \pm 24.25 \\ 58.16 \pm 24.56 \\ 64.61 \pm 24.79 \\ 70.20 \pm 24.63 \\ 75.20 \pm 23.90 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 0.99 1.00 / 0.98 1.00 / 0.97
500 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 54.30 \pm 1.13 \\ 75.10 \pm 4.92 \\ 96.94 \pm 0.42 \\ 97.67 \pm 0.58 \\ 98.00 \pm 0.00 \end{array}$	37.00 ± 5.66 46.50 ± 3.54 82.50 ± 0.71 83.50 ± 2.12 88.50 ± 0.71	$77.67 \pm 0.58 \\ 91.67 \pm 1.53 \\ 93.33 \pm 0.58 \\ 98.00 \pm 0.00 \\ 99.00 \pm 0.00$	$\begin{array}{c} 91.00 \pm 0.00 \\ 91.56 \pm 0.19 \\ 93.89 \pm 0.69 \\ 93.56 \pm 0.19 \\ 93.78 \pm 0.19 \end{array}$	$\begin{array}{c} 90.56 \pm 0.19 \\ 91.56 \pm 0.38 \\ 90.67 \pm 0.00 \\ 92.00 \pm 0.00 \\ 93.00 \pm 0.88 \end{array}$	$ \begin{vmatrix} 67.63 \pm 0.45 \\ 72.03 \pm 0.15 \\ 75.99 \pm 0.64 \\ 75.80 \pm 0.16 \\ 76.33 \pm 0.29 \end{vmatrix} $	$ \begin{vmatrix} 48.81 \pm 0.35 \\ 52.63 \pm 0.86 \\ 55.63 \pm 1.07 \\ 66.19 \pm 0.81 \\ 74.60 \pm 1.21 \end{vmatrix} $	$ \begin{vmatrix} 14.70 \pm 0.65 \\ 13.93 \pm 0.02 \\ 14.74 \pm 0.27 \\ 28.67 \pm 0.49 \\ 27.82 \pm 0.42 \end{vmatrix} $	$79.00 \pm 1.00 \\ 81.67 \pm 1.53 \\ 86.33 \pm 0.58 \\ 93.00 \pm 0.00 \\ 95.33 \pm 0.58$	$\begin{array}{c} 57.66 \pm 0.19 \\ 58.50 \pm 0.20 \\ 59.43 \pm 0.05 \\ 61.53 \pm 0.13 \\ 63.70 \pm 0.14 \end{array}$	$62.69 \pm 23.46 \\ 68.24 \pm 24.29 \\ 74.69 \pm 25.01 \\ 78.84 \pm 21.50 \\ 80.75 \pm 21.60$	1.00/1.00 0.99/0.99 0.97/0.96 0.88/0.86 0.50/0.47
500 w/clusters (C)	16 C 7 16 C 10 16 C 25 64 C 5 64 C 7	95.00 96.00 96.00 95.00 98.00	86.00 87.00 86.00 84.00 88.00	98.00 99.00 99.00 99.00 99.00	93.67 93.00 92.71 93.67 94.00	91.67 91.33 93.00 92.67 93.33	75.10 74.17 76.42 76.45 76.42	55.52 55.14 75.42 76.43 76.67	20.50 30.29 30.40 27.49 28.48	90.00 94.00 95.00 96.00 96.00	63.57 63.09 66.07 64.10 68.00	76.90 78.30 81.00 80.48 81.79	1.00/0.98 1.00/0.98 1.00/0.95 0.97/0.93 0.97/0.91
1000 w/clusters (C)	16 C 25	95.00	84.00	99.00	93.67	90.67	76.43	56.71	30.20	97.00	64.61	78.73	1.00/0.97

Table 10: Absolute In-Distribution ROUGE-1 scores for various tasks and methods

Model Type	Method Type					Ta	sks					. Average	Para. Saved
Woder Type		task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	Average	raia. Saveu
	base lora	$\begin{array}{c} 0.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.02 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{c} 0.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{ c c c c c c }\hline 0.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ \end{array}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	1.00 / 1.00 0.00 / 0.00
TIES	10 50 100 500	$\begin{array}{c} 0.69 \pm 0.00 \\ 0.45 \pm 0.00 \\ 0.41 \pm 0.00 \\ 0.22 \pm 0.00 \end{array}$	$\begin{array}{c} 0.57 \pm 0.02 \\ 0.41 \pm 0.00 \\ 0.40 \pm 0.00 \\ 0.26 \pm 0.00 \end{array}$	$\begin{array}{c} 0.45 \pm 0.04 \\ 0.18 \pm 0.05 \\ 0.20 \pm 0.05 \\ 0.01 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.10 \pm 0.01 \\ 0.03 \pm 0.01 \\ 0.01 \pm 0.02 \\ 0.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.57 \pm 0.03 \\ 0.70 \pm 0.02 \\ 0.65 \pm 0.00 \\ 0.60 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.39 \pm 0.01 \\ 0.36 \pm 0.00 \\ 0.36 \pm 0.00 \\ 0.32 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.21 \pm 0.00 \\ 0.21 \pm 0.00 \\ 0.21 \pm 0.00 \\ 0.07 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.82 \pm 0.01 \\ 0.32 \pm 0.04 \\ 0.01 \pm 0.00 \\ 0.01 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.00 \pm 0.00 \\ \end{vmatrix} $	$\begin{array}{c} 0.38 \pm 0.28 \\ 0.27 \pm 0.22 \\ 0.23 \pm 0.22 \\ 0.15 \pm 0.20 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 1.00 1.00 / 1.00
SVD	SVD 2 SVD 4 SVD 8 SVD 16	$\begin{array}{c} 0.98 \pm 0.03 \\ 0.99 \pm 0.04 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.07 \pm 0.02 \\ 1.04 \pm 0.01 \\ 1.02 \pm 0.01 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.99 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.98 \pm 0.01 \\ 1.01 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.98 \pm 0.03 \\ 1.11 \pm 0.00 \\ 1.02 \pm 0.05 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.94 \pm 0.01 \\ 0.97 \pm 0.02 \\ 1.00 \pm 0.00 \\ 0.99 \pm 0.01 \end{array}$	$ \begin{vmatrix} 1.03 \pm 0.17 \\ 0.99 \pm 0.13 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.00 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{array}{c} 0.15 \pm 0.29 \\ 0.90 \pm 0.17 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array} $	$\begin{array}{c} 0.91 \pm 0.28 \\ 1.00 \pm 0.08 \\ 1.00 \pm 0.02 \\ 1.00 \pm 0.00 \end{array}$	0.88 / 0.88 0.75 / 0.75 0.50 / 0.50 0.00 / 0.00
10 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 1.02 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.01 \\ 1.05 \pm 0.01 \\ 1.03 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.01 \pm 0.02 \\ 0.98 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.96 \pm 0.01 \\ 1.02 \pm 0.02 \\ 1.02 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.11 \pm 0.11 \\ 1.11 \pm 0.00 \\ 1.11 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.89 \pm 0.03 \\ 0.93 \pm 0.01 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 1.19 \pm 0.04 \\ 1.10 \pm 0.04 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.99 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.33 \pm 0.58 \\ 0.67 \pm 0.58 \\ 0.67 \pm 0.58 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.95 \pm 0.27 \\ 0.98 \pm 0.19 \\ 0.98 \pm 0.19 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	1.00 / 0.90 1.00 / 0.80 1.00 / 0.60 1.00 / 0.20 1.00 / -0.60
10 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 1.02 \pm 0.00 \\ 1.02 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.06 \pm 0.01 \\ 1.04 \pm 0.01 \\ 1.03 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 0.98 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.97 \pm 0.03 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.15 \pm 0.06 \\ 1.11 \pm 0.00 \\ 1.07 \pm 0.06 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.92 \pm 0.02 \\ 0.92 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.01 \end{array}$	$ \begin{vmatrix} 1.17 \pm 0.04 \\ 1.02 \pm 0.04 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 1.01 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.67 \pm 0.58 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.20 \\ 1.01 \pm 0.05 \\ 1.01 \pm 0.03 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	1.00/0.90 0.99/0.79 0.97/0.57 0.88/0.07 0.50/-1.10
50 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.91 \pm 0.06 \\ 1.00 \pm 0.02 \\ 1.02 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$\begin{array}{c} 0.98 \pm 0.01 \\ 1.02 \pm 0.02 \\ 1.05 \pm 0.02 \\ 1.08 \pm 0.01 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.91 \pm 0.09 \\ 1.00 \pm 0.01 \end{vmatrix} $	$\begin{array}{c} 0.78 \pm 0.05 \\ 0.90 \pm 0.03 \\ 0.95 \pm 0.02 \\ 0.95 \pm 0.02 \\ 1.01 \pm 0.02 \end{array}$	$\begin{array}{c} 0.89 \pm 0.29 \\ 0.85 \pm 0.42 \\ 1.15 \pm 0.17 \\ 1.04 \pm 0.06 \\ 1.11 \pm 0.00 \end{array}$	$\begin{array}{c} 0.34 \pm 0.06 \\ 0.56 \pm 0.04 \\ 0.81 \pm 0.03 \\ 0.92 \pm 0.03 \\ 0.95 \pm 0.04 \end{array}$	$\begin{array}{c} 0.50 \pm 0.45 \\ 0.98 \pm 0.23 \\ 1.14 \pm 0.00 \\ 1.21 \pm 0.07 \\ 1.02 \pm 0.04 \end{array}$	$\begin{array}{c} 0.86 \pm 0.07 \\ 0.98 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.01 \end{array}$	$ \begin{vmatrix} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.67 \pm 0.58 \\ 1.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.72 \pm 0.35 \\ 0.83 \pm 0.34 \\ 0.91 \pm 0.33 \\ 0.99 \pm 0.20 \\ 1.01 \pm 0.04 \end{array}$	1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84 1.00 / 0.68
50 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.96 \pm 0.05 \\ 1.02 \pm 0.00 \\ 1.02 \pm 0.01 \\ 1.02 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 1.00 \pm 0.02 \\ 1.06 \pm 0.02 \\ 1.06 \pm 0.01 \\ 1.02 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.01 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.95 \pm 0.04 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.01 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.87 \pm 0.01 \\ 0.92 \pm 0.03 \\ 0.96 \pm 0.02 \\ 0.99 \pm 0.02 \\ 1.00 \pm 0.01 \end{array}$	$ \begin{vmatrix} 1.04 \pm 0.06 \\ 1.15 \pm 0.06 \\ 1.22 \pm 0.00 \\ 1.15 \pm 0.06 \\ 1.04 \pm 0.06 \end{vmatrix} $	$\begin{array}{c} 0.31 \pm 0.08 \\ 0.73 \pm 0.04 \\ 0.94 \pm 0.01 \\ 0.92 \pm 0.01 \\ 0.99 \pm 0.00 \end{array}$	$\begin{array}{c} 0.98 \pm 0.23 \\ 1.17 \pm 0.04 \\ 1.17 \pm 0.04 \\ 1.10 \pm 0.08 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.97 \pm 0.02 \\ 1.01 \pm 0.02 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.01 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.81 \pm 0.35 \\ 0.90 \pm 0.33 \\ 0.94 \pm 0.33 \\ 1.02 \pm 0.07 \\ 1.01 \pm 0.02 \end{array}$	1.00/0.98 0.99/0.95 0.97/0.89 0.88/0.72 0.50/0.18
100 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.54 \pm 0.16 \\ 0.85 \pm 0.15 \\ 1.00 \pm 0.04 \\ 1.01 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.89 \pm 0.06 \\ 0.98 \pm 0.01 \\ 1.01 \pm 0.01 \\ 1.02 \pm 0.01 \\ 1.06 \pm 0.00 \end{array}$	$\begin{array}{c} 0.90 \pm 0.04 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.89 \pm 0.05 \\ 0.86 \pm 0.13 \\ 0.98 \pm 0.02 \\ 1.01 \pm 0.01 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.42 \pm 0.08 \\ 0.70 \pm 0.14 \\ 0.88 \pm 0.04 \\ 0.95 \pm 0.02 \\ 0.96 \pm 0.01 \end{array}$	$\begin{array}{c} 0.44 \pm 0.00 \\ 0.74 \pm 0.28 \\ 1.07 \pm 0.06 \\ 1.11 \pm 0.00 \\ 1.11 \pm 0.11 \end{array}$	$\begin{array}{c} 0.08 \pm 0.02 \\ 0.28 \pm 0.07 \\ 0.58 \pm 0.09 \\ 0.81 \pm 0.06 \\ 0.92 \pm 0.02 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.48 \pm 0.55 \\ 1.10 \pm 0.04 \\ 1.21 \pm 0.00 \\ 1.21 \pm 0.07 \end{array}$	$\begin{array}{c} 0.76 \pm 0.05 \\ 0.91 \pm 0.02 \\ 0.96 \pm 0.02 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.49 \pm 0.36 \\ 0.68 \pm 0.36 \\ 0.86 \pm 0.32 \\ 0.91 \pm 0.33 \\ 0.93 \pm 0.33 \end{array}$	1.00 / 0.99 1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84
100 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.85 \pm 0.03 \\ 0.99 \pm 0.02 \\ 1.02 \pm 0.00 \\ 1.01 \pm 0.01 \\ 1.01 \pm 0.01 \end{array}$	$\begin{array}{c} 0.97 \pm 0.03 \\ 0.99 \pm 0.01 \\ 1.00 \pm 0.02 \\ 1.05 \pm 0.01 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 0.97 \pm 0.03 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.95 \pm 0.06 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.01 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.80 \pm 0.02 \\ 0.90 \pm 0.03 \\ 0.94 \pm 0.01 \\ 0.98 \pm 0.01 \\ 1.02 \pm 0.01 \end{array}$	$ \begin{array}{c} 0.81 \pm 0.17 \\ 1.04 \pm 0.06 \\ 1.04 \pm 0.06 \\ 1.15 \pm 0.06 \\ 1.19 \pm 0.06 \end{array} $	$\begin{array}{c} 0.29 \pm 0.04 \\ 0.55 \pm 0.04 \\ 0.78 \pm 0.01 \\ 0.94 \pm 0.01 \\ 0.93 \pm 0.01 \end{array}$	$ \begin{array}{c} 0.60 \pm 0.34 \\ 1.07 \pm 0.07 \\ 1.14 \pm 0.00 \\ 1.21 \pm 0.00 \\ 1.02 \pm 0.04 \end{array} $	$\begin{array}{c} 0.87 \pm 0.02 \\ 0.96 \pm 0.01 \\ 0.99 \pm 0.02 \\ 1.01 \pm 0.00 \\ 1.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.33 \pm 0.58 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.71 \pm 0.33 \\ 0.85 \pm 0.32 \\ 0.89 \pm 0.31 \\ 0.97 \pm 0.28 \\ 1.02 \pm 0.06 \end{array}$	1.00/0.99 0.99/0.97 0.97/0.93 0.88/0.80 0.50/0.34
100 w/clusters (C)	16 C 5 16 C 7	1.13 ± 0.01 1.12 ± 0.01	1.03 ± 0.01 1.01 ± 0.01	1.00 ± 0.00 1.00 ± 0.00	1.00 ± 0.01 1.00 ± 0.00	0.97 ± 0.03 0.98 ± 0.03	1.24 ± 0.09 1.16 ± 0.11	0.88 ± 0.06 0.92 ± 0.05	1.42 ± 0.07 1.45 ± 0.04	1.05 ± 0.01 1.03 ± 0.01	0.65 ± 0.46 0.69 ± 0.49	1.04 ± 0.19 1.04 ± 0.18	1.00/0.95 1.00/0.93
500 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 0.22 \pm 0.10 \\ 0.27 \pm 0.18 \\ 0.40 \pm 0.04 \\ 0.61 \pm 0.04 \\ 0.95 \pm 0.02 \end{array}$	$ \begin{vmatrix} 0.55 \pm 0.03 \\ 0.55 \pm 0.08 \\ 0.63 \pm 0.11 \\ 0.92 \pm 0.02 \\ 0.99 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.81 \pm 0.05 \\ 0.82 \pm 0.02 \\ 0.89 \pm 0.04 \\ 0.97 \pm 0.03 \\ 1.00 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.33 \pm 0.49 \\ 0.49 \pm 0.37 \\ 0.91 \pm 0.01 \\ 0.93 \pm 0.05 \\ 1.00 \pm 0.01 \end{vmatrix} $	$\begin{array}{c} 0.70 \pm 0.03 \\ 0.75 \pm 0.01 \\ 0.80 \pm 0.01 \\ 0.80 \pm 0.00 \\ 0.86 \pm 0.01 \end{array}$	$ \begin{vmatrix} 0.15 \pm 0.17 \\ 0.22 \pm 0.11 \\ 0.48 \pm 0.06 \\ 0.74 \pm 0.17 \\ 0.85 \pm 0.28 \end{vmatrix} $	$\begin{array}{c} 0.03 \pm 0.01 \\ 0.05 \pm 0.05 \\ 0.13 \pm 0.04 \\ 0.22 \pm 0.11 \\ 0.28 \pm 0.06 \end{array}$	$ \begin{vmatrix} 0.00 \pm 0.00 \\ 0.02 \pm 0.04 \\ 0.05 \pm 0.08 \\ 0.12 \pm 0.08 \\ 0.55 \pm 0.39 \end{vmatrix} $	$\begin{array}{c} 0.76 \pm 0.06 \\ 0.79 \pm 0.04 \\ 0.82 \pm 0.02 \\ 0.85 \pm 0.05 \\ 0.92 \pm 0.02 \end{array}$	$ \begin{vmatrix} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.35 \pm 0.35 \\ 0.39 \pm 0.34 \\ 0.51 \pm 0.35 \\ 0.61 \pm 0.36 \\ 0.73 \pm 0.36 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 0.99 1.00 / 0.98 1.00 / 0.97
500 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 0.21 \pm 0.02 \\ 0.54 \pm 0.08 \\ 0.99 \pm 0.03 \\ 1.02 \pm 0.01 \\ 1.03 \pm 0.00 \end{array}$	$\begin{array}{c} 0.43 \pm 0.07 \\ 0.54 \pm 0.04 \\ 0.96 \pm 0.01 \\ 0.97 \pm 0.02 \\ 1.03 \pm 0.01 \end{array}$	$\begin{array}{c} 0.78 \pm 0.01 \\ 0.93 \pm 0.02 \\ 0.94 \pm 0.01 \\ 0.99 \pm 0.00 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.90 \pm 0.00 \\ 0.92 \pm 0.01 \\ 1.01 \pm 0.03 \\ 1.00 \pm 0.01 \\ 1.00 \pm 0.01 \end{array}$	$\begin{array}{c} 0.86 \pm 0.01 \\ 0.90 \pm 0.01 \\ 0.87 \pm 0.00 \\ 0.92 \pm 0.00 \\ 0.95 \pm 0.03 \end{array}$	$\begin{array}{c} 0.59 \pm 0.06 \\ 0.63 \pm 0.13 \\ 1.04 \pm 0.17 \\ 1.15 \pm 0.06 \\ 1.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.21 \pm 0.01 \\ 0.26 \pm 0.02 \\ 0.36 \pm 0.00 \\ 0.61 \pm 0.01 \\ 0.78 \pm 0.01 \end{array}$	$ \begin{vmatrix} 0.12 \pm 0.04 \\ 0.14 \pm 0.00 \\ 0.14 \pm 0.00 \\ 1.07 \pm 0.00 \\ 1.07 \pm 0.00 \end{vmatrix} $	$\begin{array}{c} 0.83 \pm 0.01 \\ 0.86 \pm 0.02 \\ 0.91 \pm 0.01 \\ 0.98 \pm 0.00 \\ 1.00 \pm 0.01 \end{array}$	$ \begin{vmatrix} 0.00 \pm 0.00 \\ \end{vmatrix} $	$\begin{array}{c} 0.50 \pm 0.34 \\ 0.57 \pm 0.34 \\ 0.71 \pm 0.39 \\ 0.87 \pm 0.33 \\ 0.88 \pm 0.31 \end{array}$	1.00/1.00 0.99/0.99 0.97/0.96 0.88/0.86 0.50/0.47
500 w/clusters (C)	16 C 7 16 C 10 16 C 25 64 C 5 64 C 7	1.08 1.10 1.10 1.09 1.13	1.00 1.01 1.00 0.98 1.02	0.99 1.00 1.00 1.00 1.00	1.00 0.98 0.99 1.00 1.01	0.92 0.91 0.97 0.96 0.98	1.01 1.01 1.12 1.12 1.12	0.39 0.37 0.81 0.83 0.90	0.62 1.51 1.42 1.33 1.42	0.98 1.02 1.03 1.04 1.04	0.00 0.00 0.00 0.00 0.00	0.80 0.89 0.95 0.94 0.96	1.00/0.98 1.00/0.98 1.00/0.95 0.97/0.93 0.97/0.91
1000 w/clusters (C)	16 C 25	1.09	0.98	1.00	1.00	0.89	1.01	0.39	1.42	1.05	0.00	0.88	1.00/0.97

Table 11: Relative In-Distribution exact match scores for various tasks and methods

Model Type	Method Type					Ta	sks					. Average	Para. Saved
wioder Type	wichiod Type	task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	Average	l ara. Saved
	base lora	8.59 ± 0.08 0.36 ± 0.01	$\begin{array}{c} 9.15 \pm 0.00 \\ 0.17 \pm 0.00 \end{array}$	$\begin{array}{c} 2.55 \pm 0.00 \\ 0.01 \pm 0.00 \end{array}$	2.88 ± 0.00 0.12 ± 0.00	2.34 ± 0.00 0.11 ± 0.00	3.46 ± 0.04 0.76 ± 0.02	6.40 ± 0.18 1.17 ± 0.07	5.55 ± 0.00 1.94 ± 0.00	8.60 ± 0.00 0.16 ± 0.00	$ \begin{vmatrix} 2.67 \pm 0.00 \\ 0.85 \pm 0.00 \end{vmatrix} $	5.19 ± 2.65 0.57 ± 0.59	1.00 / 1.00 0.00 / 0.00
SVD	SVD 2 SVD 4 SVD 8	0.32 ± 0.01 0.33 ± 0.01 0.35 ± 0.01	$\begin{array}{c} 0.15 \pm 0.00 \\ 0.16 \pm 0.00 \\ 0.17 \pm 0.00 \end{array}$	$\begin{array}{c} 0.01 \pm 0.00 \\ 0.01 \pm 0.00 \\ 0.01 \pm 0.00 \end{array}$	0.12 ± 0.00 0.12 ± 0.00 0.12 ± 0.00	0.10 ± 0.00 0.11 ± 0.00 0.11 ± 0.00	$ \begin{vmatrix} 0.76 \pm 0.02 \\ 0.76 \pm 0.02 \\ 0.77 \pm 0.02 \end{vmatrix} $	1.13 ± 0.08 1.14 ± 0.08 1.16 ± 0.07	1.94 ± 0.00 1.94 ± 0.00 1.94 ± 0.00	$\begin{array}{c} 0.13 \pm 0.00 \\ 0.14 \pm 0.00 \\ 0.15 \pm 0.00 \end{array}$	$ \begin{vmatrix} 0.97 \pm 0.00 \\ 0.86 \pm 0.00 \\ 0.84 \pm 0.00 \end{vmatrix} $	0.57 ± 0.60 0.56 ± 0.59 0.51 ± 0.58	0.88 / 0.88 0.75 / 0.75 0.50 / 0.50
	SVD 16	0.36 ± 0.01	0.17 ± 0.00	0.01 ± 0.00	0.12 ± 0.00 0.12 ± 0.00	0.11 ± 0.00	0.76 ± 0.02	1.14 ± 0.06	1.94 ± 0.00	0.16 ± 0.00	0.85 ± 0.00	0.56 ± 0.59	0.00 / 0.00
	16 D	0.33 ± 0.01	0.15 ± 0.01	0.01 ± 0.00	0.12 ± 0.00	0.10 ± 0.00	0.76 ± 0.03	1.13 ± 0.08	1.95 ± 0.01	0.14 ± 0.00	1.00 ± 0.02	0.57 ± 0.61	1.00 / 0.90
10 diagonal (D)	32 D 64 D	0.33 ± 0.01 0.35 ± 0.01	0.16 ± 0.00 0.17 ± 0.00	0.01 ± 0.00 0.01 ± 0.00	0.12 ± 0.00 0.12 ± 0.00	0.10 ± 0.00 0.11 ± 0.00	0.75 ± 0.02 0.75 ± 0.02	1.11 ± 0.07 1.11 ± 0.07	1.93 ± 0.00 1.94 ± 0.00	0.14 ± 0.01 0.15 ± 0.00	$0.88 \pm 0.00 \\ 0.84 \pm 0.00$	0.55 ± 0.60 0.55 ± 0.59	1.00 / 0.80 1.00 / 0.60
	128 D 256 D	$\begin{array}{c} 0.35 \pm 0.01 \\ 0.36 \pm 0.01 \end{array}$	$\begin{array}{c} 0.17 \pm 0.00 \\ 0.17 \pm 0.00 \end{array}$	$\begin{array}{c} 0.01 \pm 0.00 \\ 0.01 \pm 0.00 \end{array}$	$\begin{array}{c} 0.12 \pm 0.00 \\ 0.12 \pm 0.00 \end{array}$	0.11 ± 0.00 0.11 ± 0.00	0.75 ± 0.02 0.75 ± 0.02	1.11 ± 0.07 1.12 ± 0.07	1.94 ± 0.00 1.94 ± 0.00	$\begin{array}{c} 0.16 \pm 0.00 \\ 0.16 \pm 0.00 \end{array}$	$\begin{array}{c} 0.84 \pm 0.00 \\ 0.85 \pm 0.00 \end{array}$	0.56 ± 0.59 0.56 ± 0.59	1.00 / 0.20 1.00 / -0.60
	16 F 32 F	0.33 ± 0.00 0.33 ± 0.01	0.15 ± 0.00 0.16 ± 0.00	0.01 ± 0.00 0.01 ± 0.00	0.12 ± 0.00 0.12 ± 0.00	0.10 ± 0.00 0.10 ± 0.00	0.76 ± 0.02 0.75 ± 0.02	1.20 ± 0.02 1.11 ± 0.07	1.95 ± 0.00 1.94 ± 0.00	0.13 ± 0.00 0.14 ± 0.00	0.97 ± 0.00 0.86 ± 0.00	0.57 ± 0.61 0.55 ± 0.60	1.00/0.90 0.99/0.79
10 full (F)	64 F	0.34 ± 0.01	0.16 ± 0.00 0.16 ± 0.00	0.01 ± 0.00 0.01 ± 0.00	0.12 ± 0.00	0.10 ± 0.00 0.11 ± 0.00	0.75 ± 0.02	1.11 ± 0.07	1.94 ± 0.00 1.94 ± 0.00	0.14 ± 0.00 0.15 ± 0.00	0.84 ± 0.00	0.55 ± 0.60 0.55 ± 0.59	0.97/0.57
	128 F 256 F	0.35 ± 0.01 0.36 ± 0.01	$0.17 \pm 0.00 \\ 0.17 \pm 0.00$	0.01 ± 0.00 0.01 ± 0.00	0.12 ± 0.00 0.12 ± 0.00	0.11 ± 0.00 0.11 ± 0.00	0.75 ± 0.02 0.75 ± 0.02	1.12 ± 0.07 1.12 ± 0.07	1.94 ± 0.00 1.94 ± 0.00	0.16 ± 0.00 0.16 ± 0.00	0.84 ± 0.00 0.85 ± 0.00	0.56 ± 0.59 0.56 ± 0.59	0.88/0.07 0.50/-1.10
	16 D	0.61 ± 0.06	0.19 ± 0.02	0.03 ± 0.01	0.29 ± 0.04	0.36 ± 0.04	0.95 ± 0.05	1.73 ± 0.21	2.66 ± 0.22	0.32 ± 0.11	1.98 ± 0.01	0.91 ± 0.88	1.00 / 0.98
50 diagonal (D)	32 D 64 D	0.37 ± 0.02 0.33 ± 0.02	0.16 ± 0.00 0.15 ± 0.00	0.01 ± 0.00 0.01 ± 0.00	0.19 ± 0.03 0.12 ± 0.00	0.18 ± 0.01 0.10 ± 0.00	0.85 ± 0.05 0.79 ± 0.02	1.37 ± 0.14 1.12 ± 0.08	2.12 ± 0.05 1.97 ± 0.01	0.16 ± 0.00 0.13 ± 0.01	1.65 ± 0.03 1.13 ± 0.03	0.71 ± 0.73 0.59 ± 0.63	1.00 / 0.96 1.00 / 0.92
	128 D	0.33 ± 0.01	0.15 ± 0.00	0.01 ± 0.00	0.12 ± 0.00	0.10 ± 0.00	0.76 ± 0.03	1.10 ± 0.05	1.93 ± 0.01	0.14 ± 0.00	0.93 ± 0.01	0.56 ± 0.60	1.00 / 0.84
	256 D	0.34 ± 0.01	0.16 ± 0.00	0.01 ± 0.00	0.12 ± 0.00	0.10 ± 0.00	0.76 ± 0.03	1.11 ± 0.05	1.93 ± 0.00	0.15 ± 0.00	0.85 ± 0.00	0.55 ± 0.59	1.00 / 0.68
	16 F 32 F	0.47 ± 0.06 0.36 ± 0.02	0.17 ± 0.00 0.16 ± 0.00	0.02 ± 0.00 0.01 ± 0.00	0.20 ± 0.02 0.14 ± 0.00	0.19 ± 0.04 0.11 ± 0.00	0.86 ± 0.03 0.80 ± 0.03	1.71 ± 0.10 1.14 ± 0.08	2.20 ± 0.04 2.00 ± 0.01	0.17 ± 0.01 0.14 ± 0.00	1.84 ± 0.07 1.32 ± 0.02	0.78 ± 0.80 0.62 ± 0.65	1.00/0.98 0.99/0.95
50 full (F)	64 F	0.33 ± 0.01	0.15 ± 0.00	0.01 ± 0.00	0.12 ± 0.00	0.10 ± 0.00	0.77 ± 0.03	1.10 ± 0.06	1.94 ± 0.00	0.13 ± 0.00	1.02 ± 0.00	0.57 ± 0.61	0.97/0.89
	128 F 256 F	0.33 ± 0.01 0.35 ± 0.01	$\begin{array}{c} 0.16 \pm 0.00 \\ 0.16 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.01 \pm 0.00 \end{array}$	0.12 ± 0.00 0.12 ± 0.00	0.10 ± 0.00 0.11 ± 0.00	0.76 ± 0.03 0.76 ± 0.03	1.11 ± 0.05 1.11 ± 0.05	1.93 ± 0.00 1.94 ± 0.00	0.14 ± 0.00 0.15 ± 0.00	0.87 ± 0.00 0.84 ± 0.00	0.55 ± 0.60 0.55 ± 0.59	0.88/0.72 0.50/0.18
	16 D	1.69 ± 0.49	0.26 ± 0.04	0.18 ± 0.07	0.34 ± 0.02	1.01 ± 0.20	1.45 ± 0.10	3.59 ± 0.25	3.72 ± 0.72	0.44 ± 0.20	2.37 ± 0.09	1.51 ± 1.32	1.00 / 0.99
100 diagonal (D)	32 D 64 D	0.67 ± 0.24 0.39 ± 0.06	0.18 ± 0.01 0.16 ± 0.00	0.06 ± 0.05 0.01 ± 0.00	0.31 ± 0.06 0.18 ± 0.02	0.35 ± 0.08 0.14 ± 0.01	1.04 ± 0.15 0.86 ± 0.02	1.97 ± 0.13 1.39 ± 0.07	2.88 ± 0.70 2.18 ± 0.04	0.22 ± 0.01 0.17 ± 0.00	2.12 ± 0.07 1.79 ± 0.02	0.98 ± 0.98 0.73 ± 0.76	1.00 / 0.98 1.00 / 0.96
	128 D 256 D	0.32 ± 0.00 0.32 ± 0.00	0.15 ± 0.00 0.15 ± 0.00	0.01 ± 0.00 0.01 ± 0.00	0.12 ± 0.00 0.12 ± 0.00	0.10 ± 0.00 0.10 ± 0.00	0.79 ± 0.02 0.77 ± 0.02	1.19 ± 0.02 1.16 ± 0.00	2.00 ± 0.01 1.94 ± 0.00	0.14 ± 0.01 0.13 ± 0.00	1.24 ± 0.04 0.96 ± 0.01	0.61 ± 0.65 0.56 ± 0.61	1.00 / 0.92 1.00 / 0.84
	16 F	0.66 ± 0.07	0.19 ± 0.01	0.03 ± 0.01	0.25 ± 0.02 0.15 ± 0.01	0.29 ± 0.02	0.99 ± 0.07	2.50 ± 0.51	2.63 ± 0.03	0.24 ± 0.02	2.21 ± 0.08	1.00 ± 1.01	1.00/0.99
100 full (F)	32 F 64 F	0.40 ± 0.00 0.34 ± 0.01	0.17 ± 0.00 0.15 ± 0.00	0.01 ± 0.00 0.01 ± 0.00	0.13 ± 0.01 0.12 ± 0.00	0.13 ± 0.01 0.11 ± 0.00	0.85 ± 0.02 0.79 ± 0.01	1.53 ± 0.12 1.23 ± 0.07	2.17 ± 0.06 1.98 ± 0.01	0.15 ± 0.01 0.15 ± 0.00	1.93 ± 0.04 1.26 ± 0.01	0.75 ± 0.80 0.61 ± 0.65	0.99/0.97 0.97/0.93
	128 F	0.32 ± 0.00	0.15 ± 0.00	0.01 ± 0.00	0.12 ± 0.00	0.10 ± 0.00	0.77 ± 0.02	1.16 ± 0.01	1.94 ± 0.00	0.13 ± 0.00	0.99 ± 0.01	0.57 ± 0.61	0.88/0.80
	256 F 16 C 5	0.33 ± 0.00 0.34 ± 0.01	0.16 ± 0.00 0.15 ± 0.00	0.00 ± 0.00 0.01 ± 0.00	0.12 ± 0.00 0.14 ± 0.01	0.10 ± 0.00 0.11 ± 0.00	0.76 ± 0.02 0.79 ± 0.02	1.15 ± 0.01 0.97 ± 0.27	1.93 ± 0.00 1.97 ± 0.00	0.14 ± 0.00 0.13 ± 0.00	0.86 ± 0.00 0.77 ± 0.25	0.56 ± 0.60 0.54 ± 0.58	0.50/0.34
100 w/clusters (C)	16 C 7	0.34 ± 0.01	0.15 ± 0.00	0.01 ± 0.00	0.13 ± 0.00	0.10 ± 0.00	0.78 ± 0.02	0.96 ± 0.27	1.96 ± 0.00	0.14 ± 0.00	0.74 ± 0.22	0.53 ± 0.57	1.00/0.93
	16 D	2.95 ± 0.28	0.73 ± 0.29	0.27 ± 0.09	0.67 ± 0.28	0.52 ± 0.07	2.06 ± 0.30	4.85 ± 0.31	3.94 ± 0.42	0.50 ± 0.05	2.50 ± 0.03	1.94 ± 1.59	1.00 / 1.00
500 diagonal (D)	32 D 64 D	2.33 ± 0.30 1.67 ± 0.18	0.62 ± 0.17 0.43 ± 0.16	0.24 ± 0.05 0.13 ± 0.04	0.50 ± 0.16 0.29 ± 0.02	0.37 ± 0.07 0.23 ± 0.02	1.86 ± 0.25 1.32 ± 0.28	4.73 ± 0.35 3.99 ± 0.36	3.81 ± 0.59 3.41 ± 0.28	0.39 ± 0.04 0.32 ± 0.05	2.46 ± 0.05 2.35 ± 0.11	1.77 ± 1.57 1.45 ± 1.39	1.00 / 1.00 1.00 / 0.99
500 unagonai (D)	128 D	1.12 ± 0.02	0.23 ± 0.00	0.04 ± 0.03	0.21 ± 0.04	0.22 ± 0.03	1.08 ± 0.06	3.05 ± 0.87	3.09 ± 0.37	0.26 ± 0.03	2.31 ± 0.04	1.19 ± 1.21	1.00 / 0.98
	256 D	0.54 ± 0.03	0.18 ± 0.01	0.01 ± 0.00	0.16 ± 0.01	0.15 ± 0.01	0.92 ± 0.08	2.42 ± 0.14	2.51 ± 0.13	0.19 ± 0.01	2.09 ± 0.02	0.94 ± 0.99	1.00 / 0.97
	16 F 32 F	2.14 ± 0.06 1.17 ± 0.07	0.70 ± 0.04 0.48 ± 0.03	0.28 ± 0.00 0.08 ± 0.04	0.27 ± 0.01 0.21 ± 0.01	0.21 ± 0.00 0.17 ± 0.00	1.14 ± 0.04 0.99 ± 0.04	3.06 ± 0.27 2.69 ± 0.10	2.71 ± 0.01 2.47 ± 0.02	0.34 ± 0.01 0.25 ± 0.02	2.21 ± 0.01 2.11 ± 0.04	1.33 ± 1.09 1.08 ± 0.99	1.00/1.00 0.99/0.99
500 full (F)	64 F	0.51 ± 0.03	0.21 ± 0.00	0.02 ± 0.00	0.17 ± 0.01	0.14 ± 0.00	0.88 ± 0.04	2.19 ± 0.14	2.34 ± 0.03	0.20 ± 0.00	1.97 ± 0.02	0.89 ± 0.91	0.97/0.96
	128 F 256 F	0.39 ± 0.01 0.32 ± 0.01	0.16 ± 0.00 0.15 ± 0.00	0.01 ± 0.00 0.01 ± 0.00	0.13 ± 0.00 0.12 ± 0.00	0.11 ± 0.00 0.10 ± 0.00	0.81 ± 0.03 0.77 ± 0.01	1.42 ± 0.07 1.18 ± 0.04	2.03 ± 0.01 1.96 ± 0.00	0.16 ± 0.00 0.14 ± 0.01	1.71 ± 0.01 1.25 ± 0.00	0.71 ± 0.74 0.61 ± 0.65	0.88/0.86 0.50/0.47
	16 C 7	0.40	0.18	0.01	0.15	0.13	0.90	2.03	2.21	0.16	1.50	0.77	1.00/0.98
500 (1) (2)	16 C 10	0.36	0.16	0.01	0.14	0.13	0.87	2.19	2.04	0.15	1.38	0.74	1.00/0.98
500 w/clusters (C)	16 C 25 64 C 5	0.32 0.36	0.16 0.16	0.01 0.01	0.13 0.12	0.10 0.10	0.81 0.80	1.28 1.17	1.96 1.98	0.12 0.14	1.07 1.17	0.60 0.60	1.00/0.95 0.97/0.93
	64 C 7	0.34	0.15	0.01	0.12	0.10	0.79	1.14	1.96	0.13	1.08	0.58	0.97/0.91
1000 w/clusters (C)	16 C 25	0.37	0.16	0.01	0.13	0.13	0.86	2.12	2.04	0.14	1.35	0.73	1.00/0.97

Table 12: Absolute In-Distribution test loss for various tasks and methods

		<u> </u>				Ta	sks						
Model Type	Method Type	task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	Average	Para. Saved
	base lora	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 1.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c c} 0.10 \pm 0.30 \\ 100.00 \pm 0.00 \end{array}$	1.00 / 1.00 0.00 / 0.00
TIES	10 50 100 500	$\begin{array}{c} 41.00 \pm 0.00 \\ 24.00 \pm 0.00 \\ 22.00 \pm 0.00 \\ 8.00 \pm 0.00 \end{array}$	$\begin{array}{c} 53.67 \pm 0.58 \\ 38.67 \pm 0.58 \\ 38.00 \pm 0.00 \\ 25.00 \pm 0.00 \end{array}$	$44.33 \pm 4.04 \\ 17.67 \pm 4.62 \\ 18.67 \pm 4.62 \\ 1.00 \pm 0.00$	$ \begin{array}{c} 10.33 \pm 0.58 \\ 2.00 \pm 0.00 \\ 1.00 \pm 1.73 \\ 0.00 \pm 0.00 \\ \end{array} $	$46.33 \pm 4.04 \\ 56.33 \pm 0.58 \\ 51.67 \pm 4.62 \\ 59.00 \pm 0.00$	$\begin{array}{c} 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 1.00 \pm 0.00 \\ 0.00 \pm 0.00 \end{array}$	$\begin{array}{c} 8.00 \pm 0.00 \\ 8.00 \pm 0.00 \\ 8.00 \pm 0.00 \\ 3.00 \pm 0.00 \end{array}$	$\begin{array}{c} 8.00 \pm 0.00 \\ 8.00 \pm 0.00 \\ 7.33 \pm 0.58 \\ 6.00 \pm 0.00 \end{array}$	$76.67 \pm 1.15 \\ 29.67 \pm 2.89 \\ 2.00 \pm 0.00 \\ 2.00 \pm 0.00$	$\begin{array}{c} 1.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \end{array}$	$\begin{array}{c} 29.03 \pm 25.69 \\ 18.53 \pm 18.07 \\ 14.97 \pm 17.20 \\ 9.90 \pm 18.12 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 1.00 1.00 / 1.00
SVD	SVD 2 SVD 4 SVD 8 SVD 16	$\begin{array}{c} 88.33 \pm 0.65 \\ 93.00 \pm 0.00 \\ 98.89 \pm 0.60 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 91.91 \pm 0.94 \\ 96.64 \pm 0.50 \\ 98.55 \pm 0.52 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 97.25 \pm 0.45 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 92.83 \pm 0.39 \\ 96.75 \pm 0.87 \\ 99.42 \pm 0.51 \\ 100.00 \pm 0.00 \end{array}$	$76.50 \pm 1.51 88.83 \pm 1.53 93.44 \pm 0.73 99.67 \pm 0.50$	66.00 ± 1.41 90.67 ± 1.23 97.22 ± 0.44 99.50 ± 0.55	58.08 ± 1.16 72.17 ± 0.58 82.78 ± 1.64 99.67 ± 0.50	$\begin{array}{c} 98.67 \pm 0.49 \\ 98.67 \pm 0.49 \\ 99.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 5.83 \pm 0.94 \\ 16.67 \pm 1.78 \\ 60.00 \pm 0.87 \\ 98.11 \pm 0.78 \end{array}$	77.42 ± 27.69 85.24 ± 24.39 93.70 ± 11.59 99.69 ± 0.68	0.88 / 0.88 0.75 / 0.75 0.50 / 0.50 0.00 / 0.00
10 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 83.33 \pm 1.53 \\ 93.00 \pm 1.00 \\ 99.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 88.33 \pm 0.58 \\ 95.33 \pm 0.58 \\ 97.00 \pm 1.00 \\ 99.33 \pm 0.58 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 97.00 \pm 2.00 \\ 98.00 \pm 1.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 88.33 \pm 1.15 \\ 93.67 \pm 1.53 \\ 98.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 57.00 \pm 1.00 \\ 80.67 \pm 2.31 \\ 90.67 \pm 1.53 \\ 96.67 \pm 1.53 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 48.67 \pm 3.21 \\ 78.67 \pm 1.15 \\ 95.33 \pm 1.15 \\ 98.33 \pm 1.15 \\ 100.00 \pm 0.00 \end{array}$	50.67 ± 4.93 68.00 ± 1.73 79.67 ± 1.53 95.67 ± 2.31 99.33 ± 1.15	$\begin{array}{c} 97.67 \pm 1.53 \\ 98.33 \pm 0.58 \\ 99.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 5.33 \pm 1.15 \\ 14.67 \pm 2.52 \\ 55.00 \pm 4.36 \\ 91.67 \pm 3.51 \\ 95.00 \pm 1.00 \end{array}$	$\begin{array}{c} 71.63 \pm 29.53 \\ 82.03 \pm 24.99 \\ 91.37 \pm 13.78 \\ 98.17 \pm 2.94 \\ 99.43 \pm 1.57 \end{array}$	1.00 / 0.90 1.00 / 0.80 1.00 / 0.60 1.00 / 0.20 1.00 / -0.60
10 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 83.00 \pm 2.00 \\ 91.33 \pm 0.58 \\ 99.00 \pm 0.00 \\ 99.67 \pm 0.58 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 93.00 \pm 1.00 \\ 96.00 \pm 1.00 \\ 97.33 \pm 0.58 \\ 99.33 \pm 0.58 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 98.33 \pm 0.58 \\ 98.33 \pm 0.58 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 91.67 \pm 0.58 \\ 94.33 \pm 0.58 \\ 99.33 \pm 1.15 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$64.33 \pm 3.21 \\ 84.00 \pm 2.00 \\ 91.33 \pm 1.53 \\ 97.67 \pm 1.15 \\ 99.67 \pm 0.58$	$\begin{array}{c} 59.33 \pm 1.15 \\ 83.00 \pm 1.73 \\ 96.33 \pm 0.58 \\ 100.00 \pm 0.00 \\ 99.67 \pm 0.58 \end{array}$	$\begin{array}{c} 52.67 \pm 1.53 \\ 70.33 \pm 1.53 \\ 81.67 \pm 2.31 \\ 95.67 \pm 1.15 \\ 99.67 \pm 0.58 \end{array}$	$\begin{array}{c} 98.33 \pm 0.58 \\ 99.00 \pm 1.00 \\ 99.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 6.33 \pm 1.15 \\ 22.00 \pm 2.65 \\ 58.33 \pm 1.53 \\ 91.00 \pm 1.00 \\ 98.00 \pm 1.00 \end{array}$	$\begin{array}{c} 74.70 \pm 28.71 \\ 83.83 \pm 22.82 \\ 92.23 \pm 12.76 \\ 98.33 \pm 2.89 \\ 99.70 \pm 0.70 \end{array}$	1.00/0.90 0.99/0.79 0.97/0.57 0.88/0.07 0.50/-1.10
50 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 52.67 \pm 4.51 \\ 69.67 \pm 3.21 \\ 79.67 \pm 2.52 \\ 90.00 \pm 1.00 \\ 94.67 \pm 0.58 \end{array}$	$86.67 \pm 3.06 \\ 88.67 \pm 1.53 \\ 91.00 \pm 1.00 \\ 91.33 \pm 0.58 \\ 96.33 \pm 0.58$	$\begin{array}{c} 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 85.00 \pm 3.46 \\ 95.00 \pm 2.00 \\ 97.67 \pm 0.58 \\ 98.33 \pm 0.58 \\ 99.67 \pm 0.58 \end{array}$	$65.33 \pm 3.79 \\ 80.00 \pm 3.00 \\ 88.00 \pm 1.00 \\ 90.67 \pm 2.08 \\ 96.33 \pm 1.15$	$\begin{array}{c} 25.33 \pm 5.03 \\ 36.67 \pm 5.13 \\ 52.00 \pm 1.00 \\ 73.67 \pm 2.08 \\ 87.33 \pm 0.58 \end{array}$	$ \begin{array}{c} 10.00 \pm 1.00 \\ 17.00 \pm 2.65 \\ 36.67 \pm 5.69 \\ 63.67 \pm 1.53 \\ 87.00 \pm 2.65 \end{array} $	$ \begin{array}{c} 10.67 \pm 10.26 \\ 26.33 \pm 5.03 \\ 41.33 \pm 1.15 \\ 56.33 \pm 0.58 \\ 71.67 \pm 1.53 \end{array} $	$\begin{array}{c} 81.00 \pm 6.56 \\ 95.00 \pm 2.00 \\ 96.00 \pm 1.00 \\ 98.00 \pm 0.00 \\ 99.67 \pm 0.58 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.33 \pm 0.58 \\ 7.33 \pm 1.15 \\ 31.67 \pm 1.15 \end{array}$	$\begin{array}{c} 51.67 \pm 36.18 \\ 60.83 \pm 36.02 \\ 68.27 \pm 32.66 \\ 76.93 \pm 27.79 \\ 86.43 \pm 20.41 \end{array}$	1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84 1.00 / 0.68
50 full (F)	16 F 32 F 64 F 128 F 256 F	$61.67 \pm 3.06 \\ 71.00 \pm 1.00 \\ 81.67 \pm 0.58 \\ 91.00 \pm 1.00 \\ 97.00 \pm 0.00$	$\begin{array}{c} 89.67 \pm 1.15 \\ 89.00 \pm 1.73 \\ 93.67 \pm 1.15 \\ 94.33 \pm 0.58 \\ 98.00 \pm 0.00 \end{array}$	$\begin{array}{c} 99.67 \pm 0.58 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 90.67 \pm 2.52 \\ 98.00 \pm 0.00 \\ 98.33 \pm 0.58 \\ 99.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	78.33 ± 3.51 85.00 ± 1.00 90.67 ± 2.08 93.33 ± 1.53 99.67 ± 0.58	$\begin{array}{c} 34.00 \pm 1.00 \\ 47.00 \pm 1.73 \\ 61.67 \pm 1.53 \\ 81.67 \pm 0.58 \\ 92.00 \pm 0.00 \end{array}$	$\begin{array}{c} 7.00 \pm 3.46 \\ 29.00 \pm 3.00 \\ 54.33 \pm 1.53 \\ 75.00 \pm 1.73 \\ 94.33 \pm 1.15 \end{array}$	25.00 ± 6.24 35.00 ± 2.00 51.33 ± 1.15 67.67 ± 2.08 79.67 ± 1.53	$\begin{array}{c} 90.00 \pm 1.00 \\ 98.00 \pm 1.00 \\ 98.33 \pm 0.58 \\ 98.67 \pm 0.58 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 3.33 \pm 0.58 \\ 16.67 \pm 0.58 \\ 57.33 \pm 2.52 \end{array}$	$\begin{array}{c} 57.60 \pm 36.59 \\ 65.20 \pm 34.00 \\ 73.33 \pm 29.89 \\ 81.73 \pm 24.46 \\ 91.70 \pm 13.11 \end{array}$	1.00/0.98 0.99/0.95 0.97/0.89 0.88/0.72 0.50/0.18
100 diagonal (D)	16 D 32 D 64 D 128 D 256 D	33.00 ± 8.19 51.00 ± 7.81 68.00 ± 2.65 82.00 ± 2.00 90.00 ± 1.00	79.33 ± 5.69 90.00 ± 1.00 87.33 ± 1.53 90.00 ± 2.00 93.00 ± 2.00	$\begin{array}{c} 89.33 \pm 5.51 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	80.00 ± 3.61 88.00 ± 7.00 94.33 ± 4.04 97.33 ± 0.58 97.67 ± 0.58	35.33 ± 6.03 58.67 ± 11.68 80.33 ± 2.08 85.33 ± 0.58 91.67 ± 0.58	$\begin{array}{c} 4.00 \pm 1.73 \\ 17.67 \pm 10.26 \\ 38.00 \pm 3.00 \\ 55.33 \pm 2.08 \\ 71.67 \pm 5.13 \end{array}$	3.00 ± 1.00 7.67 ± 2.89 19.67 ± 4.51 34.33 ± 3.79 59.67 ± 1.53	0.00 ± 0.00 9.33 ± 12.86 28.33 ± 1.53 36.67 ± 2.52 58.00 ± 0.00	71.33 ± 4.51 86.33 ± 2.52 92.67 ± 1.15 94.67 ± 0.58 97.67 ± 0.58	0.00 ± 0.00 0.00 ± 0.00 0.33 ± 0.58 0.00 ± 0.00 4.00 ± 1.00	39.53 ± 36.15 50.87 ± 38.43 60.90 ± 34.91 67.57 ± 32.95 76.33 ± 28.88	1.00 / 0.99 1.00 / 0.98 1.00 / 0.96 1.00 / 0.92 1.00 / 0.84
100 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 49.00 \pm 2.00 \\ 65.00 \pm 3.46 \\ 72.33 \pm 0.58 \\ 84.33 \pm 1.53 \\ 91.67 \pm 1.15 \end{array}$	89.67 ± 3.21 90.33 ± 1.53 89.67 ± 1.53 92.33 ± 1.53 96.67 ± 0.58	$\begin{array}{c} 97.00 \pm 3.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$\begin{array}{c} 84.33 \pm 3.06 \\ 96.33 \pm 1.53 \\ 97.67 \pm 0.58 \\ 98.00 \pm 0.00 \\ 100.00 \pm 0.00 \end{array}$	$65.33 \pm 2.52 \\ 80.00 \pm 2.65 \\ 86.00 \pm 1.00 \\ 91.33 \pm 0.58 \\ 94.33 \pm 0.58$	$\begin{array}{c} 20.67 \pm 8.33 \\ 41.33 \pm 3.21 \\ 53.00 \pm 1.00 \\ 68.67 \pm 0.58 \\ 84.67 \pm 0.58 \end{array}$	6.33 ± 2.08 16.00 ± 0.00 35.33 ± 1.53 56.00 ± 1.00 78.00 ± 0.00	8.33 ± 4.73 29.33 ± 2.08 38.00 ± 1.73 57.67 ± 1.15 69.67 ± 0.58	$\begin{array}{c} 81.33 \pm 2.08 \\ 92.00 \pm 2.65 \\ 94.67 \pm 0.58 \\ 99.00 \pm 0.00 \\ 99.00 \pm 0.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 5.33 \pm 0.58 \\ 22.00 \pm 1.00 \end{array}$	50.20 ± 37.06 61.03 ± 35.43 66.67 ± 32.54 75.27 ± 28.67 83.60 ± 23.08	1.00/0.99 0.99/0.97 0.97/0.93 0.88/0.80 0.50/0.34
100 w/clusters (C)	16 C 5 16 C 7	74.67 ± 0.94 77.67 ± 0.47	91.00 ± 0.82 90.33 ± 1.25	$100.00 \pm 0.00 \\ 100.00 \pm 0.00$	96.67 ± 1.25 97.33 ± 0.94	87.67 ± 1.70 90.33 ± 2.05	53.67 ± 2.05 58.67 ± 0.94	40.67 ± 2.87 49.00 ± 2.16	41.00 ± 4.55 49.00 ± 0.82	97.67 ± 1.25 97.67 ± 0.47	0.67 ± 0.94 3.67 ± 1.25	68.37 ± 31.46 71.37 ± 29.48	1.00/0.95 1.00/0.93
500 diagonal (D)	16 D 32 D 64 D 128 D 256 D	$\begin{array}{c} 8.00 \pm 3.61 \\ 14.33 \pm 11.02 \\ 25.67 \pm 1.15 \\ 38.33 \pm 3.21 \\ 53.33 \pm 0.58 \end{array}$	$\begin{array}{c} 51.50 \pm 3.54 \\ 52.50 \pm 9.19 \\ 62.50 \pm 12.02 \\ 85.50 \pm 2.12 \\ 91.00 \pm 2.83 \end{array}$	79.67 ± 4.93 80.67 ± 2.08 87.33 ± 4.04 96.00 ± 3.00 100.00 ± 0.00	$\begin{array}{c} 28.00 \pm 42.44 \\ 43.00 \pm 31.48 \\ 78.33 \pm 1.15 \\ 81.33 \pm 2.31 \\ 89.00 \pm 2.65 \end{array}$	$\begin{array}{c} 56.67 \pm 2.89 \\ 60.67 \pm 0.58 \\ 65.33 \pm 3.06 \\ 65.67 \pm 1.15 \\ 76.00 \pm 2.00 \end{array}$	$\begin{array}{c} 0.67 \pm 1.15 \\ 0.67 \pm 1.15 \\ 5.33 \pm 3.51 \\ 11.67 \pm 4.73 \\ 20.67 \pm 6.81 \end{array}$	0.33 ± 0.58 1.67 ± 1.15 3.67 ± 1.15 5.33 ± 2.08 6.00 ± 1.73	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 1.00 \pm 1.73 \\ 2.00 \pm 1.00 \\ 12.00 \pm 8.00 \end{array}$	71.67 ± 6.03 74.33 ± 4.04 76.67 ± 2.31 80.00 ± 5.00 86.33 ± 2.31	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \end{array}$	$\begin{array}{c} 28.90 \pm 33.54 \\ 32.10 \pm 33.43 \\ 39.83 \pm 35.80 \\ 45.24 \pm 37.78 \\ 52.14 \pm 38.64 \end{array}$	1.00 / 1.00 1.00 / 1.00 1.00 / 0.99 1.00 / 0.98 1.00 / 0.97
500 full (F)	16 F 32 F 64 F 128 F 256 F	$\begin{array}{c} 8.33 \pm 2.08 \\ 33.67 \pm 4.16 \\ 56.00 \pm 2.65 \\ 69.33 \pm 0.58 \\ 79.67 \pm 0.58 \end{array}$	$41.00 \pm 5.66 \\ 51.00 \pm 1.41 \\ 85.50 \pm 0.71 \\ 88.50 \pm 0.71 \\ 89.50 \pm 0.71$	$76.67 \pm 0.58 \\ 92.67 \pm 1.53 \\ 94.33 \pm 0.58 \\ 99.00 \pm 0.00 \\ 100.00 \pm 0.00$	$78.00 \pm 0.00 \\ 77.00 \pm 1.73 \\ 89.33 \pm 2.89 \\ 96.33 \pm 1.53 \\ 97.67 \pm 0.58$	$72.67 \pm 0.58 \\ 75.00 \pm 2.00 \\ 74.33 \pm 1.15 \\ 80.33 \pm 1.15 \\ 87.33 \pm 0.58$	$\begin{array}{c} 6.00 \pm 0.00 \\ 14.33 \pm 1.53 \\ 36.33 \pm 1.15 \\ 45.00 \pm 2.00 \\ 57.00 \pm 1.00 \end{array}$	$\begin{array}{c} 5.67 \pm 0.58 \\ 8.00 \pm 0.00 \\ 9.00 \pm 1.00 \\ 16.33 \pm 0.58 \\ 35.00 \pm 1.00 \end{array}$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \\ 2.67 \pm 1.15 \\ 31.00 \pm 1.73 \\ 42.00 \pm 1.00 \end{array}$	$78.00 \pm 1.00 \\ 80.67 \pm 1.53 \\ 84.00 \pm 1.00 \\ 92.00 \pm 0.00 \\ 95.00 \pm 1.00$	$\begin{array}{c} 0.00 \pm 0.00 \\ 0.00 \pm 0.00 \end{array}$	36.48 ± 35.46 42.97 ± 35.80 52.03 ± 36.92 60.86 ± 35.07 67.59 ± 32.67	1.00/1.00 0.99/0.99 0.97/0.96 0.88/0.86 0.50/0.47
500 w/clusters (C)	16 C 7 16 C 10 16 C 25 64 C 5 64 C 7	63.00 69.00 79.00 77.00 76.00	90.00 93.00 90.00 88.00 90.00	99.00 100.00 100.00 100.00 100.00	96.00 98.00 97.00 98.00 97.00	78.00 81.00 88.00 89.00 89.00	31.00 34.00 53.00 56.00 60.00	9.00 8.00 38.00 39.00 48.00	15.00 33.00 48.00 42.00 49.00	89.00 95.00 98.00 99.00 99.00	1.00 1.00 0.00 0.00 3.00	57.10 61.20 69.10 68.80 71.10	1.00/0.98 1.00/0.98 1.00/0.95 0.97/0.93 0.97/0.91
1000 w/clusters (C)	16 C 25	73.00	90.00	100.00	98.00	77.00	39.00	8.00	34.00	96.00	1.00	61.60	1.00/0.97

Table 13: Absolute In-Distribution agreement for various tasks and methods

Model Type	Method Type	Tasks										
I		task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	Average
	SVD 2	0.29 ± 0.00	0.43 ± 0.00	0.31 ± 0.00	0.40 ± 0.00	0.38 ± 0.00	0.31 ± 0.00	0.37 ± 0.00	0.31 ± 0.00	0.42 ± 0.00	0.30 ± 0.00	0.35 ± 0.05
SVD	SVD 4	0.16 ± 0.00	0.24 ± 0.00	0.16 ± 0.00	0.25 ± 0.00	0.23 ± 0.00	0.17 ± 0.00	0.22 ± 0.00	0.16 ± 0.00	0.25 ± 0.00	0.16 ± 0.00	0.20 ± 0.04
	SVD 8	0.06 ± 0.00	0.09 ± 0.00	0.06 ± 0.00	0.11 ± 0.00	0.10 ± 0.00	0.07 ± 0.00	0.09 ± 0.00	0.06 ± 0.00	0.11 ± 0.00	0.06 ± 0.00	0.08 ± 0.02
	16 D	0.37 ± 0.02	0.51 ± 0.02	0.36 ± 0.01	0.57 ± 0.02	0.55 ± 0.00	0.39 ± 0.02	0.49 ± 0.01	0.36 ± 0.02	0.53 ± 0.03	0.39 ± 0.01	0.45 ± 0.08
10 diagonal (D)	32 D 64 D	0.21 ± 0.01 0.10 ± 0.00	0.28 ± 0.00 0.11 ± 0.01	0.20 ± 0.01 0.09 ± 0.00	0.35 ± 0.00 0.18 ± 0.01	0.33 ± 0.01 0.18 ± 0.00	0.22 ± 0.01 0.10 ± 0.00	0.31 ± 0.01 0.15 ± 0.01	0.20 ± 0.01 0.09 ± 0.00	0.32 ± 0.01 0.14 ± 0.00	0.22 ± 0.00 0.09 ± 0.00	0.26 ± 0.06 0.12 ± 0.04
10 diagonai (D)	128 D	0.10 ± 0.00 0.02 ± 0.00	0.01 ± 0.00	0.02 ± 0.00 0.02 ± 0.00	0.03 ± 0.00	0.16 ± 0.00 0.04 ± 0.00	0.10 ± 0.00 0.02 ± 0.00	0.03 ± 0.00	0.02 ± 0.00	0.14 ± 0.00 0.02 ± 0.00	0.02 ± 0.00	0.12 ± 0.04 0.03 ± 0.01
	256 D	0.00 ± 0.00										
	16 F	0.35 ± 0.00	0.46 ± 0.00	0.34 ± 0.00	0.51 ± 0.00	0.47 ± 0.01	0.36 ± 0.01	0.45 ± 0.01	0.35 ± 0.01	0.49 ± 0.00	0.35 ± 0.01	0.41 ± 0.06
10 full (F)	32 F 64 F	0.20 ± 0.00 0.10 ± 0.00	0.24 ± 0.00 0.10 ± 0.00	0.20 ± 0.00 0.09 ± 0.00	0.30 ± 0.00 0.13 ± 0.00	0.29 ± 0.00 0.13 ± 0.00	0.22 ± 0.00 0.10 ± 0.00	0.27 ± 0.00 0.12 ± 0.00	0.20 ± 0.00 0.09 ± 0.00	0.27 ± 0.00 0.12 ± 0.00	0.21 ± 0.00 0.10 ± 0.00	0.24 ± 0.04 0.11 ± 0.02
10 Iuli (F)	128 F	0.10 ± 0.00 0.02 ± 0.00	0.10 ± 0.00 0.02 ± 0.00	0.09 ± 0.00 0.02 ± 0.00	0.13 ± 0.00 0.01 ± 0.00	0.13 ± 0.00 0.02 ± 0.00	0.10 ± 0.00 0.02 ± 0.00	0.12 ± 0.00 0.02 ± 0.00	0.09 ± 0.00 0.02 ± 0.00	0.12 ± 0.00 0.01 ± 0.00	0.10 ± 0.00 0.02 ± 0.00	0.11 ± 0.02 0.02 ± 0.00
	256 F	0.00 ± 0.00										
	16 D	0.66 ± 0.01	0.69 ± 0.01	0.88 ± 0.01	0.76 ± 0.03	0.95 ± 0.02	0.91 ± 0.01	0.83 ± 0.02	0.88 ± 0.03	0.72 ± 0.02	0.88 ± 0.02	0.82 ± 0.10
	32 D	0.50 ± 0.01	0.52 ± 0.02	0.73 ± 0.01	0.58 ± 0.03	0.88 ± 0.03	0.79 ± 0.03	0.72 ± 0.01	0.75 ± 0.01	0.57 ± 0.02	0.75 ± 0.01	0.68 ± 0.12
50 diagonal (D)	64 D 128 D	0.34 ± 0.01 0.21 ± 0.01	0.37 ± 0.01 0.22 ± 0.01	0.52 ± 0.00 0.31 ± 0.00	0.38 ± 0.01 0.22 ± 0.00	0.71 ± 0.02 0.51 ± 0.01	0.58 ± 0.01 0.42 ± 0.01	0.54 ± 0.00 0.38 ± 0.00	0.56 ± 0.00 0.39 ± 0.00	0.44 ± 0.01 0.27 ± 0.00	0.58 ± 0.01 0.40 ± 0.00	0.50 ± 0.11 0.33 ± 0.10
	256 D	0.21 ± 0.01 0.10 ± 0.00	0.22 ± 0.01 0.12 ± 0.00	0.31 ± 0.00 0.16 ± 0.00	0.22 ± 0.00 0.10 ± 0.00	0.31 ± 0.01 0.29 ± 0.01	0.42 ± 0.01 0.21 ± 0.00	0.38 ± 0.00 0.19 ± 0.00	0.39 ± 0.00 0.23 ± 0.01	0.27 ± 0.00 0.15 ± 0.00	0.40 ± 0.00 0.20 ± 0.00	0.33 ± 0.10 0.18 ± 0.06
	16 F	0.57 ± 0.01	0.60 ± 0.01	0.86 ± 0.01	0.71 ± 0.02	0.95 ± 0.01	0.88 ± 0.01	0.81 ± 0.00	0.83 ± 0.01	0.67 ± 0.01	0.86 ± 0.01	0.78 ± 0.12
	32 F	0.47 ± 0.01	$0.48 \pm {\scriptstyle 0.01}$	0.71 ± 0.00	0.55 ± 0.01	$0.78 \pm \scriptstyle{0.01}$	0.69 ± 0.01	0.69 ± 0.00	0.65 ± 0.01	0.53 ± 0.01	0.71 ± 0.00	0.63 ± 0.11
50 full (F)	64 F	0.33 ± 0.00	0.35 ± 0.00	0.45 ± 0.00	0.36 ± 0.00	0.56 ± 0.00	0.50 ± 0.01	0.47 ± 0.00	0.49 ± 0.00	0.39 ± 0.01	0.49 ± 0.00	0.44 ± 0.08
	128 F 256 F	0.19 ± 0.00 0.09 ± 0.00	0.21 ± 0.00 0.10 ± 0.00	0.25 ± 0.00 0.10 ± 0.00	0.19 ± 0.00 0.08 ± 0.00	0.35 ± 0.00 0.16 ± 0.00	0.30 ± 0.00 0.13 ± 0.00	0.28 ± 0.00 0.12 ± 0.00	0.31 ± 0.00 0.15 ± 0.00	0.24 ± 0.00 0.11 ± 0.00	0.30 ± 0.00 0.13 ± 0.00	0.26 ± 0.05 0.12 ± 0.02
1	16 D	0.90 ± 0.01	0.85 ± 0.01	0.87 ± 0.03	0.88 ± 0.02	0.68 ± 0.02	0.91 ± 0.01	0.97 ± 0.01	0.98 ± 0.01	0.96 ± 0.01	1.00 ± 0.00	0.90 ± 0.09
	32 D	0.90 ± 0.01 0.83 ± 0.02	0.03 ± 0.01 0.77 ± 0.00	0.67 ± 0.03 0.77 ± 0.01	0.08 ± 0.02 0.78 ± 0.00	0.55 ± 0.02 0.55 ± 0.02	0.71 ± 0.01 0.79 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.87 ± 0.00	0.98 ± 0.01	0.82 ± 0.03
100 diagonal (D)	64 D	0.67 ± 0.01	0.63 ± 0.00	0.59 ± 0.02	0.63 ± 0.01	0.40 ± 0.00	0.62 ± 0.01	0.86 ± 0.02	0.82 ± 0.02	0.71 ± 0.03	0.93 ± 0.00	0.68 ± 0.15
	128 D 256 D	0.49 ± 0.01	0.47 ± 0.00	0.42 ± 0.01	0.45 ± 0.00	0.27 ± 0.02	0.44 ± 0.01	0.73 ± 0.01	0.69 ± 0.02	0.59 ± 0.02	0.80 ± 0.02	0.53 ± 0.16
		0.32 ± 0.00	0.31 ± 0.00	0.26 ± 0.01	0.30 ± 0.00	0.15 ± 0.01	0.28 ± 0.00	0.51 ± 0.02	0.51 ± 0.02	0.40 ± 0.01	0.61 ± 0.01	0.36 ± 0.14
	16 F 32 F	0.88 ± 0.00 0.78 ± 0.00	0.82 ± 0.00 0.72 ± 0.00	0.84 ± 0.01 0.73 ± 0.00	0.86 ± 0.00 0.74 ± 0.00	0.67 ± 0.01 0.52 ± 0.00	0.88 ± 0.01 0.74 ± 0.01	0.99 ± 0.00 0.94 ± 0.01	0.96 ± 0.01 0.89 ± 0.00	0.91 ± 0.01 0.77 ± 0.02	1.00 ± 0.00 0.99 ± 0.00	0.88 ± 0.09 0.78 ± 0.13
100 full (F)	64 F	0.60 ± 0.00	0.57 ± 0.00	0.57 ± 0.00	0.57 ± 0.00	0.39 ± 0.00	0.56 ± 0.00	0.76 ± 0.00	0.73 ± 0.00	0.60 ± 0.00	0.83 ± 0.01	0.62 ± 0.12
	128 F	0.40 ± 0.00	0.38 ± 0.00	0.35 ± 0.00	0.37 ± 0.00	0.25 ± 0.00	0.37 ± 0.00	0.52 ± 0.00	0.54 ± 0.00	0.45 ± 0.00	0.60 ± 0.00	0.42 ± 0.10
	256 F	0.21 ± 0.00	0.20 ± 0.00	0.18 ± 0.00	0.19 ± 0.00	0.13 ± 0.00	0.19 ± 0.00	0.30 ± 0.00	0.34 ± 0.00	0.26 ± 0.00	0.38 ± 0.00	0.24 ± 0.08
100 w/clusters (C)	16 C 5 16 C 7	0.46 ± 0.00 0.41 ± 0.01	0.46 ± 0.00 0.42 ± 0.01	0.45 ± 0.00 0.39 ± 0.01	0.47 ± 0.01 0.43 ± 0.01	0.61 ± 0.01 0.51 ± 0.01	0.65 ± 0.01 0.56 ± 0.01	0.61 ± 0.00 0.53 ± 0.01	0.64 ± 0.02 0.55 ± 0.00	0.45 ± 0.00 0.42 ± 0.01	0.59 ± 0.01 0.54 ± 0.01	0.54 ± 0.08 0.48 ± 0.06
	16 D				1.00 ± 0.00				0.92 ± 0.00	ı	1	0.94 ± 0.08
	32 D	0.97 ± 0.00 0.96 ± 0.00	0.73 ± 0.00 0.70 ± 0.00	0.96 ± 0.00 0.92 ± 0.01	0.98 ± 0.00	0.99 ± 0.01 0.96 ± 0.01	0.96 ± 0.01 0.93 ± 0.01	0.90 ± 0.00 0.86 ± 0.00	0.92 ± 0.00 0.89 ± 0.00	1.00 ± 0.00 1.00 ± 0.00	1.00 ± 0.00 1.00 ± 0.00	0.94 ± 0.08 0.92 ± 0.09
500 diagonal (D)	64 D	0.90 ± 0.01	0.65 ± 0.00	0.86 ± 0.01	0.96 ± 0.02	0.90 ± 0.01	0.87 ± 0.01	0.81 ± 0.00	0.83 ± 0.01	0.99 ± 0.01	1.00 ± 0.00	0.88 ± 0.10
	128 D	0.82 ± 0.01	0.60 ± 0.00	0.76 ± 0.00	0.90 ± 0.02	0.83 ± 0.01	0.78 ± 0.02	0.74 ± 0.00	0.74 ± 0.01	0.97 ± 0.01	1.00 ± 0.00	0.81 ± 0.12
	256 D	0.59 ± 0.02	0.51 ± 0.00	0.56 ± 0.01	0.81 ± 0.02	0.70 ± 0.02	0.55 ± 0.01	0.57 ± 0.01	0.54 ± 0.01	0.91 ± 0.01	1.00 ± 0.01	0.67 ± 0.17
	16 F 32 F	0.94 ± 0.00 0.88 ± 0.00	0.67 ± 0.00 0.61 ± 0.00	0.88 ± 0.00 0.81 ± 0.00	1.00 ± 0.00 0.97 ± 0.01	0.98 ± 0.00 0.94 ± 0.00	0.90 ± 0.00 0.84 ± 0.00	0.82 ± 0.00 0.75 ± 0.00	0.83 ± 0.00 0.77 ± 0.00	1.00 ± 0.00 0.99 ± 0.00	1.00 ± 0.00 0.99 ± 0.00	0.90 ± 0.10 0.86 ± 0.12
500 full (F)	64 F	0.80 ± 0.00 0.80 ± 0.00	0.51 ± 0.00 0.55 ± 0.00	0.31 ± 0.00 0.72 ± 0.00	0.97 ± 0.01 0.86 ± 0.00	0.94 ± 0.00 0.82 ± 0.01	0.34 ± 0.00 0.76 ± 0.00	0.73 ± 0.00 0.67 ± 0.00	0.77 ± 0.00 0.70 ± 0.00	0.99 ± 0.00 0.94 ± 0.00	0.99 ± 0.00 0.99 ± 0.00	0.30 ± 0.12 0.78 ± 0.13
	128 F	0.64 ± 0.00	0.46 ± 0.00	0.60 ± 0.00	0.74 ± 0.00	0.65 ± 0.00	0.63 ± 0.00	0.56 ± 0.00	0.58 ± 0.00	0.85 ± 0.00	0.96 ± 0.00	0.67 ± 0.14
	256 F	0.43 ± 0.00	0.35 ± 0.00	0.44 ± 0.00	0.55 ± 0.00	0.49 ± 0.00	0.45 ± 0.00	0.40 ± 0.00	0.42 ± 0.00	0.67 ± 0.00	0.84 ± 0.00	0.50 ± 0.14
	16 C 7 16 C 10	0.68 0.61	0.70 0.65	0.64 0.61	0.72 0.66	0.85 0.84	0.90 0.86	0.93 0.88	0.92 0.84	0.71 0.62	0.83 0.76	0.79 0.73
500 w/clusters (C)	16 C 10 16 C 25	0.61	0.65	0.61	0.66	0.84	0.86	0.88	0.84	0.62	0.76	0.73
	64 C 5	0.49	0.49	0.45	0.51	0.64	0.66	0.62	0.67	0.50	0.65	0.57
	64 C 7	0.45	0.45	0.41	0.45	0.56	0.58	0.55	0.59	0.44	0.57	0.51
1000 w/clusters (C)	16 C 25	0.58	0.64	0.54	0.64	0.81	0.87	0.90	0.84	0.57	0.74	0.71

Table 14: Reconstruction error In-Distribution for various tasks and methods

Model Type	ype Method Type Tasks									. Average		
		task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	
	16 F	0.46 ± 0.01	0.63 ± 0.00	0.50 ± 0.00	0.55 ± 0.01	0.50 ± 0.00	0.49 ± 0.01	0.50 ± 0.01	0.50 ± 0.01	0.61 ± 0.01	0.47 ± 0.01	0.52 ± 0.06
	32 F	0.30 ± 0.01	0.37 ± 0.00	0.31 ± 0.00	0.35 ± 0.00	0.34 ± 0.00	0.31 ± 0.00	0.33 ± 0.00	0.31 ± 0.00	0.38 ± 0.00	0.30 ± 0.00	0.33 ± 0.03
10 full (F)	64 F	0.15 ± 0.00	0.15 ± 0.00	0.16 ± 0.00	0.17 ± 0.00	0.17 ± 0.00	0.16 ± 0.00	0.16 ± 0.00	0.16 ± 0.00	0.17 ± 0.00	0.15 ± 0.00	0.16 ± 0.01
	16 F	0.80 ± 0.02	0.82 ± 0.01	0.85 ± 0.01	0.90 ± 0.02	0.78 ± 0.01	0.95 ± 0.01	0.76 ± 0.01	0.75 ± 0.00	0.79 ± 0.01	0.82 ± 0.01	0.82 ± 0.06
	32 F	0.65 ± 0.01	0.67 ± 0.01	0.72 ± 0.01	0.76 ± 0.02	0.65 ± 0.01	0.82 ± 0.02	0.66 ± 0.01	0.65 ± 0.01	0.67 ± 0.02	0.69 ± 0.00	0.69 ± 0.06
50 full (F)	64 F	0.50 ± 0.01	0.52 ± 0.00	0.52 ± 0.00	0.55 ± 0.01	0.52 ± 0.01	0.62 ± 0.00	0.54 ± 0.01	0.51 ± 0.00	0.54 ± 0.01	0.57 ± 0.00	0.54 ± 0.03
	16 F	0.93 ± 0.02	0.90 ± 0.02	0.93 ± 0.01	0.91 ± 0.02	0.88 ± 0.03	0.98 ± 0.01	0.96 ± 0.01	0.78 ± 0.00	0.82 ± 0.00	0.93 ± 0.02	0.90 ± 0.06
	32 F	0.87 ± 0.01	0.81 ± 0.01	0.85 ± 0.02	0.80 ± 0.01	0.79 ± 0.02	0.91 ± 0.00	0.90 ± 0.01	0.74 ± 0.01	0.70 ± 0.02	0.85 ± 0.02	0.82 ± 0.07
100 full (F)	64 F	0.65 ± 0.04	0.69 ± 0.01	0.71 ± 0.01	0.67 ± 0.01	0.64 ± 0.01	0.76 ± 0.01	0.77 ± 0.01	0.67 ± 0.00	0.61 ± 0.00	0.75 ± 0.06	0.69 ± 0.06
	16 F	0.98 ± 0.04	0.98 ± 0.01	0.99 ± 0.01	1.00 ± 0.00	0.99 ± 0.00	0.96 ± 0.05	0.93 ± 0.10	0.94 ± 0.09	1.00 ± 0.00	0.99 ± 0.00	0.98 ± 0.05
	32 F	0.92 ± 0.07	0.84 ± 0.20	0.92 ± 0.10	0.98 ± 0.02	0.97 ± 0.02	0.89 ± 0.08	0.82 ± 0.13	0.84 ± 0.11	0.99 ± 0.00	0.99 ± 0.02	0.92 ± 0.10
500 full (F)	64 F	0.80 ± 0.00	0.67 ± 0.21	0.78 ± 0.11	0.90 ± 0.07	0.86 ± 0.08	0.76 ± 0.00	0.67 ± 0.00	0.70 ± 0.00	0.96 ± 0.03	0.99 ± 0.00	0.81 ± 0.13

Table 15: **Reconstruction error on random LoRAs** The error is larger in comparison to reconstructing trained (i.e., non-random) LoRAs in Table 14 for the corresponding compression methods.

Model Type	Method Type	Tasks						Average				
		task039	task190	task280	task290	task391	task442	task620	task1342	task1391	task1598	
	base	24.44	1.60	19.13	39.22	10.27	35.46	7.85	6.22	17.82	38.87	20.09
	lora	95.00	86.00	99.00	93.67	94.33	74.88	74.40	26.68	95.00	50.32	78.93
	32 F	97.00	90.00	99.00	93.33	94.67	74.09	72.13	27.83	94.00	50.71	79.28
10 full (F)	64 F	95.00	89.00	99.00	93.67	94.67	74.29	74.80	26.63	96.00	51.04	79.41
	32 F	96.00	88.00	99.00	93.67	92.33	72.30	75.97	29.89	94.00	45.68	78.68
50 full (F)	64 F	98.00	89.00	99.00	93.67	93.33	72.74	76.50	29.33	96.00	45.71	79.33
	32 F	92.10	83.00	99.00	93.67	92.00	71.09	63.29	27.87	88.00	42.36	75.24
100 full (F)	64 F	97.00	87.00	99.00	93.67	92.33	72.23	74.69	29.98	95.00	44.71	78.56
-	32 F	68.92	43.00	87.00	91.67	90.67	70.08	51.16	14.40	83.00	41.97	64.19
500 full (F)	64 F	93.50	78.00	91.00	92.33	90.33	72.55	57.49	15.44	85.00	42.31	71.80

Table 16: Performance with convergence In-Distribution Rouge-L

Table 17: Agreement Comparison. 100 LoRAs

	Agreement (%)
	83.015
S	100.000
Rank 8	87.032
Rank 16	88.908
Rank 32	91.545
Rank 64	94.659
Rank 8	87.686
Rank 16	90.163
Rank 32	94.018
Rank 64	96.918
	Rank 8 Rank 16 Rank 32 Rank 64 Rank 8 Rank 16 Rank 32

Table 18: Performance Comparison. 100 LoRAs

Configuration		Average Performance
Base Model		32.28
Uncompressed LoRAs		48.32
Join Compression		
Diagonal	Rank 8	41.90
-	Rank 16	45.44
	Rank 32	46.89
	Rank 64	47.43
Full	Rank 8	43.88
	Rank 16	45.79
	Rank 32	46.83
	Rank 64	47.66

Table 19: Task-Based Performance Evaluation Across Different Models and Ranks

Task	Base Model	LoRA	Diagonal R8	Diagonal R16	Diagonal R32	Diagonal R64
Causal Judgement	57.47	64.37	55.17	58.62	58.62	58.62
Date Understanding	15.33	23.33	20.67	22.00	21.33	22.67
Formal Fallacies	51.33	56.00	52.67	52.67	53.33	54.67
Hyperbaton	6.67	68.00	57.33	63.33	67.33	68.00
Logical Deduction (5 Objects)	21.33	37.33	32.00	36.67	37.33	37.33
Logical Deduction (7 Objects)	12.67	44.00	31.33	42.67	44.67	45.33
Movie Recommendation	62.67	67.33	62.00	64.67	66.67	67.33
Object Counting	34.67	38.00	35.33	36.67	36.67	38.00
Snarks	50.00	61.54	53.85	56.41	58.97	57.69
Temporal Sequences	16.67	23.33	18.67	20.67	24.00	24.67
Average	32.88	48.32	41.90	45.44	46.89	47.43

Table 20: Task-Based Performance Evaluation Across Different Models and Ranks

Task	Base Model	LoRA	Full R8	Full R16	Full R32	Full R64
Causal Judgement	57.47	64.37	56.32	57.47	58.62	60.92
Date Understanding	15.33	23.33	19.33	22.00	22.67	22.67
Formal Fallacies	51.33	56.00	51.33	52.67	53.33	56.00
Hyperbaton	6.67	68.00	63.33	66.00	69.33	68.00
Logical Deduction (5 Objects)	21.33	37.33	35.33	36.00	35.33	37.33
Logical Deduction (7 Objects)	12.67	44.00	40.00	44.67	44.67	44.67
Movie Recommendation	62.67	67.33	63.33	65.33	67.33	67.33
Object Counting	34.67	38.00	35.33	36.67	37.33	37.33
Snarks	50.00	61.54	53.85	55.13	57.69	58.97
Temporal Sequences	16.67	23.33	20.67	22.00	22.00	23.33
Average	32.88	48.32	43.88	45.79	46.83	47.66