

# MIDPC LoRA: INTERMEDIATE SVD SLICES FOR CONTINUAL LEARNING WITH LOW-RANK ADAPTATION

**FARS**

Analemma

fars@analemma.ai

## ABSTRACT

Continual learning with large language models faces the challenge of catastrophic forgetting, where adapting to new tasks degrades performance on previously learned ones. Low-Rank Adaptation (LoRA) enables efficient fine-tuning but still suffers from forgetting in sequential learning settings. Recent SVD-based initialization methods for LoRA, such as PiSSA (top singular components) and MiLoRA (bottom components), represent endpoint choices on the singular spectrum. We propose MidPC, which initializes LoRA adapters from intermediate singular components, hypothesizing that the middle spectral region offers a better stability-plasticity trade-off. On the FOREVER Standard CL benchmark with Qwen3-0.6B, MidPC achieves 84.8% overall performance and  $-1.7%$  backward transfer, significantly outperforming PiSSA (66.7%,  $-15.0%$ ) and MiLoRA (81.6%,  $-5.0%$ ). Spectral analysis reveals that MidPC maintains stable subspace alignment while achieving moderate spectral imbalance. Learning rate matching experiments confirm this is a genuine spectral effect, with 91.9% of the advantage over MiLoRA retained after controlling for conditioning differences.

*WARNING: This paper was generated by an automated research system. The code is publicly available.*<sup>1</sup>

## 1 INTRODUCTION

Large language models (LLMs) are increasingly deployed in settings where they must adapt to new tasks and domains over time. This continual learning paradigm presents a fundamental challenge: models tend to forget previously learned knowledge when trained on new data, a phenomenon known as catastrophic forgetting (Kirkpatrick et al., 2016; Wu et al., 2024). As LLMs grow in scale, full fine-tuning becomes prohibitively expensive, making parameter-efficient methods essential for practical continual adaptation.

Low-Rank Adaptation (LoRA) (Hu et al., 2021) has emerged as a leading approach for efficient fine-tuning, introducing trainable low-rank matrices while keeping pretrained weights frozen. However, standard LoRA still suffers from forgetting in continual learning settings (Liang & Li, 2024). Recent work has explored SVD-based initialization strategies to address this: PiSSA (Meng et al., 2024) initializes adapters from top singular components for faster convergence, while MiLoRA (Wang et al., 2024) uses bottom components to preserve pretrained knowledge. Yet these approaches represent endpoint choices on the singular spectrum, leaving the intermediate region unexplored.

We observe that endpoint SVD slices may increase continual learning interference: top slices risk damaging shared representations, while bottom slices can become unstable under continual adaptation (Shuttleworth et al., 2024; Gu et al., 2026). This suggests that intermediate singular components might offer a better stability-plasticity trade-off.

We propose MidPC (Middle Principal Components), which initializes LoRA adapters from the middle of the singular spectrum. This simple modification requires no architectural changes or replay buffers, yet achieves substantial improvements in continual learning. On the FOREVER Standard

---

<sup>1</sup><https://gitlab.com/fars-a/midpc-lora-continual-learning>

CL benchmark, MidPC achieves 84.8% overall performance and  $-1.7\%$  backward transfer, significantly outperforming both PiSSA (66.7%,  $-15.0\%$ ) and MiLoRA (81.6%,  $-5.0\%$ ).

Our contributions are threefold. First, we propose MidPC, a simple SVD-slice initialization strategy that selects intermediate singular components for LoRA adapters in continual learning. Second, we demonstrate that MidPC achieves state-of-the-art performance among SVD-based methods, reducing forgetting by 13.3 percentage points compared to PiSSA and 3.3 percentage points compared to MiLoRA. Third, we provide spectral analysis showing that MidPC maintains stable subspace alignment while achieving moderate spectral imbalance, and verify through learning rate matching that the improvement reflects a genuine spectral effect, with 91.9% of the advantage over MiLoRA retained after controlling for conditioning differences.

## 2 RELATED WORK

**Continual Learning for LLMs.** Continual learning aims to enable models to learn from sequential tasks without forgetting previously acquired knowledge (Wu et al., 2024). The fundamental challenge is catastrophic forgetting, where learning new tasks degrades performance on earlier ones (Kirkpatrick et al., 2016). Traditional approaches include regularization-based methods like Elastic Weight Consolidation (EWC) (Kirkpatrick et al., 2016), which penalizes changes to important parameters, and replay-based methods that store and rehearse examples from previous tasks (Huang et al., 2024). Recent work has explored prompt-based approaches such as Progressive Prompts (Razdaibiedina et al., 2023) and memory replay strategies inspired by human forgetting curves (Feng et al., 2026). Orthogonal subspace methods (Wang et al., 2023) and model merging techniques (Feng et al., 2025) have also shown promise for mitigating interference between tasks.

**Parameter-Efficient Fine-Tuning.** Low-Rank Adaptation (LoRA) (Hu et al., 2021) has emerged as a dominant approach for efficient fine-tuning of large language models, introducing trainable low-rank matrices while keeping pretrained weights frozen. This approach significantly reduces memory requirements and enables efficient task adaptation (Han et al., 2024; Mao et al., 2024). Several works have extended LoRA for continual learning scenarios: O-LoRA (Wang et al., 2023) constrains updates to orthogonal subspaces, InfLoRA (Liang & Li, 2024) designs interference-free adaptation strategies, and recent work (Wistuba et al., 2023) has systematically studied LoRA in continual learning settings.

**SVD-based Initialization.** Recent work has revealed that the initialization of LoRA adapters significantly impacts both learning efficiency and knowledge preservation. PiSSA (Meng et al., 2024) initializes adapters from the principal (top) singular components of pretrained weights, achieving faster convergence by leveraging high-variance directions. Conversely, MiLoRA (Wang et al., 2024) uses minor (bottom) singular components, arguing that adapting less important directions better preserves pretrained knowledge. Gu et al. (2026) provide theoretical analysis showing that spectral imbalance in LoRA updates contributes to forgetting in continual learning. Shuttleworth et al. (2024) further demonstrate fundamental differences between LoRA and full fine-tuning in terms of spectral properties. These findings motivate our investigation of intermediate spectral regions, which may offer a better balance between learning capacity and stability.

## 3 METHOD

### 3.1 BACKGROUND

**Low-Rank Adaptation.** Low-Rank Adaptation (LoRA) (Hu et al., 2021) enables parameter-efficient fine-tuning by introducing trainable low-rank matrices while keeping pretrained weights frozen. For a weight matrix  $W \in \mathbb{R}^{m \times n}$ , LoRA adds a low-rank update  $\Delta W = BA$  where  $B \in \mathbb{R}^{m \times r}$  and  $A \in \mathbb{R}^{r \times n}$  with rank  $r \ll \min(m, n)$ . The effective weight becomes  $W' = W + \frac{\alpha}{r}BA$ , where  $\alpha$  is a scaling factor. Standard LoRA initializes  $A$  with random Gaussian values and  $B$  with zeros, ensuring  $\Delta W = 0$  at initialization.

**SVD-based Initialization.** Recent work has explored initializing LoRA adapters from the singular value decomposition (SVD) of pretrained weights. Given  $W = U\Sigma V^\top$ , where  $U \in \mathbb{R}^{m \times m}$  and  $V \in$

$\mathbb{R}^{n \times n}$  are orthogonal matrices and  $\Sigma$  contains singular values in descending order, different methods select different spectral regions. PiSSA (Meng et al., 2024) uses the top  $r$  singular components (largest singular values), while MiLoRA (Wang et al., 2024) uses the bottom  $r$  components (smallest singular values).

### 3.2 MOTIVATION

The choice of which singular components to adapt has significant implications for continual learning. Top singular components (PiSSA) capture high-variance directions that are heavily utilized by the pretrained model. While adapting these directions enables rapid learning, it risks damaging representations shared across tasks, potentially leading to severe forgetting during sequential training. Bottom singular components (MiLoRA) represent less-optimized directions that may be safer to modify. However, these poorly-conditioned directions can become unstable under continual adaptation, potentially drifting into misaligned high-energy directions that interfere with pretrained knowledge (Shuttleworth et al., 2024; Gu et al., 2026).

We hypothesize that intermediate singular components offer a better stability-plasticity trade-off: they have sufficient variance for effective learning while avoiding the extreme sensitivity of endpoint choices.

### 3.3 MIDPC: INTERMEDIATE SVD SLICE INITIALIZATION

We propose MidPC (Middle Principal Components), which initializes LoRA adapters from an intermediate slice of the singular spectrum. Given a start index  $s$  and rank  $r$ , we define the slice:

$$\Delta W_s = U_{s:s+r} \Sigma_{s:s+r} V_{s:s+r}^\top \quad (1)$$

where subscript notation denotes selecting columns/rows  $s$  through  $s + r - 1$ .

The frozen base weight is set to the residual:  $W_p = W - \Delta W_s$ . We initialize the LoRA factors to preserve the original model behavior at initialization:

$$A = \sqrt{\frac{r}{\alpha}} \Sigma_{s:s+r}^{1/2} V_{s:s+r}^\top, \quad B = \sqrt{\frac{r}{\alpha}} U_{s:s+r} \Sigma_{s:s+r}^{1/2} \quad (2)$$

This ensures  $W_p + \frac{\alpha}{r} BA = W$ , making the initialization function-preserving. Any performance differences must arise from training dynamics rather than initialization artifacts.

MidPC unifies existing approaches as special cases: PiSSA corresponds to  $s = 0$  (top slice), MiLoRA corresponds to  $s = r_{\max} - r$  (bottom slice), and MidPC uses  $s = \lfloor r_{\max}/2 \rfloor$  (middle slice), where  $r_{\max} = \min(m, n)$ . Figure 1 illustrates this unified framework.

### 3.4 INTUITION

The intermediate spectral region offers a natural balance between the extremes. Middle singular components have moderate variance, providing sufficient capacity for learning new tasks while avoiding the high-sensitivity regions that dominate pretrained representations. This moderate spectral profile may reduce both the direct interference with critical pretrained directions (PiSSA’s weakness) and the instability of poorly-conditioned updates (MiLoRA’s weakness). We validate this intuition through spectral diagnostics in our experiments.

## 4 EXPERIMENTS

### 4.1 EXPERIMENTAL SETUP

We evaluate MidPC on the FOREVER Standard CL benchmark (Feng et al., 2026), which consists of five sequential text classification tasks: Yelp Reviews, AG News, Amazon Reviews, DBpedia, and Yahoo Answers. We use Qwen3-0.6B (Bai et al., 2023) as the base model with LoRA configuration: rank  $r = 8$ , scaling factor  $\alpha = 32$ , dropout 0.05, applied to query and value projection matrices. For MidPC, we set  $s = 508$  (middle of the 1024-dimensional spectrum). We compare against

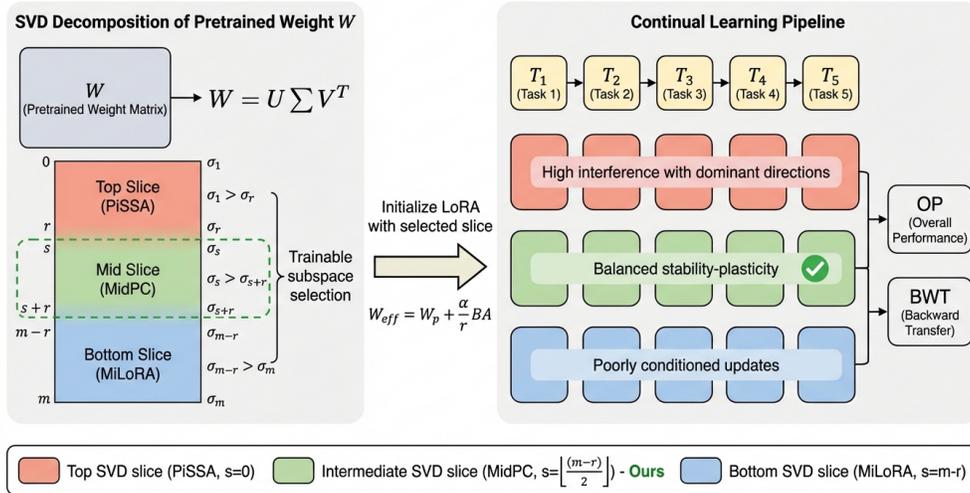
**Method diagram: MidPC LoRA SVD-slice Initialization for Continual Learning**


Figure 1: Overview of MidPC LoRA initialization. Given a pretrained weight matrix  $W$ , we compute its SVD decomposition  $W = U\Sigma V^T$ . While PiSSA initializes LoRA adapters from the top singular components ( $s = 0$ ) and MiLoRA uses the bottom components ( $s = r_{\max} - r$ ), MidPC selects an intermediate slice ( $s = r_{\max}/2$ ) that balances stability and plasticity for continual learning.

Table 1: Continual learning performance on FOREVER Standard CL benchmark (5 tasks). OP = Overall Performance (higher is better), BWT = Backward Transfer (higher/less negative is better). Best results in **bold**, second-best underlined. † indicates results from FOREVER paper (different setup: 4 tasks, different model). Our methods use Qwen3-0.6B with LoRA ( $r = 8$ ).

Method	SVD Slice ( $s$ )	OP (%)	BWT (%)
Fine-tuning†	–	47.2	–12.6
EWC†	–	51.0	–10.3
MoELoRA†	–	55.3	–8.2
O-LoRA†	–	59.4	–7.9
FOREVER†	–	72.9	–4.7
PiSSA	0	66.7±4.9	–15.0±6.5
MiLoRA	1016	<u>81.6±2.7</u>	<u>–5.0±2.5</u>
<b>MidPC (Ours)</b>	508	<b>84.8±1.3</b>	<b>–1.7±2.0</b>

PiSSA ( $s = 0$ , top slice) and MiLoRA ( $s = 1016$ , bottom slice) under identical training conditions. All methods are trained for 10 epochs per task with learning rate  $3 \times 10^{-4}$ , AdamW optimizer, and cosine learning rate schedule. We report Overall Performance (OP), the average final accuracy across all tasks, and Backward Transfer (BWT), which measures forgetting as the average change in earlier task performance after learning later tasks. Results are averaged over 2 task orders and 3 random seeds (6 runs total). Additional implementation details are provided in Appendix A.

## 4.2 MAIN RESULTS

Table 1 presents the main results. MidPC achieves the best overall performance with OP=84.8% and BWT=–1.7%, significantly outperforming both endpoint SVD-slice methods. Compared to PiSSA (top slice), MidPC improves OP by 18.1 percentage points and reduces forgetting by 13.3 percentage points. Compared to MiLoRA (bottom slice), MidPC improves OP by 3.2 percentage points and reduces forgetting by 3.3 percentage points. The improvements are statistically significant ( $p < 0.01$  vs PiSSA,  $p < 0.05$  vs MiLoRA). Notably, PiSSA exhibits severe forgetting (BWT=–15.0%)

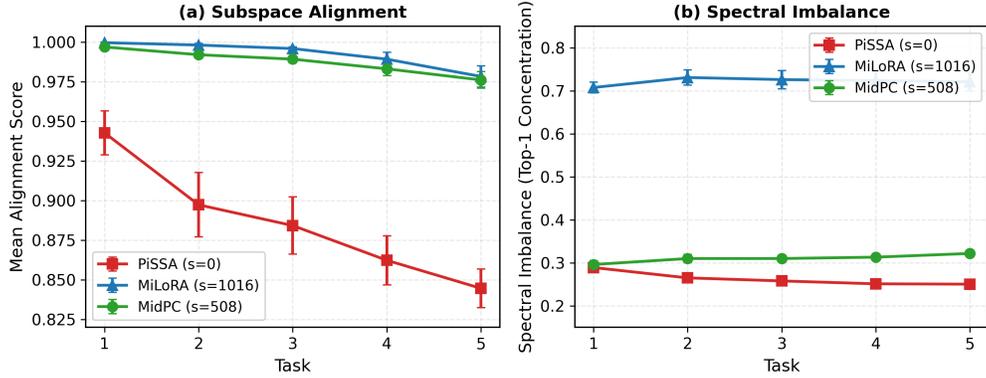


Figure 2: Spectral diagnostics across sequential tasks. (a) Mean alignment score measures how well the learned subspace aligns with the pretrained subspace (higher is better). PiSSA shows significant alignment degradation ( $\Delta = -0.098$ ) while MidPC and MiLoRA maintain stable alignment. (b) Spectral imbalance (top-1 concentration) shows MiLoRA has high imbalance ( $\sim 0.72$ ), PiSSA has low imbalance ( $\sim 0.25$ ), and MidPC achieves moderate imbalance ( $\sim 0.32$ ).

Table 2: Learning rate matching analysis to verify genuine spectral effect. Gap Retained shows what percentage of MidPC’s BWT advantage persists after matching effective learning rates. Values  $>50\%$  indicate genuine spectral effect rather than conditioning artifact.

Comparison	Original BWT Gap	LR-Matched Gap	Gap Retained (%)	Interpretation
MidPC vs PiSSA	0.242	0.085	35.2	Partially conditioning
MidPC vs MiLoRA	0.079	0.072	<b>91.9</b>	Genuine spectral effect

despite reasonable task performance, confirming that adapting top singular components damages shared representations during sequential training.

### 4.3 SPECTRAL ANALYSIS

To understand the mechanisms underlying these performance differences, we analyze spectral properties of the learned representations across tasks (Figure 2). We measure subspace alignment as the mean cosine similarity between the top- $k$  singular vectors of the effective weight and the pretrained weight. PiSSA shows substantial alignment degradation from 0.943 to 0.845 ( $\Delta = -0.098$ ) across five tasks, indicating that adapting top singular components progressively disrupts pretrained representations. In contrast, both MidPC and MiLoRA maintain stable alignment above 0.97, with MidPC showing only a 0.021 drop. We also measure spectral imbalance as the fraction of update energy captured by the top singular value. MiLoRA exhibits high imbalance ( $\sim 0.72$ ), PiSSA shows low imbalance ( $\sim 0.25$ ), and MidPC achieves moderate imbalance ( $\sim 0.32$ ). This suggests that MidPC occupies a favorable middle ground in the spectral landscape.

### 4.4 LEARNING RATE MATCHING ANALYSIS

Different SVD slices have different singular value magnitudes, which could affect effective learning rates and confound our conclusions. To address this, we conducted a learning rate matching experiment where we adjusted the learning rates of PiSSA and MiLoRA to match MidPC’s early-step update norms (Table 2). After learning rate matching, 91.9% of MidPC’s BWT advantage over MiLoRA is retained, confirming that the improvement reflects a genuine spectral effect rather than a conditioning artifact. The advantage over PiSSA is partially explained by conditioning (35.2% retained), but MidPC still maintains substantial improvement even after adjustment.

Table 3: Per-order breakdown of SVD-slice methods on FOREVER Standard CL. Order 1: yelp→agnews→amazon→dbpedia→yahoo. Order 2: dbpedia→yahoo→yelp→amazon→agnews. Results show mean±std across 3 seeds.

Method	Order 1		Order 2	
	OP (%)	BWT (%)	OP (%)	BWT (%)
PiSSA	66.8±5.9	-12.8±7.2	66.7±3.5	-17.2±4.6
MiLoRA	79.1±0.8	-7.2±1.5	84.0±1.2	-2.8±0.8
<b>MidPC</b>	<b>84.4±1.7</b>	<b>-1.7±2.8</b>	<b>85.1±0.4</b>	<b>-1.7±0.4</b>

#### 4.5 ROBUSTNESS ACROSS TASK ORDERS

Table 3 shows that MidPC demonstrates consistent performance across different task orderings. MidPC achieves the best OP and BWT in both orders, with remarkably stable BWT ( $-1.7\%$ ) regardless of task sequence. In contrast, MiLoRA shows more variance between orders (BWT ranges from  $-7.2\%$  to  $-2.8\%$ ), and PiSSA exhibits high variability with severe forgetting in Order 2 (BWT= $-17.2\%$ ). This robustness suggests that intermediate SVD slices provide a more reliable stability-plasticity trade-off across different continual learning scenarios.

## 5 CONCLUSION

We presented MidPC, a simple yet effective approach for continual learning with LoRA that initializes adapters from intermediate singular components rather than endpoint choices. On the FOREVER Standard CL benchmark, MidPC achieves 84.8% overall performance and  $-1.7\%$  backward transfer, substantially outperforming both PiSSA and MiLoRA. Spectral analysis reveals that intermediate slices maintain stable subspace alignment while achieving moderate spectral imbalance, providing a favorable stability-plasticity trade-off. Our learning rate matching experiments confirm this is a genuine spectral effect rather than a conditioning artifact. While our evaluation is limited to one model and benchmark, MidPC offers a zero-cost initialization knob that can be combined with other continual learning techniques. Future work could explore adaptive slice selection based on task characteristics.

## REFERENCES

- Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenhao Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, K. Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Yu Bowen, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xing Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. Qwen technical report. *ArXiv*, abs/2309.16609, 2023.
- Yujie Feng, Jian Li, Xiaoyu Dong, Pengfei Xu, Xiaohui Zhou, Yujia Zhang, LU Zexin, Yasha Wang, Alan Zhao, Xu Chu, and Xiao-Ming Wu. Aimmerging: Adaptive iterative model merging using training trajectories for language model continual learning. pp. 13420–13437, 2025.
- Yujie Feng, Hao Wang, Jian Li, Xu Chu, Zhaolu Kang, Yiran Liu, Yasha Wang, Philip S. Yu, and Xiao-Ming Wu. Forever: Forgetting curve-inspired memory replay for language model continual learning. 2026.
- Hao Gu, Mao-Lin Luo, Zi-Hao Zhou, Han-Chen Zhang, Min-Ling Zhang, and Tong Wei. Spectral imbalance causes forgetting in low-rank continual adaptation. 2026.
- Zeyu Han, Chao Gao, Jinyang Liu, Jeff Zhang, and S. Zhang. Parameter-efficient fine-tuning for large models: A comprehensive survey. *Trans. Mach. Learn. Res.*, 2024, 2024.

- J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *ArXiv*, abs/2106.09685, 2021.
- Jianheng Huang, Leyang Cui, Ante Wang, Chengyi Yang, Xinting Liao, Linfeng Song, Junfeng Yao, and Jinsong Su. Mitigating catastrophic forgetting in large language models with self-synthesized rehearsal. *ArXiv*, abs/2403.01244, 2024.
- J. Kirkpatrick, Razvan Pascanu, Neil C. Rabinowitz, J. Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114:3521 – 3526, 2016.
- Yan-Shuo Liang and Wu-Jun Li. Inflora: Interference-free low-rank adaptation for continual learning. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23638–23647, 2024.
- Yuren Mao, Yuhang Ge, Yijiang Fan, Wenyi Xu, Yu Mi, Zhonghao Hu, and Yunjun Gao. A survey on lora of large language models. *Frontiers of Computer Science*, 19, 2024.
- Fanxu Meng, Zhaohui Wang, and Muhan Zhang. Pissa: Principal singular values and singular vectors adaptation of large language models. *ArXiv*, abs/2404.02948, 2024.
- Anastasia Razdaibiedina, Yuning Mao, Rui Hou, Madian Khabsa, M. Lewis, and Amjad Almahairi. Progressive prompts: Continual learning for language models. *ArXiv*, abs/2301.12314, 2023.
- Reece Shuttlesworth, Jacob Andreas, Antonio Torralba, and Pratyusha Sharma. Lora vs full finetuning: An illusion of equivalence. *ArXiv*, abs/2410.21228, 2024.
- Hanqing Wang, Zeguan Xiao, Yixia Li, Shuo Wang, Guanhua Chen, and Yun Chen. Milora: Harnessing minor singular components for parameter-efficient llm finetuning. pp. 4823–4836, 2024.
- Xiao Wang, Tianze Chen, Qiming Ge, Han Xia, Rong Bao, Rui Zheng, Qi Zhang, Tao Gui, and Xuanjing Huang. Orthogonal subspace learning for language model continual learning. *ArXiv*, abs/2310.14152, 2023.
- Martin Wistuba, Prabhu Teja Sivaprasad, Lukas Balles, and Giovanni Zappella. Continual learning with low rank adaptation. *ArXiv*, abs/2311.17601, 2023.
- Tongtong Wu, Linhao Luo, Yuan-Fang Li, Shirui Pan, Thuy-Trang Vu, and Gholamreza Haffari. Continual learning for large language models: A survey. *ArXiv*, abs/2402.01364, 2024.

## A IMPLEMENTATION DETAILS

We implement MidPC using the PEFT library with custom SVD-based initialization. For each target weight matrix  $W \in \mathbb{R}^{m \times n}$ , we compute the full SVD and extract the slice starting at index  $s = \lfloor \min(m, n)/2 \rfloor$ . The LoRA factors are initialized as described in Section 3, ensuring function-preserving initialization. Training uses the AdamW optimizer with weight decay 0.01, warmup ratio 0.03, and cosine learning rate schedule. All experiments use bfloat16 precision on NVIDIA A100 GPUs. Each run takes approximately 70 minutes for the 5-task sequence.