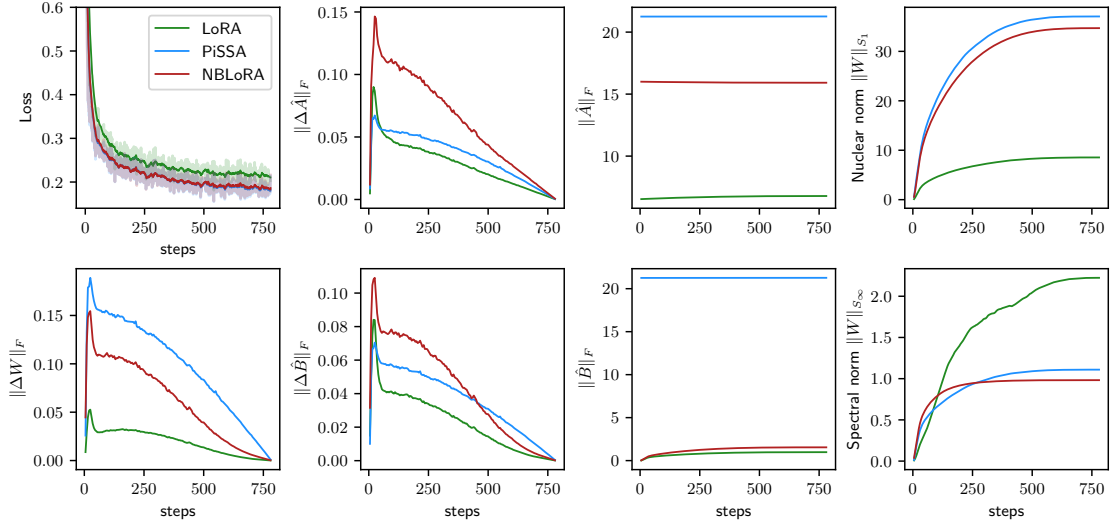


Additional Figures

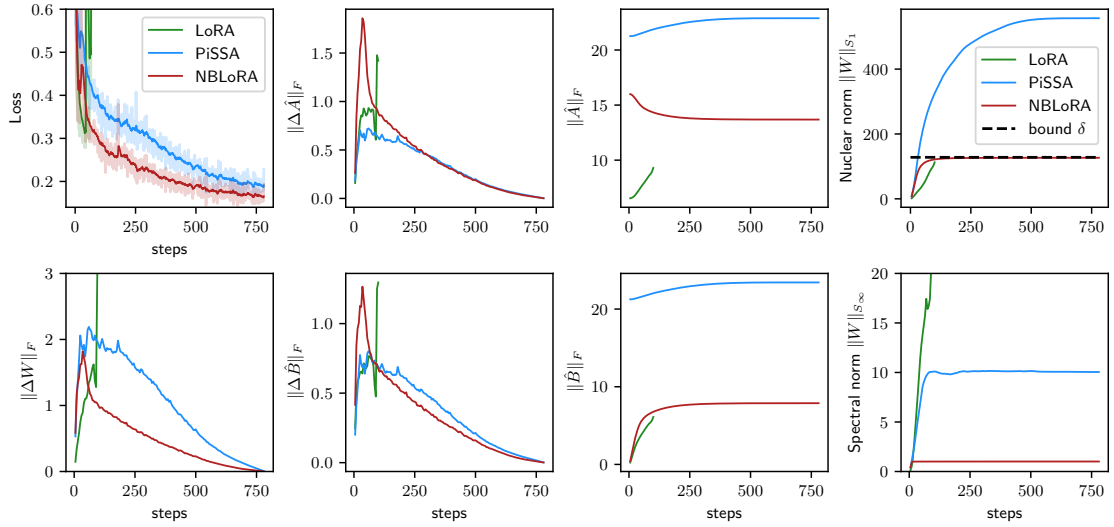
- Fig. 1 contains the training dynamics analysis for LoRA, PiSSA and NBLoRA, see the response to Reviewer SKyQ for detail discussion.
- Fig. 2 adds the norm ratio plot for each experimental setup, see the response to Reviewer SKyQ for detail discussion.
- Fig. 3 contains the training dynamics analysis for DeLoRa and NBLoRA, see the response to Reviewer smpu for detail discussion.
- Fig. 4 contains the zoom-in plots for the training dynamics of LoRA, see the response to Reviewer smpu for detail discussion.

Table 1: Fine-tuning three base models based on LoRA (Lo), DoRA (Do), PiSSA (Pi) and NB-LoRA (NB) over different learning rates ($\{1\text{e-}5, 5\text{e-}5, 1\text{e-}4, \textcolor{blue}{2\text{e-}4}\}$ for Mistral and $\{5\text{e-}5, 1\text{e-}4, 5\text{e-}4, \textcolor{blue}{7\text{e-}4}\}$ for LLaMA). We report the minimum, maximum and averaged test results, where the metrics for math and coding are $\frac{1}{2}(\text{GSM8K} + \text{MATH})$ and $\frac{1}{2}(\text{HumanEval} + \text{MBPP})$, respectively.

Base Model		Mistral-7B-v0.1				LLaMA-3-8B				LLaMA-2-13B				Model Avg.			
Method		Lo	Do	Pi	NB	Lo	Do	Pi	NB	Lo	Do	Pi	NB	Lo	Do	Pi	NB
Math	min	43.9	44.2	42.2	42.0	49.5	49.6	37.9	47.8	35.1	35.1	36.8	38.1	42.9	43.0	38.9	42.6
	max	49.1	48.1	47.0	47.9	51.5	51.8	52.0	52.9	41.8	41.0	40.4	42.2	47.4	47.0	46.5	47.7
	avg	47.2	46.8	45.4	46.0	50.5	50.9	45.6	50.3	38.9	38.8	38.7	40.4	45.5	45.5	43.2	45.6
Code	min	52.4	53.7	52.9	53.6	56.3	56.5	44.4	57.4	42.5	42.5	40.1	44.0	50.4	50.9	45.8	51.7
	max	57.8	59.2	59.0	59.7	63.2	62.6	63.0	68.1	46.6	47.2	45.6	49.4	55.9	56.4	55.9	59.1
	avg	56.1	57.0	56.0	57.5	60.3	60.5	52.6	62.0	44.7	45.0	43.9	47.6	53.7	54.2	50.8	55.7
Task Avg.	min	48.1	48.9	47.5	47.8	52.9	53.1	41.2	52.6	38.8	38.8	38.5	41.0	46.6	46.9	42.4	47.2
	max	53.5	53.7	53.0	53.8	57.4	57.2	57.5	60.5	44.2	44.1	43.0	45.8	51.7	51.7	51.2	53.4
	avg	51.7	51.9	50.7	51.8	55.4	55.7	49.1	56.2	41.8	41.9	41.3	44.0	49.6	49.8	47.0	50.6

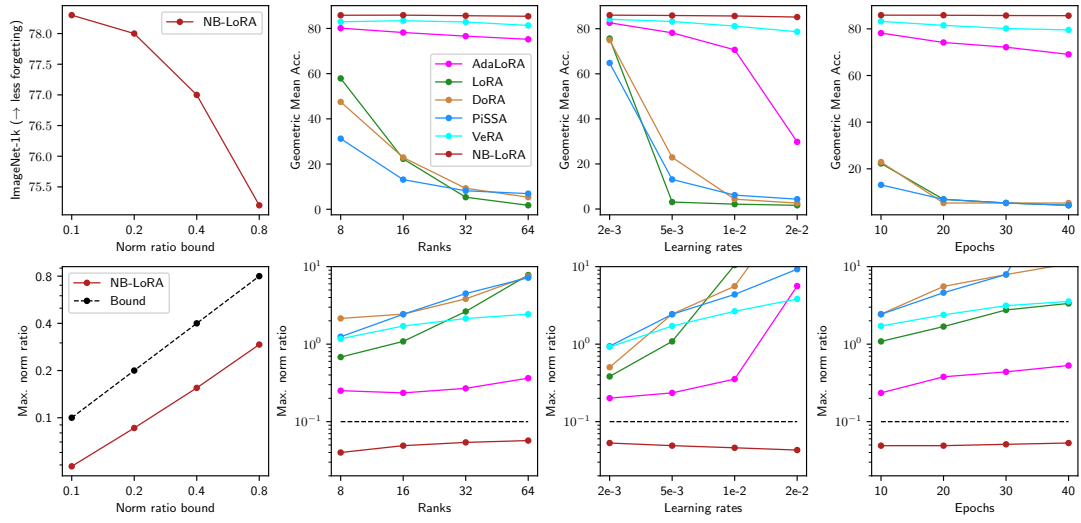


(a) Learning rate of 5e-5

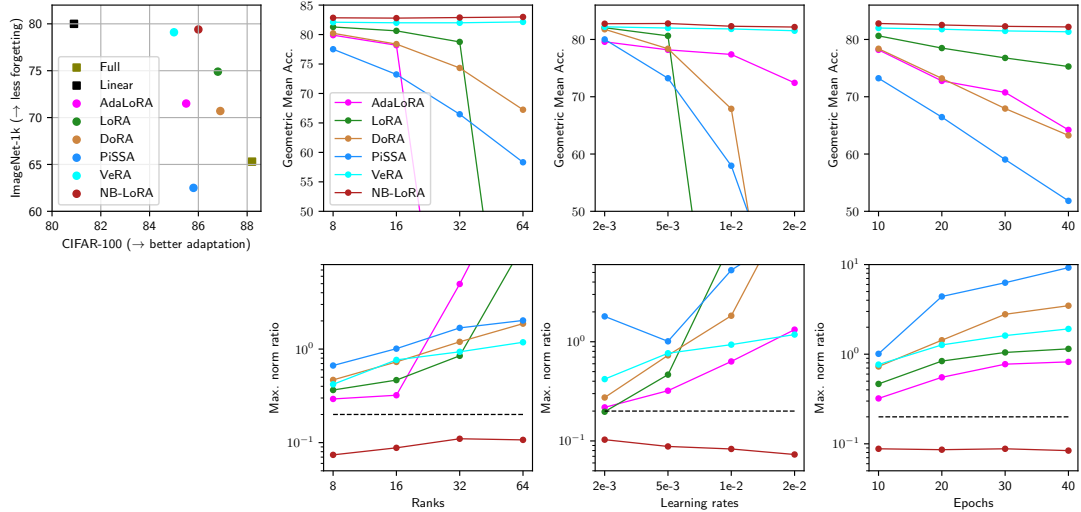


(b) Learning rate of 1e-3

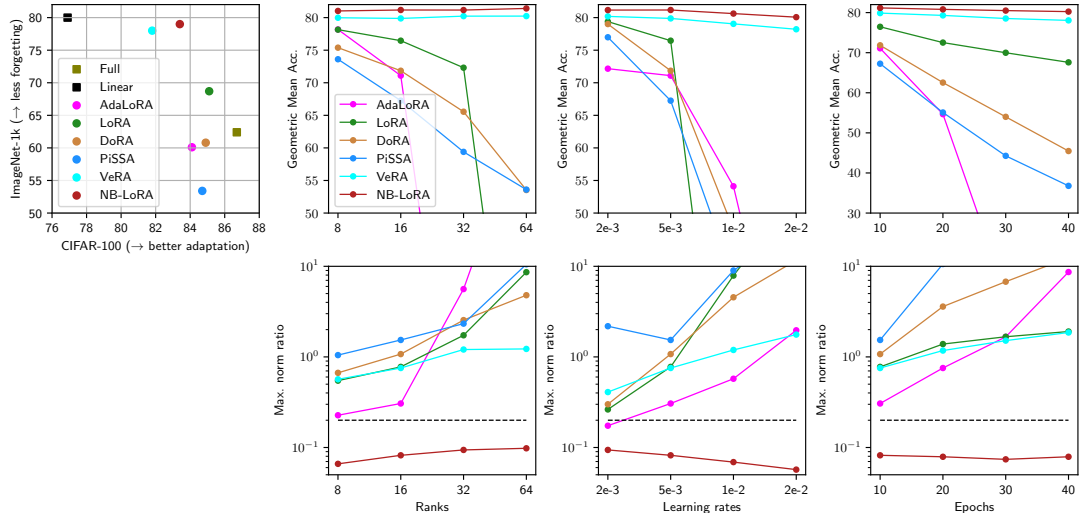
Figure 1: Training dynamics for LoRA, PiSSA and NBLoRA two learning rates.



(a) SVHN



(b) CIFAR-100



(c) Food-101

Figure 2: Geometric mean of source and target accuracies and norm-ratio of weight adapter with different adapters on various of hyper-parameter setup.

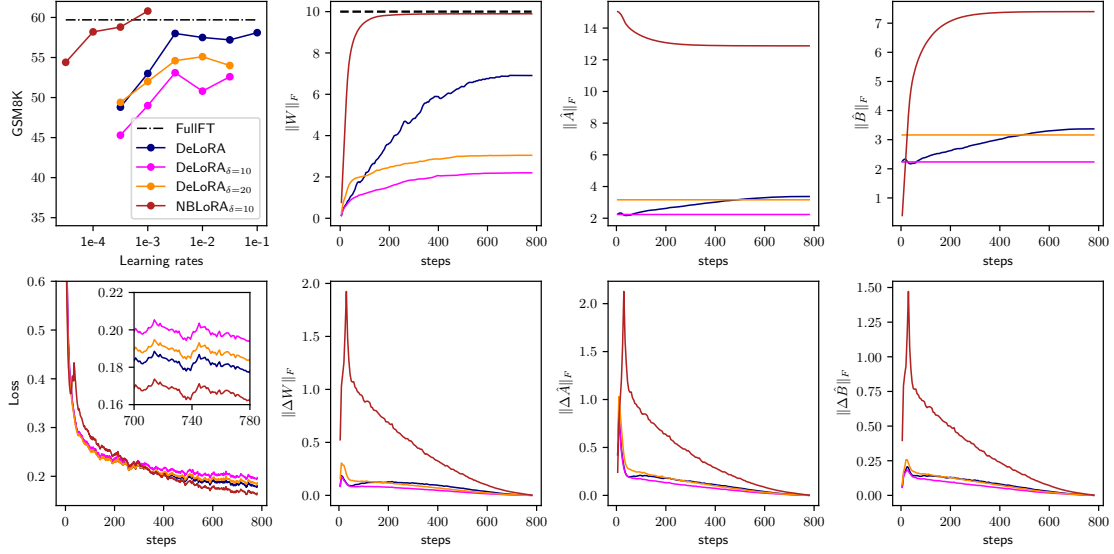


Figure 3: Training dynamics of NB-LoRA and DeLoRA. Note that Figure 4 of the paper reports matrix norms of DeLoRA ($lr=1e-1$), DeLoRA $_{\delta=10}$ ($lr=5e-2$) and DeLoRA $_{\delta=20}$ ($lr=5e-2$), whereas the figure above shows the corresponding result under $lr=5e-3$.

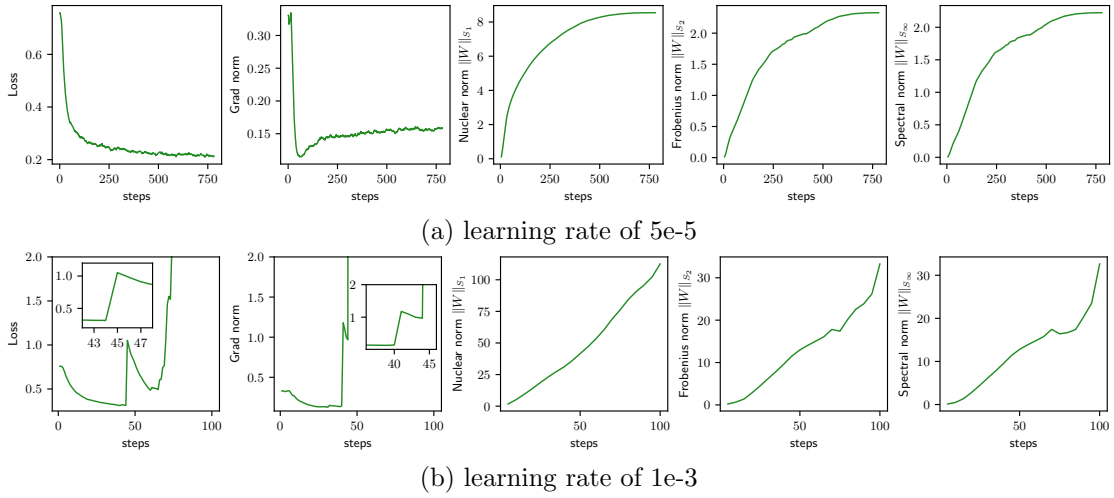


Figure 4: Training dynamics for LoRA under two learning rates