A Additional Experimental Details

A.1 Computational hardware

All our experiments were carried out on Linux servers equipped with an AMD EPYC 7763 64-Core CPU processor, 512GB RAM, and NVIDIA RTX 6000 ADA 48G / A800 80G GPU.

A.2 Hyperparameters

Table 6: Hyperparameter configurations for GLUE benchmark.

Model	Hyperparameter	SST-2	MRPC	CoLA	QNLI	RTE	STS-B
	Optimizer	AdamW					
	LR Scheduler	Linear					
	Warmup Ratio	0.06					
	Max Seq. Len.	512					
	Spectral Coefficients n	{250,500}					
Base	Rank r	32	64	64	8	32	256
	Epochs	50	30	100	40	100	90
	Batch Size	128	32	128	32	32	32
	LR (Head)	6E-4	6E-4	3E-4	6E-5	3E-4	2E-4
	LR (FoRA)	2E-2	4E-2	4E-2	7E-2	3E-2	2E-2
Large	Rank r	32	32	32	32	32	32
	Epochs	20	50	100	30	70	40
	Batch Size	128	32	128	8	32	32
	LR (Head)	1E-4	2E-4	4E-4	4E-4	3E-4	7E-5
	LR (FoRA)	3E-2	5E-2	4E-2	2E-2	2E-2	3E-2

Table 7: Hyperparameter configurations for mathematical reasoning.

	LLaMA2_{7B}		LLaMA2_{13B}		LLaMA3_{8B}	
Hyperparameter	GSM8k	MATH	GSM8k	MATH	GSM8k	MATH
Optimizer	AdamW					
LR Scheduler	Cosine					
Batch Size	16					
Warmup Ratio	0.05					
Dropout	0.05					
Epochs	3					
Where	Q,V					
Spectral Coefficients n	20000		30000		20000	
Rank r (FoRA)	256	128	256	128	256	128
Rank r (DFoRA)	256	128	256	128	128	128
LR (LoRA)	5E-4	5E-4	5E-4	6E-4	5E-4	5E-4
LR (DoRA)	4E-4	5E-4	4E-4	6E-4	6E-4	2E-4
LR (FoRA)	6E-3	5E-3	5E-3	5E-3	1E-3	9E-4
LR (DFoRA)	5E-3	3E-3	6E-3	6E-3	1E-3	9E-4

Table 8: Hyperparameter configurations for commonsense reasoning.

	$\text{LLaMA}_{7\text{B}}$		$\text{LLaMA}_{13\text{B}}$		$\text{LLaMA2}_{7\text{B}}$		$\text{LLaMA3}_{8\text{B}}$	
Hyperparameter	FoRA	DFoRA	FoRA	DFoRA	FoRA	DFoRA	FoRA	DFoRA
Optimizer	AdamW							
LR Scheduler	Linear							
Batch Size	16							
Warmup Steps	100							
Dropout	0.05							
Epochs	3							
Rank r	32							
Alpha α	64							
Where	Q,K,V,Up,Down							
Spectral Coefficients n	30000		40000		30000		30000	
LR	1E-3	1.4E-3	9E-4	9E-4	8E-4	8E-4	5E-4	5E-4

Table 9: Hyperparameter configurations for finetuning ViT on the image classification datasets.

Model	Hyperparameter	OxfordPets	StanfordCars	DTD	EuroSAT	FGVC	RESISC
	Optimizer	AdamW					
	Epochs	10					
	Batch Size	64					
	Rank r (LoRA)	16					
	Spectral Coefficients n	8000					
BASE	Rank r (FoRA)	32	128	64	64	256	32
	LR (Head)	8E-3	1E-2	1E-2	1E-4	1E-2	1E-2
	LR (FoRA)	4E-3	5E-2	5E-3	2E-2	5E-2	2E-2
	Weight Decay	4E-2	1E-5	2E-4	4E-3	2E-2	9E-2
LARGE	Rank r (FoRA)	64	128	128	64	256	32
	LR (Head)	6E-3	5E-3	1E-2	1E-3	1E-2	1E-2
	LR (FoRA)	5E-3	3E-2	4E-3	3E-2	8E-2	1E-2
	Weight Decay	3E-4	2E-5	3E-5	3E-3	1E-2	1E-3

A.3 Parameter Count of Sparse Learning Strategies

As the rank increases, the number of learnable parameters in LoRA grows linearly, leading to a significant parameter overhead. While VeRA exhibits a minimal increase in parameters, its strong dependence on the size of its adaptation matrices limits its flexibility in adapting to more complex tasks. In contrast, both FoRA and random masking maintain a fixed number of learnable parameters across different ranks, providing greater flexibility by allowing parameter adjustments based on task complexity.

Table 10: Comparison of learnable parameters across different compression strategies.

		Rank r					
	Methods	2^3	2^4	2^5	2^6	2^7	2^8
$\text{RoBERTa}_{\text{BASE}}$	LoRA	6,144	12,288	24,576	49,152	98,304	196,608
	VeRA	776	784	800	832	896	1024
	FoRA/Mask	500	500	500	500	500	500
ViT_{BASE}	LoRA	6,144	12,288	24,576	49,152	98,304	196,608
	VeRA	776	784	800	832	896	1024
	FoRA/Mask	6,144	8000	8000	8000	8000	8000

B Implemetation

Algorithm 1 presents the PyTorch implementation of FoRA. Our approach allows for a straightforward plug-in, with the only modification needed being the replacement of the PyTorch linear module with the FoRA linear module. Additionally, we create a cached empty spectral matrix that matches the size of the adaptation matrix. This caching strategy not only accelerates GPU computations but also minimizes GPU overhead, as only the sparse spectral coefficients require backward propagation.

Algorithm 1 FoRA PyTorch Implementation

```
class PEFT(nn.Module):
    def __init__(self, ...):
        if LoRA:
            A, B = torch.nn.Linear(d1, r), torch.nn.Linear(r, d2)
        elif FoRA:
            A, B = FoRALinear(n, d1, r), FoRALinear(n, r, d2)

class FoRALinear(nn.Module):
    def __init__(self, n, din, dout):
        self.n = n # number of spectral coefficients
        self.din = din # input dimension
        self.dout = dout # output dimension
        # spectral location
        self.L = torch.randperm(din * dout)[:n]
        self.L = torch.stack([self.L // din, self.L % dout], dim=0)
        # spectral coefficients
        self.s = nn.Parameter(torch.zeros(n))
        # cached spectral matrix
        self.cached_F = torch.zeros(self.din, self.dout)

    def weight(self):
        # spectral assignment
        F = self.cached_F.clone()
        F[self.L[0], self.L[1]] = self.s
        # spatial conversion
        W = torch.fft.ifft2(F).real()
        return W
```

C Generation Examples

Instruction Tuning (MATH)

How many inches are in the circumference of a great circle of a sphere which has a surface area of 196π sq in? Express your answer in terms of π.
Note: To measure the circumference, take the largest circle on the surface of the sphere.
Answer the above question. First think step by step and then answer the final number.

Response

Model	**Response**
LLaMA2$_{7B}$ (FoRA)	The surface area of a sphere is πr^2, where r is the radius. So, the radius is $r = \sqrt{196\pi} = 12\sqrt{3}$. The circumference of the sphere is $2\pi r = \boxed{24\pi}$. The final answer is: 24π.
LLaMA3$_{8B}$ (FoRA)	The surface area of a sphere is $4\pi r^2$, so $r^2 = 49$ and $r = 7$. The circumference of the great circle is $2\pi r = \boxed{14\pi}$. The final answer is: 14π.

Instruction Tuning (HellaS.)

Please choose the correct ending to complete the given sentence: Roof shingle removal: A man is sitting on a roof. he
Ending1: is using wrap to wrap a pair of skis. Ending2: is ripping level tiles off. Ending3: is holding a rubik's cube. Ending4: starts pulling up roofing on a roof.
Answer format: ending1/ending2/ending3/ending4

Response

Model	**Response**
LLaMA2$_{7B}$ (FoRA)	the correct answer is ending2
LLaMA3$_{8B}$ (FoRA)	the correct answer is ending4