Table A.1: Score on AlpacaEval2.0 with LLaMA2-7B. The reference model is Text-davinci-003.

Method	#Params.	Finetuning Data	Win Rate (%)
LoRA	0.83%	10K cleaned Alpaca	61.55
LoReFT	0.03%	10K cleaned Alpaca	60.21
RoAd_1	0.02%	10K cleaned Alpaca	**62.64**
LoReFT	0.03%	UltraFeedback	61.68
RoAd_1	0.02%	UltraFeedback	**62.60**

Table A.2: Visual instruction tuning results for LLaVA1.5-7B. The mean of three random runs is reported here for all methods.

Method	#Params.	GQA	SQA	VQAT	POPE	Avg.
LoRA	4.61%	62.4	**68.5**	56.9	**86.0**	**68.5**
RoAd_4	0.08%	60.0	66.9	53.3	85.5	66.4
RoAd_1 + LoRA	1.19%	**62.5**	68.2	**57.4**	85.8	**68.5**

Table A.3: Performance on the GLUE benchmark.

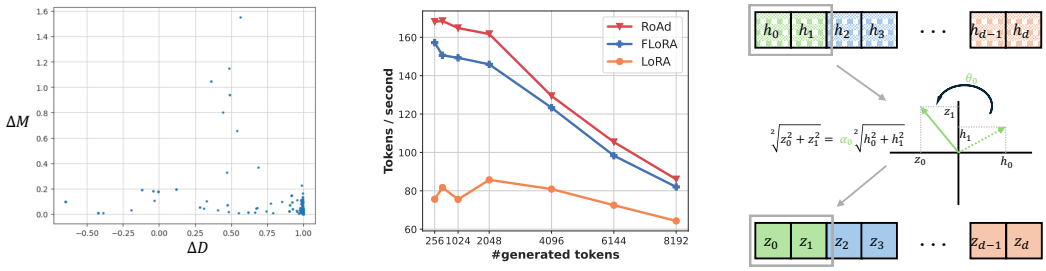
Row	Model	Method	#Params.	RTE	MRPC	STS-B	CoLA	SST-2	QNLI	QQP	MNLI	Avg.
1	RoBERTa-base	Full FT	100.00%	78.3	87.9	90.6	62.4	94.4	**92.5**	**91.7**	**87.3**	85.6
2	RoBERTa-base	RoAd_4	0.27%	**80.8**	**90.7**	90.5	**65.2**	93.6	91.8	90.0	86.8	**86.2**
3	RoBERTa-base	RoAd_2	0.13%	80.3	90.2	**90.8**	64.9	93.9	92.0	90.1	86.7	86.1
4	RoBERTa-base	RoAd_1	0.07%	78.9	89.2	90.5	64.4	93.9	91.9	89.6	86.3	85.6
5	RoBERTa-base	RoAd_1 (fc1)	0.03%	79.1	90.2	90.2	60.9	**94.6**	91.6	88.7	85.4	85.1
6	RoBERTa-large	Full FT	100.00%	85.8	**91.7**	**92.6**	**68.2**	96.0	93.8	**91.5**	88.8	88.6
7	RoBERTa-large	RoAd_4	0.25%	89.0	90.7	92.1	66.6	96.0	93.8	91.4	**90.1**	88.7
8	RoBERTa-large	RoAd_2	0.12%	**89.7**	91.3	92.0	66.7	95.8	**94.7**	91.4	90.0	**88.9**
9	RoBERTa-large	RoAd_1	0.06%	89.2	91.0	91.7	66.1	**96.3**	94.4	91.0	89.7	88.7
10	RoBERTa-large	RoAd_1 (fc1)	0.03%	88.7	91.5	91.9	68.1	96.1	94.5	90.2	89.6	88.8
11	T5-base	Full FT	220M	71.9	90.2	89.7	61.8	94.6	93.0	**91.6**	**86.8**	84.9
12	T5-base	ATTEMPT-m	96K	82.7	87.3	90.8	**64.3**	94.3	**93.2**	90.1	83.7	85.8
13	T5-base	RoAd_1-m	111K	**85.6**	**92.0**	**90.9**	59.3	**94.8**	92.5	90.1	85.9	**86.4**

Table A.4: Accuracy on the commonsense reasoning tasks.

Row	Model	Method	#Params.	BoolQ	PIQA	SIQA	HellaS.	WinoG.	ARC-e	ARC-c	OBQA	Avg.
1	LLaMA-7B	RoAd_1 + LoRA	0.84%	71.2	84.0	81.4	94.0	84.2	86.7	71.2	84.8	82.2
2	LLaMA-13B	RoAd_1 + LoRA	0.68%	73.4	87.4	83.3	95.8	86.6	90.0	78.2	88.8	85.4
3	LLaMA-7B	SSF	0.04%	67.8	82.3	78.8	91.5	79.1	83.1	65.6	77.4	78.2
4	LLaMA-13B	SSF	0.03%	71.0	85.0	82.0	93.9	85.3	86.4	71.8	82.6	82.3
5	LLaMA-7B	LoRA	0.04%	65.0	80.5	77.6	80.3	79.0	76.0	63.0	72.4	74.2

Table A.5: Accuracy on the arithmetic reasoning tasks.

Row	Model	Method	#Params.	AQuA	GSM8K	MAWPS	SVAMP	Avg.
1	LLaMA-7B	RoAd_1 + LoRA	0.84%	24.4	35.5	84.5	55.4	50.0
2	LLaMA-13B	RoAd_1 + LoRA	0.68%	24.4	46.5	86.6	62.9	55.1
3	LLaMA-7B	SSF	0.04%	20.1	17.9	66.4	39.3	35.9
4	LLaMA-13B	SSF	0.03%	22.4	27.9	72.7	48.7	42.9
5	LLaMA-7B	LoRA	0.04%	18.5	27.0	75.3	46.1	41.7



(a) Pilot study on LlaMA2-7B (b) Batching efficiency (c) Overview of RoAd

Figure A.1: **(a)** The change in magnitude and angle of representations between pretrained and finetuned LLM using LoRA. **(b)** Throughput on LLaMA-7B. FLoRA is new here.