

# Rebuttal Responses

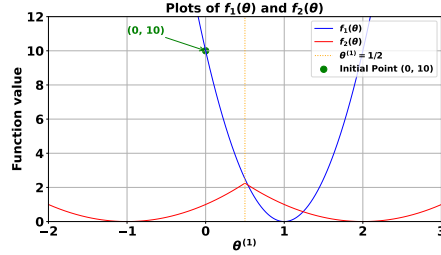
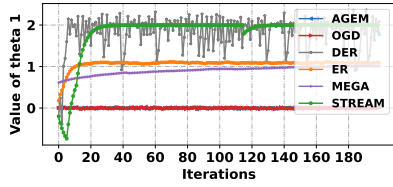
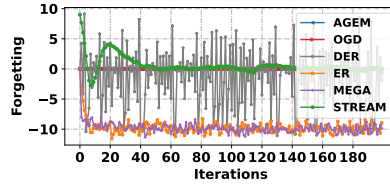


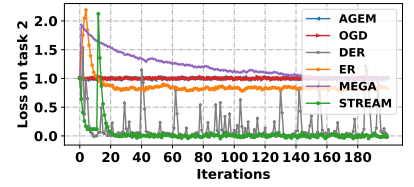
Figure 1: Loss function of two tasks.  $f_1(\theta)$  and  $f_2(\theta)$  have different minimums and  $f_2(\theta)$  is dominant across most values of  $\theta_1$ .



(a)  $\theta^{(1)}$  value vs. iterations.



(b) The forgetting vs. iterations.

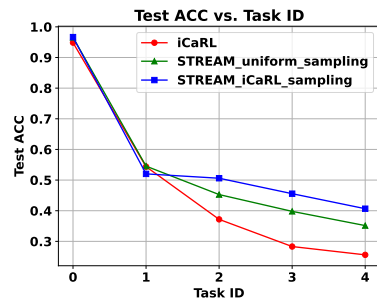


(c) Loss on task 2 vs. iterations.

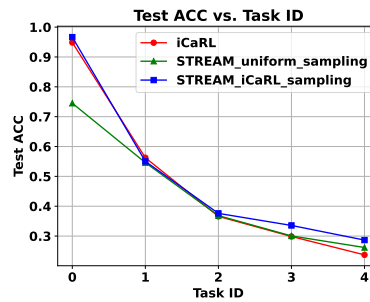
Figure 2: Synthetic experiment for the counterexample. We add Gaussian random noise to simulate the stochastic gradient descent. STREAM can find the optimal  $\theta^{(1)}$  and achieve minimal forgetting and loss on the new task.

Table 1: Results on Multiple Dataset, Split CIFAR100, and Split Tiny Imagenet.

Methods	Multiple Dataset		Split CIFAR100		Split Tiny-Imagenet	
	ACC ( $\uparrow$ )	FGT ( $\downarrow$ )	ACC ( $\uparrow$ )	FGT ( $\downarrow$ )	ACC ( $\uparrow$ )	FGT ( $\downarrow$ )
NCL	46.64 $\pm$ 2.32	0.334 $\pm$ 0.035	44.42 $\pm$ 1.35	0.325 $\pm$ 0.024	20.43 $\pm$ 0.64	0.301 $\pm$ 0.009
SGP	55.68 $\pm$ 1.23	0.305 $\pm$ 0.153	56.55 $\pm$ 1.64	<b>0.100<math>\pm</math>0.001</b>	26.86 $\pm$ 0.90	0.122 $\pm$ 0.008
STREAM	<b>72.08<math>\pm</math>1.40</b>	<b>0.152<math>\pm</math>0.035</b>	<b>64.06<math>\pm</math>0.86</b>	0.132 $\pm$ 0.010	<b>31.36<math>\pm</math>0.71</b>	<b>0.121<math>\pm</math>0.008</b>



(a) Average testing ACC on CIFAR10.

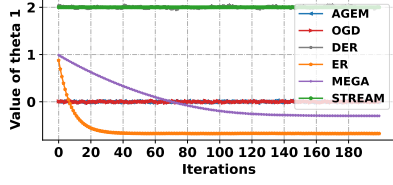


(b) Average testing ACC on CIFAR10 with label noise  $p = 0.2$ .

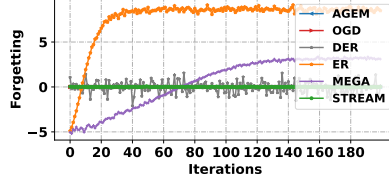
Figure 3: Class-incremental learning comparison with different sampling methods.

Table 2: Continual learning process (retain two decimal places for the numerical values).

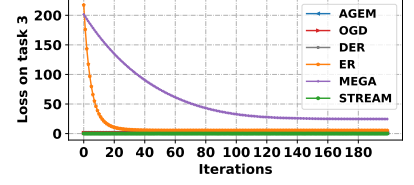
Task	$\theta^{(1)}$		$\theta^{(2)}$		FGT↓		Loss↓	
	ER	STREAM	ER	STREAM	ER	STREAM	ER	STREAM
$f_1$	0.00	0.00	0.00	0.00	0.00	0.00	10.00	10.00
$f_2$	1.09	2.00	0.00	0.00	-9.20	0.00	0.83	0.00
$f_3$	-0.67	2.00	0.00	0.00	8.62	0.00	0.00	0.00



(a)  $\theta^{(1)}$  value vs. iterations.



(b) The forgetting on task 1 and task 2 vs. iterations.



(c) Loss on task 3 vs. iterations.

Figure 4: Synthetic experiment on task 3 for the counterexample. Each algorithm starts from their points at the end of the task2.

Table 3: Results on Multiple Dataset, Split CIFAR100, and Split Tiny Imagenet.

Methods	Multiple Dataset		Split CIFAR100		Split Tiny-Imagenet	
	ACC (↑)	FGT (↓)	ACC (↑)	FGT (↓)	ACC (↑)	FGT (↓)
ER	59.03±0.90	0.244±0.024	48.56±1.74	0.342±0.025	17.54±0.83	0.230±0.019
"vanilla" ER	61.33± 0.94	0.329±0.043	60.92±0.93	0.142±0.009	23.38±0.56	0.170±0.011
STREAM	<b>72.08±1.40</b>	<b>0.152±0.035</b>	<b>64.06±0.86</b>	<b>0.132±0.010</b>	<b>31.36±0.71</b>	<b>0.121±0.008</b>

Table 4: Running time/memory on Multiple Dataset and Split CIFAR-100.

Methods	Multiple Dataset (hours/MB)	Split CIFAR-100 (hours/MB)
EWC	0.16/552	1.31/930
MAS	0.17/562	1.31/963
AGEM	0.16/570	1.30/1102
OGD	0.47/558	3.01/926
DER	0.18/778	1.27/1146
GDumb	0.13/572	0.92/1034
MEGA	0.15/560	1.05/1032
STREAM	0.11/572	0.79/1046