Rebuttal Responses

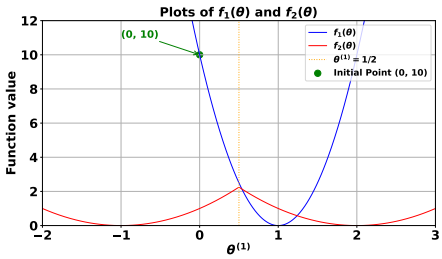
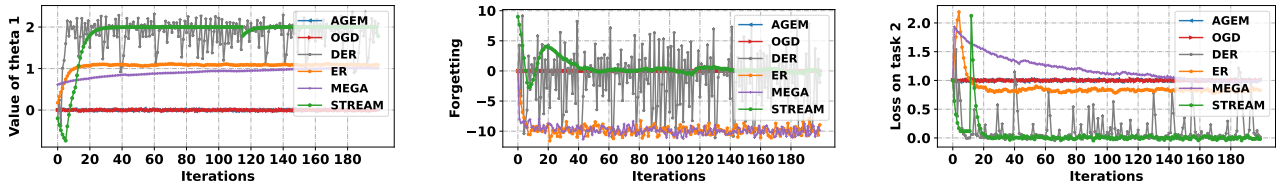


Figure 1: Loss function of two tasks. $f_1(\theta)$ and $f_2(\theta)$ have different minimums and $f_2(\theta)$ is dominant across most values of θ_1.



(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 (↑)	FGT (↓)	ACC (↑)	FGT (↓)	ACC (↑)	FGT (↓)
NCL	46.64±2.32	0.334±0.035	44.42±1.35	0.325±0.024	20.43±0.64	0.301±0.009
SGP	55.68±1.23	0.305±0.153	56.55±1.64	**0.100±0.001**	26.86±0.90	0.122±0.008
STREAM	**72.08±1.40**	**0.152±0.035**	**64.06±0.86**	0.132±0.010	**31.36±0.71**	**0.121±0.008**