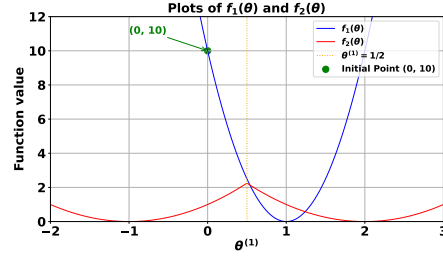# Rebuttal Response Figures
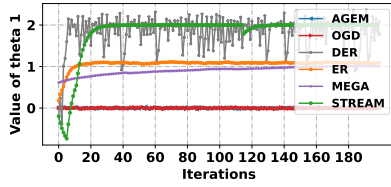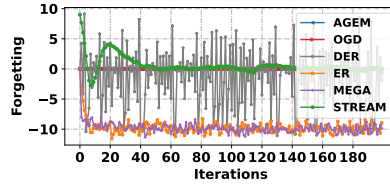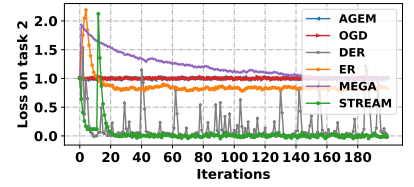

Figure 1: Loss function of two tasks. $f_1(\theta)$ and $f_2(\theta)$ have different minimums and $f_2(\theta)$ is dominant across most values of θ_1.

(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.