

A Additional Experiment Conditions

A.1 Abductive Reasoning

We set the hypothesis length $N = 15$ in the backward pass and allow the forward pass to generate $N*2$ tokens for complete sentences. We run $T = 20$ forward-backward iterations, with each backward pass performing 20 gradient updates using a small step size $\lambda = 0.0003$. The mixing weight of forward/backward logits is $\gamma = 0.88$. We use greedy decoding to produce a single candidate at each iteration T .

A.2 Counterfactual Reasoning

We use a step size $\lambda = 0.0004$ in backward pass and a mixing weight $\gamma = 0.92$. One difference from the abductive task is that, here we vary

the number of forward-backward iterations within $\{5, 10\}$ and the number of backward gradient updates within $\{5, 8, 10, 15\}$. Each configuration produces one candidate at the end of the algorithm. So for each example, we produce 8 candidates for ranking. We found such a generation-ranking protocol gives better performance on the counterfactual task.

Since we need to generate 3 sentences, the number of tokens N is relatively large. For the effectiveness of backpropagation and forward computation, we split the generation into 3 segments, one for each sentence, and perform the forward-backward passes for each segment separately. A sentence that was generated for the i th segment, is then appended to the context when generating the $i+1$ segment.