TOWARDS ELIMINATING CATASTROPHIC FORGETTING IN THE CURRICULUM LEARNING OF MATH REASONING TASKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Curriculum learning has found broad application across numerous domains. Nevertheless, its effectiveness is intrinsically curtailed by catastrophic forgetting, driven by the shifts in model parameter distributions between curriculum tasks. In this paper, we investigate the phenomenon of catastrophic forgetting in this training paradigm, building on the established efficacy of curriculum learning. Our theoretical analyses of parameter update dynamics demonstrate that catastrophic forgetting in curriculum learning stems from the divergence of task optima, which is generally essential to the faster convergence of curriculum learning; therefore, forgetting cannot be completely eliminated. Based on this finding, we augment the training process and propose IV-EWC, which incorporates Elastic Weight Consolidation (EWC) into the curriculum learning objective to curb catastrophic forgetting in mathematical reasoning, a prototypical curriculum learning scenario. IV-EWC employs the influence function to construct a representative validation set from the curriculum's training data, which is used to drive dynamic regularization during training. We further present an extended theoretical analysis to show that EWC-based regularization methods mitigate catastrophic forgetting in curriculum learning, thereby providing theoretical support for IV-EWC. Empirical evaluations on three backbone models and three benchmarks indicate that curriculum learning exhibits catastrophic forgetting. IV-EWC alleviates this issue, reducing forgetting by 162% on average relative to vanilla curriculum learning and yielding positive backward transfer, as evidenced by improved performance on easier tasks after subsequent training on challenging tasks.

1 INTRODUCTION

In the field of machine learning, curriculum learning can help improve generalization, accelerate convergence and find better local minima (Bengio et al., 2009; Guo et al., 2018; Kumar et al., 2010; Hacohen & Weinshall, 2019). Nowadays, it has been widely adopted in the post-training scenario, combined with new methods such as reinforcement learning to improve the reasoning capabilities (Xi et al., 2024; Wen et al., 2025; Parashar et al., 2026) of large language models.

Curriculum learning decomposes the original problem into an ordered sequence of subtasks. The resulting non-stationary training distribution can induce catastrophic forgetting (McCloskey & Cohen, 1989; French, 1999; Hetherington & Seidenberg, 2014): the accuracy on easier tasks drops and the model is inclined to apply strategies learned for challenging tasks to easier tasks, where they may be suboptimal. Yet most methods optimize only the curriculum design for forward progress and overlook backward stability, leaving performance vulnerable to forgetting.

Catastrophic forgetting can be understood as optimization-driven interference in the parameter space. Training on a newly introduced task directs gradient updates toward parameter regions that minimize the new task's loss. When the loss minima across tasks differ from each other, these updates move the model away from regions that previously achieved low loss on earlier tasks, leading to performance degradation. Importantly, forgetting is defined relative to task delineations and their ordering: the extent of interference depends on the task objectives and data distributions that

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

constitute the curriculum. Thus, overparameterization alone does not prevent forgetting; even large language models trained sequentially can exhibit path dependence that favors challenging tasks.

In this study, we conduct theoretical analyses of curriculum learning in an idealized regime in which the loss of the target problem and all intermediate curriculum problems are smooth and strongly convex. Examining the parameter updating dynamics, we find that curriculum learning's benefit derives from accelerated convergence. At the same time, catastrophic forgetting arises from non-coincident optima across tasks. In particular, achieving a reduced iteration count generally requires that the optima of the easier tasks do not coincide with those of the challenging tasks. This entails an inherent trade-off: the very differences that enable faster convergence also induce forgetting, rendering it impossible to completely eliminate forgetting in this setting.

To this end, we address catastrophic forgetting within curriculum learning by intervening in the training process. We introduce Influence-Validation EWC (IV-EWC), which dynamically applies EWC regularization (Kirkpatrick et al., 2017) using a validation set selected from the training data of curriculum tasks via the influence function (Koh & Liang, 2017), thereby protecting parameters critical to easier tasks of the curriculum. We focus on mathematical reasoning tasks because they exhibit clear difficulty hierarchies and explicit solution steps, enabling objective curriculum construction and transparent assessment of the reasoning process. Building on the analytical framework of forgetting in curriculum learning, we further show that EWC-based regularization alleviates this phenomenon under the same setting, backing up IV-EWC in theory. To summarize, our main contributions are described as follows.

- We identify and formalize the phenomenon of catastrophic forgetting in curriculum learning, providing theoretical analysis from the parameter update perspective in gradient descent to establish that forgetting can be attributed to the differences of optima across curriculum tasks, which is required for fewer iterations in curriculum learning, thus forgetting exists and cannot be eliminated.
- We propose IV-EWC, which integrates influence-function-guided selection of representative training samples with a dynamic EWC regularizer, aiming to stabilize parameters important to easier curriculum tasks while adapting to more challenging ones in mathematical reasoning tasks.
- Within the same theoretical setting, we investigate EWC-based regularization in curriculum learning and prove that it attenuates catastrophic forgetting by constraining the distribution shifts of the parameters, thereby providing a theoretical justification for IV-EWC.
- We conduct extensive experiments with three base models on three math reasoning datasets; the results corroborate the presence of forgetting in curriculum learning and demonstrate that IV-EWC substantially mitigates this degradation.

2 Related Work

2.1 Curriculum Learning on Reasoning Tasks

Curriculum learning (Bengio et al., 2009) organizes training from easier to more difficult instances or tasks and has inspired self-paced, teacher-student, and automated curricula (Graves et al., 2017; Matiisen et al., 2020; Jiang et al., 2018; Hacohen & Weinshall, 2019). These strategies have proved effective across deep learning, including neural machine translation (Kocmi & Bojar, 2017; Zhang et al., 2019; Platanios et al., 2019; Liu et al., 2020) and computer vision with noisy labels (Guo et al., 2018; Xiao et al., 2024). Curriculum learning has also been explored in medical report generation (Liu et al., 2021), biological sequence prediction (Zhang et al., 2025), as well as robotic control (Florensa et al., 2017; Wang et al., 2019; 2020; Leyendecker et al., 2021). With the emergence of reinforcement learning methods such as DPO (Rafailov et al., 2023) and GRPO (Shao et al., 2024), there has been a growing body of research that combines curriculum learning with reinforcement learning to improve the reasoning capabilities of models. R^3 (Xi et al., 2024) uses outcome supervision and achieves the effect of process supervision by constructing a reverse curriculum sequence. Light-R1 (Wen et al., 2025) constructs a three-stage curriculum post-training strategy that leverages carefully selected reasoning data to train long-CoT models. E2H Reasoner (Parashar et al., 2026) designs a scheduling strategy that utilizes data with difficulty annotations to automate the curriculum learning process, thereby improving the performance on complex tasks.

2.2 Mitigating Forgetting in Continual Learning

When investigating forgetting in curriculum learning, some methods in continual learning (Thrun, 1995) can serve as references. iCaRL (Rebuffi et al., 2017) employs a herding-based strategy to select representative samples for replay while adding a distillation loss to the objective. EWC (Kirkpatrick et al., 2017) applies parameter-wise regularization based on the Fisher Information Matrix. Deep Generative Replay (Shin et al., 2017) trains another generative model to generate replay samples that mimic past tasks. CLEAR (Rolnick et al., 2019) leverages off-policy learning and behavioral cloning for replay and on-policy learning for new samples. GEM (Lopez-Paz & Ranzato, 2017) utilizes episodic memory to control the gradient update process by projecting the gradient onto the feasible region defined by previous tasks' memories. A-GEM (Chaudhry et al., 2019) simplifies the constraints of GEM and updates gradients according to sampled memories, achieving better efficiency. UPGD (Elsayed & Mahmood, 2024) uses utility to assess the importance of each parameter in the training process and parameters with high utility are protected. Jiang et al. (2025) investigate the correlation between the function vector and forgetting, applying a function-vector based regularization to minimize forgetting.

3 Problem Formulation

In this section, we establish the effectiveness of curriculum learning for optimization problems with smooth and strongly convex loss functions. It further shows, from the parameter updating process, that forgetting is intrinsic to curriculum learning and impairs its eventual performance.

Let the loss of the original problem be a smooth and strongly convex function $f(\boldsymbol{\theta})$, with the global optimal parameter $\boldsymbol{\theta}^*$ and the condition number κ. Define the loss of the easier curriculum problem as $f_e(\boldsymbol{\theta})$, with its global optimal parameter $\boldsymbol{\theta}_e^*$ and its condition number κ_e. There exists a $\delta > 0$ which satisfies $||\boldsymbol{\theta}_e^* - \boldsymbol{\theta}^*|| \leq \delta$. Both the original and the easier problems are optimized using gradient descent, where $\boldsymbol{\theta}_0$ is the initial parameter, T_f and T_e are the iterations when the original problem and the easier curriculum problem converge respectively.

Under standard smoothness and strong convexity assumptions, we first bound the distance between the current iterate and the optimal parameter in Lemma 1 via the update recursion, then we quantify the number of iterations required to reduce the parameter error below a prescribed tolerance in Lemma 2. These lemmas provide the foundational results for the subsequent theorems on the effectiveness of curriculum learning and the phenomenon of forgetting therein.

Lemma 1. *For the target problem with a μ-strongly convex and L-smooth loss, let t be the optimization step, $\kappa = \frac{L}{\mu}$ be the condition number of the loss, and $\boldsymbol{\theta}^t$, $\boldsymbol{\theta}^*$, $\boldsymbol{\theta}_0$ be the parameter at step t, the optimal parameter of the target problem and the initial parameter of the target problem, respectively. The distance between parameters follows:*

$$||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 \leq \left(1 - \frac{1}{\kappa}\right)^t ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||^2 \tag{1}$$

The proof of Lemma 1 is detailed in Appendix A.1, where we first analyze the distance between consecutive iterates using the smoothness and strong convexity and then derive the final estimate for the distance from the current iterate to the optimal parameter.

Lemma 2. *For the target problem with a μ-strongly convex and L-smooth loss, define the convergence condition $||\boldsymbol{\theta}_T - \boldsymbol{\theta}^*|| \leq \varepsilon$ where $\boldsymbol{\theta}^T$ and $\boldsymbol{\theta}^*$ are the parameter at convergence and the optimal parameter of the target problem respectively, let $C > 0$ be a constant and κ be the condition number of the loss. The number of iterations required by gradient descent can be expressed as:*

$$T \approx C\kappa \log\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}{\varepsilon}\right) \tag{2}$$

The proof of Lemma 2 is detailed in Appendix A.2, which relies on Lemma 1 to obtain a refined upper bound on the parameter distance and derive an estimate of the number of iterations required.

From the perspective of parameter updating, we utilize Lemma 2 to estimate the iterations required for curriculum learning and for non-curriculum training under the same convergence condition in

Theorem 1. Under the assumptions that the loss of the easier curriculum problem has a lower condition number than the loss of the original problem and that their optimal parameters are close in distance, we establish that initializing on the easier problem and then transitioning to the original problem accelerates convergence, which is the effectiveness of curriculum learning.

Theorem 1. *Given the convergence condition of the easier curriculum problem* $||\boldsymbol{\theta}_{T_e} - \boldsymbol{\theta}_e^*|| \leq \varepsilon_e$ *and the convergence condition of the original problem* $||\boldsymbol{\theta}_T - \boldsymbol{\theta}^*|| \leq \varepsilon$*, for curriculum satisfying* $||\boldsymbol{\theta}_e^* - \boldsymbol{\theta}^*|| \ll ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*|| - \varepsilon_e$ *and* $\kappa_e < \kappa$*, curriculum learning requires fewer iterations to converge compared with learning at once.*

The detailed proof of Theorem 1 is given in Appendix A.3. First we employ Lemma 2 to establish the relation of the iteration counts and the parameter distance during curriculum learning, then we formulate the iterations saved to show the effectiveness of curriculum learning.

Theorem 1 demonstrates that appropriately designed curricula can reduce the number of iterations required to converge. Complementing this, we take a specific curriculum as an example to analyze why and when forgetting emerges during curriculum learning in Theorem 2, which shows that discrepancies between optima of curriculum tasks inevitably produce forgetting in curriculum learning. This forgetting is essential to the construction of curricula in Theorem 1 with losses satisfying increasing condition numbers, which drive accelerated convergence. Hence, achieving acceleration via curriculum learning necessarily entails some degree of forgetting.

Theorem 2. *For the curricula with loss functions* $f_t(\boldsymbol{\theta}) = \frac{L_t}{2}||\boldsymbol{\theta} - \boldsymbol{\theta}_t^*||^2$*, where* $\boldsymbol{\theta}_t^*$ *is the optimal parameter,* $t = 1, 2, ..., T$*,* $L_t > L_{t-1} > 0$*,* $||\boldsymbol{\theta}_t^* - \boldsymbol{\theta}_{t-1}^*||^2 > 0$ *and* $||\boldsymbol{\theta}_{t-1}^*||^2 > 0$*, when optimized starting from* $\boldsymbol{\theta}_0 = \mathbf{0}$ *with learning rate* $\eta > 0$*, forgetting exists and cannot be eliminated.*

Proof. Forgetting is defined as the average loss of previous tasks. Let the parameter after learning the last task be $\boldsymbol{\theta}_T$, and forgetting can be formulated as:

$$\mathrm{Fgt} = \frac{1}{T}\sum_{t=1}^{T}\left(f_t(\boldsymbol{\theta}_T) - f_t^*(\boldsymbol{\theta}_t^*)\right) \tag{3}$$

In accordance with IPM (Cai & Diakonikolas, 2025), the parameter update for optimizing the t-th task with learning rate η is given by:

$$\boldsymbol{\theta}_t = \mathrm{prox}_{\eta f_t}(\boldsymbol{\theta}_t) = \underset{\boldsymbol{\theta}\in\mathbb{R}^d}{\arg\min}\left(\frac{1}{2\eta}||\boldsymbol{\theta} - \boldsymbol{\theta}_{t-1}||^2 + f_t(\boldsymbol{\theta})\right) \tag{4}$$

Using Eq. 4, for the specified curriculum tasks we have:

$$\boldsymbol{\theta}_t = \underset{\boldsymbol{\theta}\in\mathbb{R}^d}{\arg\min}\left(\frac{1}{2\eta}||\boldsymbol{\theta} - \boldsymbol{\theta}_{t-1}||^2 + \frac{L_t}{2}||\boldsymbol{\theta} - \boldsymbol{\theta}_t^*||^2\right) \tag{5}$$

The optimization problem in Eq. 5 is quadratic and possesses a global minimum. Setting its derivative to zero yields

$$\begin{aligned}\left(\frac{1}{\eta} + L_t\right)\boldsymbol{\theta}_t &= \frac{\boldsymbol{\theta}_{t-1}}{\eta} + L_t\boldsymbol{\theta}_t^* \\ \boldsymbol{\theta}_t &= \frac{\boldsymbol{\theta}_{t-1} + \eta L_t\boldsymbol{\theta}_t^*}{1 + \eta L_t}\end{aligned} \tag{6}$$

To simplify the expression for $\boldsymbol{\theta}_t$, we define $\alpha_t = \frac{1}{1+\eta L_t}$. Eq. 6 can be rewritten as:

$$\boldsymbol{\theta}_t = \alpha_t\boldsymbol{\theta}_{t-1} + (1 - \alpha_t)\boldsymbol{\theta}_t^* \tag{7}$$

From Eq. 7, the process of parameter updating consists of two parts. $\alpha_t\boldsymbol{\theta}_{t-1}$ is the starting point of the update representing the regularization loss, while $(1 - \alpha_t)\boldsymbol{\theta}_t^*$ corresponds to the target of the update representing the target loss.

The curriculum design satisfies an ascending L_t and a constant η, thus α_t decreases with respect to t. As the curriculum concludes, $\boldsymbol{\theta}_t$ approaches $\boldsymbol{\theta}_t^*$.

We now turn to the forgetting in the curriculum. From Eq. 3, we have:

$$\mathrm{Fgt} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{L_t}{2}||\boldsymbol{\theta}_T - \boldsymbol{\theta}_t^*||^2\right) \tag{8}$$

Based on the analysis of Eq. 7, since at the end of the curriculum $\boldsymbol{\theta}_T \approx \boldsymbol{\theta}_T^*$, we have

$$\mathrm{Fgt}^{vanilla} \approx \frac{1}{T}\sum_{t=1}^{T}\left(\frac{L_t}{2}||\boldsymbol{\theta}_T^* - \boldsymbol{\theta}_t^*||^2\right) \tag{9}$$

Following the curriculum design we have $L_t > 0$ and $||\boldsymbol{\theta}_T^* - \boldsymbol{\theta}_t^*||^2 > 0$. Consequently, $\mathrm{Fgt} > 0$ always holds, and forgetting cannot be eliminated.

□

4 IV-EWC METHOD

In this section, we introduce IV-EWC based on the preceding analysis of the effectiveness and forgetting in curriculum learning. We first detail the augmented training strategy of IV-EWC. We then theoretically analyze the effect of incorporating the Elastic Weight Consolidation (EWC) loss term, which utilizes the Fisher Information Matrix computed from the previous training data to provide parameter-wise regularization, characterizing the ability of EWC-based methods to mitigate catastrophic forgetting in curriculum learning.

4.1 INFLUENCE-VALIDATION EWC

As formalized in Eq. 9, forgetting in curriculum learning arises because different tasks induce distinct optima for the model parameters. Since tasks in the curriculum generally do not share the same optima, forgetting cannot be completely eliminated through curriculum design alone. Nevertheless, constraining parameter updates during training can limit drift toward the optima of the challenging tasks, providing a new perspective to eliminate catastrophic forgetting. Motivated by this observation and the relatedness among tasks in the curriculum, IV-EWC employs the influence function to select a progressively expanding validation set from the curriculum's training data to compute the Fisher Information Matrix and apply EWC regularization, thereby attenuating the impact of noisy samples that contribute little to an informative regularization signal.

In curriculum learning, challenging tasks generally subsume the knowledge required for easier tasks. Excessive regularization during training on the challenging tasks can bias the model toward simple representations that favor the easier tasks, thereby limiting its ability to capture the additional complexity of the challenging tasks. To mitigate this effect, we employ a dynamic regularization coefficient $\lambda = \frac{\gamma}{t}$, where λ is the regularization weight of EWC loss, $\gamma > 0$ is a hyperparameter controlling the initial regularization weight, and t denotes the task index starting from 0. Regularization starts from task number 1 and progressively reduces as the curriculum advances, enabling more effective parameter updating while preserving previously acquired knowledge during the training of challenging tasks. The loss function of the t-th task during training is

$$L(\boldsymbol{\theta}_t) = \begin{cases} L_0(\boldsymbol{\theta}_0), & t = 0 \\ L_t(\boldsymbol{\theta}_t) + \sum_i \lambda \boldsymbol{F}_{\boldsymbol{\theta}_i}||\boldsymbol{\theta}_{t,i} - \boldsymbol{\theta}_{t-1,i}||^2, & t > 0 \end{cases} \tag{10}$$

where $\boldsymbol{\theta}_t$ is the model parameter trained during task t, $\boldsymbol{\theta}_{t-1}$ is the model parameter trained after task $t-1$, $\boldsymbol{\theta}_0$ is the model parameter during task 0 training, L_t is the task loss of task t, L_0 is the task loss of task 0, $\boldsymbol{F}$ is the Fisher Information Matrix computed using the validation set after task $t-1$ and i labels the model parameters.

4.2 THEORETICAL ANALYSIS

To assess the effectiveness of IV-EWC, we examine EWC regularization in curriculum learning under the setup of Theorem 2. Using forgetting defined in Eq. 8, Theorem 3 shows that EWC-based regularization reduces forgetting and can eliminate it entirely with a suitably tuned regularization weight, providing a formal justification for IV-EWC.

Theorem 3. *For the curricula with loss functions $f_t(\boldsymbol{\theta}) = \frac{L_t}{2}||\boldsymbol{\theta} - \boldsymbol{\theta}_t^*||^2$, where $\boldsymbol{\theta}_t^*$ is the optimal parameter, $t = 1, 2, ..., T$, $L_t > L_{t-1} > 0$, $||\boldsymbol{\theta}_t^* - \boldsymbol{\theta}_{t-1}^*||^2 > 0$ and $||\boldsymbol{\theta}_{t-1}^*||^2 > 0$, when the training data of task t follows a Gaussian distribution $\mathcal{N}(\boldsymbol{\theta}_t^*, \frac{1}{L_t})$ and the optimization starts from $\boldsymbol{\theta}_0 = \mathbf{0}$ with learning rate $\eta > 0$, incorporating an EWC loss calculated from data across previous tasks with the regularization weight λ ($0 < \lambda \leq 1$) is able to alleviate forgetting.*

Proof. To better compare curriculum learning with EWC regularization and vanilla curriculum learning, we first simplify Eq. 8. Under the specified curriculum setting, Fgt has a global minimum, where the optimal $\boldsymbol{\theta}_T^{opt}$ is given by

$$\boldsymbol{\theta}_T^{opt} = \frac{\sum_{t=1}^{T} L_t \boldsymbol{\theta}_t^*}{\sum_{t=1}^{T} L_t} \tag{11}$$

Then forgetting of the last task in the curriculum can be simplified as follows

$$\begin{aligned} \mathrm{Fgt} &= \frac{1}{2T}\left(\sum_{t=1}^{T} L_t||\boldsymbol{\theta}_T||^2 - 2\sum_{t=1}^{T} L_t\langle\boldsymbol{\theta}_T^{opt}, \boldsymbol{\theta}_T\rangle + \sum_{t=1}^{T} L_t||\boldsymbol{\theta}_t^*||^2\right) \\ &= \frac{1}{2T}\sum_{t=1}^{T} L_t||\boldsymbol{\theta}_T - \boldsymbol{\theta}_T^{opt}||^2 + C' \end{aligned} \tag{12}$$

where $\langle\cdot\rangle$ is the dot product of two vectors, $C' = \frac{1}{2T}\left(\sum_{t=1}^{T} L_t||\boldsymbol{\theta}_t^*||^2 - \sum_{t=1}^{T} L_t||\boldsymbol{\theta}_T^{opt}||^2\right)$ is a constant defined by the curricula.

Under the assumption of the t-th task's training data distribution $\mathcal{N}(\boldsymbol{\theta}_t^*, \frac{1}{L_t})$, the Fisher Information Matrix used in EWC can be simplified to the second derivative of the loss, which is L_t. Adding an EWC loss across previous tasks to the original task yields the following new loss function for task t:

$$f_t^{EWC}(\boldsymbol{\theta}) = \frac{L_t}{2}||\boldsymbol{\theta} - \boldsymbol{\theta}_t^*||^2 + \frac{\lambda}{2}\sum_{t'=1}^{t-1}\left(L_{t'}||\boldsymbol{\theta} - \boldsymbol{\theta}_{L_{t'}}^*||^2\right) \tag{13}$$

The loss of the last task in the curriculum with EWC is $f_T^{EWC}(\boldsymbol{\theta})$, setting its derivative to zero gives

$$\boldsymbol{\theta}_T^{EWC} = \frac{L_T\boldsymbol{\theta}_T^* + \lambda\sum_{t'=1}^{T-1} L_{t'}\boldsymbol{\theta}_{t'}^*}{L_T + \lambda\sum_{t'=1}^{T-1} L_{t'}} \tag{14}$$

Forgetting of the last task with EWC regularization is

$$\mathrm{Fgt}^{EWC} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{L_t}{2}||\boldsymbol{\theta}_T^{EWC} - \boldsymbol{\theta}_t^*||^2\right) \tag{15}$$

Comparing Eq. 9 with Eq. 15, let $\Delta = \mathrm{Fgt}^{vanilla} - \mathrm{Fgt}^{EWC}$. Using Eq. 12 to simplify Δ we have

$$\Delta \approx \frac{1}{2T}\sum_{t=1}^{T} L_t\left(||\boldsymbol{\theta}_T^* - \boldsymbol{\theta}_T^{opt}||^2 - ||\boldsymbol{\theta}_T^{EWC} - \boldsymbol{\theta}_T^{opt}||^2\right) \tag{16}$$

When $0 < \lambda \leq 1$, $\Delta > 0$ is granted, since $||\boldsymbol{\theta}_T^{EWC} - \boldsymbol{\theta}_T^{opt}||^2 \geq 0$ and $||\boldsymbol{\theta}_T^* - \boldsymbol{\theta}_T^{EWC}||^2 > 0$. Curriculum learning using EWC demonstrates reduced forgetting. If λ is appropriately chosen, specifically $\lambda = 1$, forgetting can be reduced to zero.

□

5 EXPERIMENTS

In this section we present empirical evidence that substantiates the preceding theoretical analysis of forgetting in curriculum learning. We further benchmark IV-EWC against four continual learning methods designed to mitigate forgetting in this setting.

Table 1: Evaluation on three base models across three math reasoning tasks. **Bold** and underlined values indicate the **best** and second-best results, respectively.

	GSM8K		MATH		cn_k12	
	Forget	ACC	Forget	ACC	Forget	ACC
Qwen2.5-3B-Instruct	0.0106	0.8084	0.0238	0.7191	0.0146	0.4421
+ Random Replay	0.0153	0.7989	0.0496	0.6950	0.0137	0.4380
+ EWC	0.0050	0.8258	-0.0031	0.7406	0.0172	0.4435
+ A-GEM	0.0067	0.8054	0.0011	0.7203	0.0221	0.4377
+ UPGD	-0.0008	0.8127	-0.0037	0.7366	0.0060	0.4422
+ IV-EWC	**-0.0103**	**0.8298**	**-0.0075**	**0.7461**	**-0.0095**	**0.4446**
Qwen2.5-7B-Instruct	0.0062	0.8998	0.0114	0.8389	0.0157	0.4907
+ Random Replay	0.0091	0.9025	0.0141	0.8410	0.0064	0.4944
+ EWC	0.0021	0.9018	0.0053	0.8417	-0.0001	0.4950
+ A-GEM	0.0036	0.8828	0.0224	0.8116	0.0093	0.4930
+ UPGD	0.0054	**0.9081**	0.0108	0.8265	0.0006	0.4991
+ IV-EWC	**-0.0066**	0.9042	**-0.0028**	**0.8464**	**-0.0113**	**0.5005**
Llama-3.2-3B-Instruct	0.0047	**0.7483**	0.0183	0.5446	0.0147	0.3532
+ Random Replay	0.0067	0.7454	0.0293	0.5346	0.0221	0.3502
+ EWC	0.0058	0.7434	0.0026	0.5471	0.0244	0.3538
+ A-GEM	0.0014	0.7398	0.0081	0.5316	0.0143	0.3530
+ UPGD	0.0030	0.7445	0.0111	0.5459	0.0193	0.3516
+ IV-EWC	**-0.0028**	0.7465	**-0.0007**	**0.5512**	**-0.0141**	**0.3610**

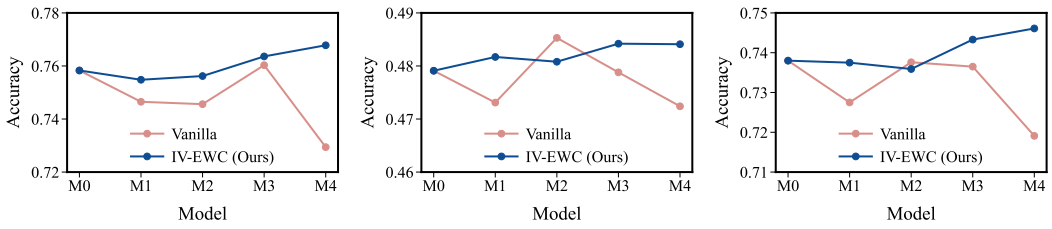
5.1 Experimental Settings

Datasets. Following previous works (Parashar et al., 2026; Xi et al., 2024), we selected two widely used math reasoning datasets: GSM8K (Cobbe et al., 2021) and MATH (Hendrycks et al., 2021). We used the classified version of GSM8K created by Parashar et al. (2026), where the difficulties of questions are evaluated by zero-shot performance. We also created a cn_k12 dataset from the NuminaMath-CoT (Li et al., 2024) dataset to test the model's ability to solve problems with diverse formats. Detailed dataset descriptions are provided in Appendix B.

Metrics. Following Jiang et al. (2025), we adopted the following metrics: (1) $\mathbf{FP} = \frac{1}{T}\sum_{j=1}^{T} a_T^j$, which is the final average performance across all tasks learned by the model, T denotes the index of the last learned task and a_T^j refers to the accuracy on the test set of task j evaluated after learning task T. (2) $\mathbf{AP} = \frac{1}{T}\sum_{j=1}^{T} a_j^j$, which summarizes the average performance measured immediately after the learning of each task. (3) $\mathbf{Forget} = \frac{FP-AP}{AP}$, which is the relative divergence between FP and AP, evaluating overall performance differences between tasks. (4) $\mathbf{ACC}$, which represents the overall accuracy of the model on the whole test set spanning all curriculum tasks, unlike FP, it accounts for the uneven distribution of test sample difficulty that can arise in certain datasets.

Baselines. To investigate the effectiveness of IV-EWC, we compared it with five baselines: (1) **vanilla curriculum**, where the model is trained directly on a sequence of curriculum tasks with the given learning rate. (2) **random replay**, where the training data includes a fixed number of selected samples from the training data of the previous task starting from the second task. (3) **EWC** (Kirkpatrick et al., 2017), where an EWC loss is added to the original loss function and it uses the same regularization weight as IV-EWC. (4) **A-GEM** (Chaudhry et al., 2019), which utilizes reference samples selected from the training set of the previous task to control gradient updates; the number of reference samples is the same as random replay. (5) **UPGD** (Elsayed & Mahmood, 2024), which uses utility to measure the significance of each parameter and instruct the learning process.

Implementation Details. In the experiments we adopted Qwen2.5-3B-Instruct, Qwen2.5-7B-Instruct and Llama-3.2-3B-Instruct as base models. The hyperparameter settings in the experiments are detailed in Appendix C. All algorithms were implemented in Python and executed using the Python 3.12 interpreter. The 3B model experiments were carried out on one NVIDIA RTX PRO 6000 and the 7B model experiments were carried out on two NVIDIA RTX PRO 6000s. To ensure reliability, experiment data are presented as the mean of four independent runs.



(a) Easy Task Accuracy (b) Hard Task Accuracy (c) Overall Accuracy

Figure 1: Accuracy of models trained after each task using vanilla curriculum and IV-EWC on the MATH dataset based on Qwen2.5-3B-Instruct. Easy task accuracy denotes the average accuracy of level 2 or lower, hard task accuracy is the average accuracy of level 3 and higher.

Table 2: Evaluation of the replay method with varying replay data using Qwen2.5-3B-Instruct.

	GSM8K		MATH		cn_k12	
	Forget	ACC	Forget	ACC	Forget	ACC
Vanilla Replay	0.0088	0.8223	0.0121	0.7375	0.0143	0.4785
+ Select Across Levels	0.0042	0.8336	0.0105	0.7359	0.0091	0.4750
+ Influence Function	0.0026	0.8336	0.0058	0.7360	0.0075	0.4775

5.2 Comparison Study

Table 1 summarizes performance for three base models over three tasks. The results indicate forgetting within the curriculum learning process, with greater model capacity corresponding to less forgetting and higher accuracy. Several approaches exhibit negative forgetting, where the performance on previously learned, easier tasks improves following training on more challenging tasks. This phenomenon is partly attributable to the inherent representation sharing in large language models and to the strong inter-level task relatedness induced by curriculum learning. IV-EWC achieves superior performance relative to all baselines on both forgetting and overall accuracy in nearly all tested configurations, with more substantial gains in smaller models. Negative forgetting is also observed in all scenarios using IV-EWC, confirming its effectiveness.

To better understand the curriculum learning process, we collected evaluation data on the test sets from models trained after each task. Figure 1 shows the grouped accuracy of vanilla curriculum and IV-EWC on the MATH dataset with Qwen2.5-3B-Instruct. While accuracy on the easy tasks deteriorates during vanilla curriculum learning, IV-EWC preserves performance on these tasks. On challenging tasks, vanilla curriculum learning exhibits pronounced fluctuations in accuracy, whereas IV-EWC maintains stable performance with a slight upward trend. Contrary to intuition, training on the hard task does not consistently yield accuracy gains on that task. We hypothesize that this pattern arises from the complex optimization dynamics of large language models and nontrivial inter-task dependencies. In comparison with the vanilla baseline, two key advantages are observed for IV-EWC: improved learning of new tasks, manifested as increasing accuracy on the hard tasks, and diminished forgetting, reflected in stable performance on easy tasks. More details about the performance during curriculum learning can be found in Appendix D.

5.3 Ablation Study

To assess the effectiveness of the validation set used in IV-EWC, an ablation experiment was conducted. This experiment used the replay method with Qwen2.5-3B-Instruct. To avoid overfitting, the learning rate was set to 3e-8, 5.5e-8 and 1e-7 for GSM8K, MATH and cn_k12 datasets respectively. The replay sample count is incremented by 100 with each subsequent task, yielding 100, 200, 300, and 400 samples for Tasks 1 to 4, respectively. Vanilla replay is the baseline method, which randomly selects 100 samples from task $t-1$ as replay samples for task t; replay samples are repeated to match the size of the training set. Select across levels method keeps the previously selected replay samples and adds 100 random samples of the previous task into replay before the

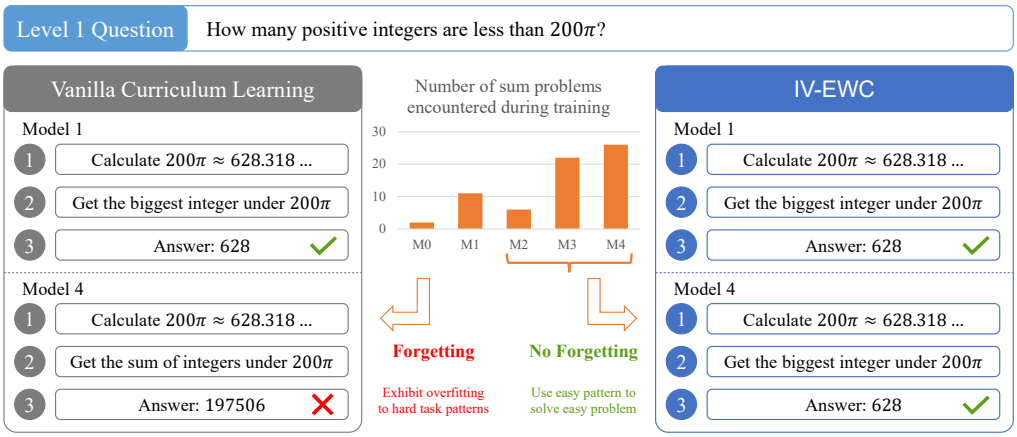


Figure 2: Case study on a MATH test example comparing Qwen2.5-3B-Instruct models trained via vanilla curriculum learning and via IV-EWC. Following training on tasks from Level 2 to Level 4, the vanilla curriculum model shows overfitting to hard-task patterns, while the model trained with IV-EWC avoids such overfitting.

current task. Influence function method directly uses the validation set of the corresponding task in IV-EWC for replay. The experiment results are illustrated in Table 2.

Keeping replay samples across levels decreases forgetting by roughly 0.0038 (34.0% of vanilla replay) on average. On this basis, using the validation set as replay samples can further decrease forgetting by about 0.0026 (22.7% of vanilla replay) on average. Results indicate that replaying an aggregated set of training examples drawn across curriculum tasks, alongside replaying examples selected via influence functions, better preserves knowledge acquired in earlier tasks and thereby reduces forgetting; employing both strategies in combination yields the strongest performance.

5.4 Case Study

To examine forgetting in curriculum learning and assess how IV-EWC mitigates it, we collected responses of models at different training stages. Figure 2 presents side-by-side the responses of the Qwen2.5-3B-Instruct model, trained on MATH with vanilla curriculum learning versus IV-EWC, for a representative problem from the MATH test set. After training on the Level 1 task, both models acquire the appropriate strategy for solving the task. After learning on the task with Level 4 data, the model trained with vanilla curriculum learning erroneously applies the heuristic for sum problems to the example, yielding an incorrect prediction. This misgeneralization is partly attributable to the training data distribution: during the training on tasks from Level 2 to Level 4, the model encounters more sum problems. Because vanilla curriculum learning introduces no explicit regularization, the model overfits to patterns characteristic of the more difficult tasks. In contrast, the model trained via IV-EWC resolves the task using the task-relevant pattern, suggesting that regularization limits overfitting to the solution approaches of challenging curriculum tasks and thus mitigates forgetting. Additional cases from the study are provided in Appendix F.

6 Conclusion

In this paper we employ both theoretical and empirical approaches to analyze the phenomenon of catastrophic forgetting in curriculum learning and propose a novel framework IV-EWC, which incorporates EWC regularization into curriculum learning. IV-EWC utilizes the influence function to select training samples, forming a validation set for regularization. A dynamic regularization mechanism is further applied to facilitate effective learning. Theoretical investigation of EWC-based methods provides a foundation for IV-EWC in eliminating forgetting within curriculum learning. Empirical results under multiple settings demonstrate that IV-EWC consistently outperforms baselines, highlighting its effectiveness.

AI USE STATEMENT

In this work, we used generative AI tools for assisting with translation. We have not used generative AI tools for helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, implementing methods, cleaning and reformatting datasets, supporting qualitative and thematic data analysis, interpreting results. Generating synthetic data sets is not applicable to this work. We have reviewed all AI-assisted work. We have carefully checked the translations generated by AI tools to ensure that the text reflects the real ideas authors would like to convey to readers. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Naman Agarwal, Brian Bullins, and Elad Hazan. Second-order stochastic optimization for machine learning in linear time. *J. Mach. Learn. Res.*, 18(1):4148–4187, January 2017. ISSN 1532-4435.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pp. 41–48, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161. doi: 10.1145/1553374.1553380. URL https://doi.org/10.1145/1553374.1553380.

Xufeng Cai and Jelena Diakonikolas. Last iterate convergence of incremental methods as a model of forgetting. In *International Conference on Learning Representations*, volume 2025, pp. 102613–102647, 2025.

Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. In *International Conference on Learning Representations*, 2019.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems, 2021. URL https://arxiv.org/abs/2110.14168.

Mohamed Elsayed and A. Rupam Mahmood. Addressing loss of plasticity and catastrophic forgetting in continual learning. In *International Conference on Learning Representations*, 2024.

Carlos Florensa, David Held, Markus Wulfmeier, Michael Zhang, and Pieter Abbeel. Reverse curriculum generation for reinforcement learning. In Sergey Levine, Vincent Vanhoucke, and Ken Goldberg (eds.), *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78 of *Proceedings of Machine Learning Research*, pp. 482–495. PMLR, 13–15 Nov 2017. URL https://proceedings.mlr.press/v78/florensa17a.html.

Robert M. French. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4):128–135, 1999. ISSN 1364-6613. doi: https://doi.org/10.1016/S1364-6613(99)01294-2. URL https://www.sciencedirect.com/science/article/pii/S1364661399012942.

Alex Graves, Marc G. Bellemare, Jacob Menick, Rémi Munos, and Koray Kavukcuoglu. Automated curriculum learning for neural networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1311–1320. JMLR.org, 2017.

Sheng Guo, Weilin Huang, Haozhi Zhang, Chenfan Zhuang, Dengke Dong, Matthew R. Scott, and Dinglong Huang. Curriculumnet: Weakly supervised learning from large-scale web images. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (eds.), *Computer Vision – ECCV 2018*, pp. 139–154, Cham, 2018. Springer International Publishing. ISBN 978-3-030-01249-6.

Guy Hacohen and Daphna Weinshall. On the power of curriculum learning in training deep networks. In *International Conference on Machine Learning*, 2019.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Phil A Hetherington and Mark S Seidenberg. Is there "catastrophic interference" in connectionist networks? In *11th Annual Conference Cognitive Science Society Pod*, pp. 26–33. Psychology Press, 2014.

Gangwei Jiang, Zhaoyi Li, Siqiao Xue, JUN ZHOU, Linqi Song, Defu Lian, Ying Wei, et al. Unlocking the power of function vectors for characterizing and mitigating catastrophic forgetting in continual instruction tuning. In *International Conference on Learning Representations*, volume 2025, pp. 47218–47243, 2025.

Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. MentorNet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2304–2313. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/jiang18c.html.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. doi: 10.1073/pnas.1611835114. URL https://www.pnas.org/doi/abs/10.1073/pnas1611835114.

Tom Kocmi and Ondřej Bojar. Curriculum learning and minibatch bucketing in neural machine translation. In Ruslan Mitkov and Galia Angelova (eds.), *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pp. 379–386, Varna, Bulgaria, September 2017. INCOMA Ltd. doi: 10.26615/978-954-452-049-6_050. URL https://aclanthology.org/R17-1050/.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1885–1894. JMLR.org, 2017.

M. Pawan Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. In *Proceedings of the 24th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'10, pp. 1189–1197, Red Hook, NY, USA, 2010. Curran Associates Inc.

Lars Leyendecker, Markus Schmitz, Hans Aoyang Zhou, Vladimir Samsonov, Marius Rittstieg, and Daniel Lütticke. Deep reinforcement learning for robotic control in high-dexterity assembly tasks - a reward curriculum approach. In *2021 Fifth IEEE International Conference on Robotic Computing (IRC)*, pp. 35–42, 2021. doi: 10.1109/IRC52146.2021.00012.

Jia Li, Edward Beeching, Lewis Tunstall, Ben Lipkin, Roman Soletskyi, Shengyi Costa Huang, Kashif Rasul, Longhui Yu, Albert Jiang, Ziju Shen, Zihan Qin, Bin Dong, Li Zhou, Yann Fleureau, Guillaume Lample, and Stanislas Polu. Numinamath. [https://huggingface.co/datasets/AI-MO/NuminaMath-CoT](https://github.com/project-numina/aimo-progress-prize/blob/main/report/numina_dataset.pdf), 2024.

Fenglin Liu, Shen Ge, and Xian Wu. Competence-based multimodal curriculum learning for medical report generation. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 3001–3012, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.234. URL https://aclanthology.org/2021.acl-long.234/.

Xuebo Liu, Houtim Lai, Derek F. Wong, and Lidia S. Chao. Norm-based curriculum learning for neural machine translation. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault

(eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 427–436, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.41. URL https://aclanthology.org/2020.acl-main.41/.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6470–6479, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Tambet Matiisen, Avital Oliver, Taco Cohen, and John Schulman. Teacher-student curriculum learning. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9):3732–3740, 2020. doi: 10.1109/TNNLS.2019.2934906.

Michael McCloskey and Neal J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In Gordon H. Bower (ed.), *Psychology of Learning and Motivation*, volume 24, pp. 109–165. Academic Press, 1989. doi: https://doi.org/10.1016/S0079-7421(08)60536-8. URL https://www.sciencedirect.com/science/article/pii/S0079742108605368.

Shubham Parashar, Shurui Gui, Xiner Li, Hongyi Ling, Sushil Vemuri, Blake Olson, Eric Li, Yu Zhang, James Caverlee, Dileep Kalathil, et al. Curriculum reinforcement learning from easy to hard tasks improves llm reasoning. In *International Conference on Learning Representations*, volume 2026, pp. 156944–156972, 2026.

Emmanouil Antonios Platanios, Otilia Stretcu, Graham Neubig, Barnabas Poczos, and Tom Mitchell. Competence-based curriculum learning for neural machine translation. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 1162–1172, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1119. URL https://aclanthology.org/N19-1119/.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. icarl: Incremental classifier and representation learning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5533–5542, 2017. doi: 10.1109/CVPR.2017.587.

David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy P. Lillicrap, and Greg Wayne. Experience replay for continual learning. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019. Curran Associates Inc.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024. URL https://arxiv.org/abs/2402.03300.

Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 2994–3003, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Sebastian Thrun. Is learning the n-th thing any easier than learning the first? In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, NIPS'95, pp. 640–646, Cambridge, MA, USA, 1995. MIT Press.

Rui Wang, Joel Lehman, Jeff Clune, and Kenneth O. Stanley. Paired open-ended trailblazer (poet): Endlessly generating increasingly complex and diverse learning environments and their solutions, 2019. URL https://arxiv.org/abs/1901.01753.

Rui Wang, Joel Lehman, Aditya Rawal, Jiale Zhi, Yulun Li, Jeff Clune, and Kenneth O. Stanley. Enhanced poet: open-ended reinforcement learning through unbounded invention of learning challenges and their solutions. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

Liang Wen, Yunke Cai, Fenrui Xiao, Xin He, Qi An, Zhenyu Duan, Yimin Du, Junchen Liu, Lifu Tang, Xiaowei Lv, Haosheng Zou, Yongchao Deng, Shousheng Jia, and Xiangzheng Zhang. Light-r1: Curriculum SFT, DPO and RL for long COT from scratch and beyond. In Georg Rehm and Yunyao Li (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, pp. 318–327, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-288-6. doi: 10.18653/v1/2025.acl-industry.24. URL https://aclanthology.org/2025.acl-industry.24/.

Zhiheng Xi, Wenxiang Chen, Boyang Hong, Senjie Jin, Rui Zheng, Wei He, Yiwen Ding, Shichun Liu, Xin Guo, Junzhe Wang, Honglin Guo, Wei Shen, Xiaoran Fan, Yuhao Zhou, Shihan Dou, Xiao Wang, Xinbo Zhang, Peng Sun, Tao Gui, Qi Zhang, and Xuanjing Huang. Training large language models for reasoning through reverse curriculum reinforcement learning. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Linhui Xiao, Xiaoshan Yang, Fang Peng, Ming Yan, Yaowei Wang, and Changsheng Xu. Clip-vg: Self-paced curriculum adapting of clip for visual grounding. *IEEE Transactions on Multimedia*, 26:4334–4347, 2024. doi: 10.1109/TMM.2023.3321501.

Xiang Zhang, Jiaqi Wei, Zijie Qiu, Sheng Xu, Nanqing Dong, Zhiqiang Gao, and Siqi Sun. Curriculum learning for biological sequence prediction: the case of de novo peptide sequencing. In *Proceedings of the 42nd International Conference on Machine Learning*, ICML'25. JMLR.org, 2025.

Xuan Zhang, Pamela Shapiro, Gaurav Kumar, Paul McNamee, Marine Carpuat, and Kevin Duh. Curriculum learning for domain adaptation in neural machine translation. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 1903–1915, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1189. URL https://aclanthology.org/N19-1189/.

A Additional Theoretical Analysis

A.1 Parameter Distance of Gradient Descent

Lemma 1. *For the target problem with a μ-strongly convex and L-smooth loss, let t be the optimization step, $\kappa = \frac{L}{\mu}$ be the condition number of the loss, and $\boldsymbol{\theta}^t$, $\boldsymbol{\theta}^*$, $\boldsymbol{\theta}_0$ be the parameter at step t, the optimal parameter of the target problem and the initial parameter of the target problem, respectively. The distance between parameters follows:*

$$||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 \leq \left(1 - \frac{1}{\kappa}\right)^t ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||^2 \tag{1}$$

Proof. Define the loss of the target problem as f, it is μ-strongly convex and L-smooth so the following inequalities hold for any x and y:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu ||x - y||^2 \tag{17}$$

$$||\nabla f(x) - \nabla f(y)||^2 \leq L \langle \nabla f(x) - \nabla f(y), x - y \rangle \tag{18}$$

Consider gradient descent with constant step size $\frac{1}{L}$, the parameter update follows:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \frac{1}{L} \nabla f(\boldsymbol{\theta}_t)$$

$$||\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*||^2 = ||\boldsymbol{\theta}_t - \frac{1}{L} \nabla f(\boldsymbol{\theta}_t) - \boldsymbol{\theta}^*||^2 \tag{19}$$

Note that $\nabla f(\boldsymbol{\theta}^*) = 0$, Eq. 19 can be written as:

$$\begin{aligned} ||\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*||^2 &= ||\boldsymbol{\theta}_t - \boldsymbol{\theta}^* - \frac{1}{L}\left(\nabla f(\boldsymbol{\theta}_t) - \nabla f(\boldsymbol{\theta}^*)\right)||^2 \\ &= ||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 - \frac{2}{L}\langle \nabla f(\boldsymbol{\theta}_t) - \nabla f(\boldsymbol{\theta}^*), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle + \frac{1}{L^2}||\nabla f(\boldsymbol{\theta}_t) - \nabla f(\boldsymbol{\theta}^*)||^2 \end{aligned} \tag{20}$$

Using Eq. 18 to replace $\frac{1}{L^2}||\nabla f(\boldsymbol{\theta}_t) - \nabla f(\boldsymbol{\theta}^*)||^2$ we have:

$$||\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*||^2 \leq ||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 - \frac{1}{L}\langle \nabla f(\boldsymbol{\theta}_t) - \nabla f(\boldsymbol{\theta}^*), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle \tag{21}$$

Using Eq. 17 to replace $\frac{1}{L}\langle \nabla f(\boldsymbol{\theta}_t) - \nabla f(\boldsymbol{\theta}^*), \boldsymbol{\theta}_t - \boldsymbol{\theta}^* \rangle$, we have:

$$\begin{aligned} ||\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}^*||^2 &\leq ||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 - \frac{\mu}{L}||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 \\ &= \left(1 - \frac{\mu}{L}\right)||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 \\ &= \left(1 - \frac{1}{\kappa}\right)||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 \end{aligned} \tag{22}$$

When the iteration proceeds from 1 to t, the distance between parameters satisfies:

$$\begin{aligned} ||\boldsymbol{\theta}_t - \boldsymbol{\theta}^*||^2 &\leq \left(1 - \frac{1}{\kappa}\right)||\boldsymbol{\theta}_{t-1} - \boldsymbol{\theta}^*||^2 \\ &\leq \left(1 - \frac{1}{\kappa}\right)^2||\boldsymbol{\theta}_{t-2} - \boldsymbol{\theta}^*||^2 \\ &\leq ... \\ &\leq \left(1 - \frac{1}{\kappa}\right)^{t-1}||\boldsymbol{\theta}_1 - \boldsymbol{\theta}^*||^2 \\ &\leq \left(1 - \frac{1}{\kappa}\right)^{t}||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||^2 \end{aligned} \tag{23}$$

□

A.2 Iteration Counts of Gradient Descent

Lemma 2. *For the target problem with a μ-strongly convex and L-smooth loss, define the convergence condition $||\boldsymbol{\theta}_T - \boldsymbol{\theta}^*|| \leq \varepsilon$ where $\boldsymbol{\theta}^T$ and $\boldsymbol{\theta}^*$ are the parameter at convergence and the optimal parameter of the target problem respectively, let $C > 0$ be a constant and κ be the condition number of the loss. The number of iterations required by gradient descent can be expressed as:*

$$T \approx C\kappa \log\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}{\varepsilon}\right) \tag{2}$$

Proof. From the upper bound of parameter distance in Lemma 1, when the gradient descent with constant step size $\frac{1}{L}$ converges at iteration T, the following should hold:

$$||\boldsymbol{\theta}_T - \boldsymbol{\theta}^*||^2 \leq \left(1 - \frac{1}{\kappa}\right)^T||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||^2 \tag{24}$$

Taking square roots for both sides yields

$$||\boldsymbol{\theta}_T - \boldsymbol{\theta}^*|| \leq \left(1 - \frac{1}{\kappa}\right)^{\frac{T}{2}}||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*|| \tag{25}$$

The convergence condition is defined as $||\boldsymbol{\theta}_T - \boldsymbol{\theta}^*|| \leq \varepsilon$. To achieve this the following should hold:

$$\left(1 - \frac{1}{\kappa}\right)^{\frac{T}{2}} ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*|| \leq \varepsilon$$
$$\left(1 - \frac{1}{\kappa}\right)^{\frac{T}{2}} \leq \frac{\varepsilon}{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||} \tag{26}$$

Taking the logarithm of both sides, we have

$$\frac{T}{2} \ln\left(1 - \frac{1}{\kappa}\right) \leq \ln\left(\frac{\varepsilon}{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}\right) \tag{27}$$

Since $\kappa \geq 1$, we have $0 \leq 1 - \frac{1}{\kappa} < 1$, taking the negative of both sides gives

$$\frac{T}{2} \ln\left(\frac{\kappa}{\kappa - 1}\right) \geq \ln\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}{\varepsilon}\right)$$
$$\frac{T}{2} \geq \frac{\ln\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}{\varepsilon}\right)}{\ln\left(\frac{\kappa}{\kappa-1}\right)} \tag{28}$$

Since $\ln(x) \leq x - 1$ holds for any x, $\ln\left(\frac{\kappa}{\kappa-1}\right) \leq \frac{\kappa}{\kappa-1} - 1 = \frac{1}{\kappa-1}$, and Eq. 28 can be written as:

$$\frac{T}{2} \geq \frac{\ln\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}{\varepsilon}\right)}{\frac{1}{\kappa-1}}$$
$$T \geq 2(\kappa - 1) \ln\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}{\varepsilon}\right) \tag{29}$$

The iteration complexity to reach $||\boldsymbol{\theta}_T - \boldsymbol{\theta}^*|| \leq \varepsilon$ is $\Theta\left(\kappa \log\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}{\varepsilon}\right)\right)$, with the hidden constants depending on the step size choice and whether one measures distance or function suboptimality. It is in correspondence with Eq. 2.

□

A.3 Effectiveness of Curriculum Learning

Theorem 1. *Given the convergence condition of the easier curriculum problem $||\boldsymbol{\theta}_{T_e} - \boldsymbol{\theta}_e^*|| \leq \varepsilon_e$ and the convergence condition of the original problem $||\boldsymbol{\theta}_T - \boldsymbol{\theta}^*|| \leq \varepsilon$, for curriculum satisfying $||\boldsymbol{\theta}_e^* - \boldsymbol{\theta}^*|| \ll ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*|| - \varepsilon_e$ and $\kappa_e < \kappa$, curriculum learning requires fewer iterations to converge compared with learning at once.*

Proof. After convergence on the easier curriculum task, the distance between the current model parameter and the optimal parameter of the original problem is bounded via the triangle inequality by the sum of the distances to the curriculum optimum and between the two optima:

$$||\boldsymbol{\theta}_{T_e} - \boldsymbol{\theta}^*|| \leq ||\boldsymbol{\theta}_{T_e} - \boldsymbol{\theta}_e^*|| + ||\boldsymbol{\theta}_e^* - \boldsymbol{\theta}^*|| \leq \varepsilon_e + \delta \tag{30}$$

where $\delta = ||\boldsymbol{\theta}_e^* - \boldsymbol{\theta}^*||$ denotes the distance between the two optima. The distance in Eq. 30 equals the distance between the initial and optimal parameters for the original problem when its optimization starts in the curriculum.

Next we consider the numbers of iterations T_e and T_f for the easier problem and the original problem in the curriculum respectively. Using Lemma 2, we have:

$$T_e \approx C\kappa_e \log\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_e^*||}{\varepsilon_e}\right) \tag{31}$$

$$T_f \approx C\kappa \log\left(\frac{||\boldsymbol{\theta}_{T_e} - \boldsymbol{\theta}^*||}{\varepsilon}\right) \lesssim C\kappa \log\left(\frac{\varepsilon_e + \delta}{\varepsilon}\right) \tag{32}$$

The total number of iterations in curriculum learning includes the iterations of the easier curriculum problem and the original problem, which is

$$T_{curriculum} = T_e + T_f \tag{33}$$

Vanilla curriculum learning directly optimizes the original problem. The number of iterations $T_{vanilla}$ satisfies Lemma 2. Curriculum learning's effectiveness lies in its faster convergence, i.e., fewer iterations. To prove the effectiveness of curriculum learning, the following should hold:

$$T_{curriculum} - T_{vanilla} < 0 \tag{34}$$

That is to prove:

$$C\kappa_e \log\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_e^*||}{\varepsilon_e}\right) + C\kappa \log\left(\frac{\varepsilon_e + \delta}{\varepsilon}\right) - C\kappa \log\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}{\varepsilon}\right) < 0 \tag{35}$$

Expanding the logarithm, that is to prove:

$$C\kappa \log\left(\frac{\varepsilon_e + \delta}{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||}\right) + C\kappa_e \log\left(\frac{||\boldsymbol{\theta}_0 - \boldsymbol{\theta}_e^*||}{\varepsilon_e}\right) < 0 \tag{36}$$

the left-hand side of the inequality comprises two terms: the first reflects the reduction in iteration count after introducing curriculum learning, the second corresponds to the iterations required to learn the easier curriculum problem.

According to the curriculum design $||\boldsymbol{\theta}_e^* - \boldsymbol{\theta}^*|| \ll ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*|| - \varepsilon_e$, using δ to substitute the left part we have $\varepsilon_e + \delta \ll ||\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*||$. The first term on the left side of Eq. 36 is negative. Furthermore, the curriculum satisfies $\kappa_e < \kappa$ so Eq. 36 always holds; the given curriculum converges faster than learning at once.

□

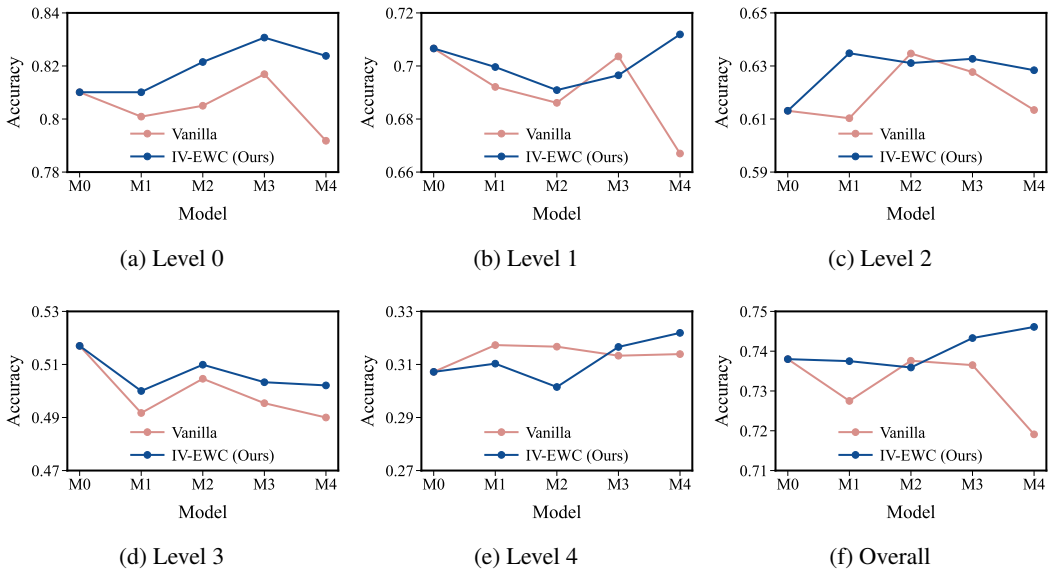
B Dataset Details

In this section, we provide more information about datasets used in the experiments. Three math reasoning datasets are adopted to evaluate the model, whose details are described as follows.

The GSM8K dataset (Cobbe et al., 2021) contains easy mathematical questions with their contexts from daily lives. In the experiments we adopt the version by Parashar et al. (2026), where the questions are graded and split into four levels. The training set contains 7473 samples with 1630, 1747, 2327, 1769 samples at Level 0 to Level 3 respectively. The test set contains 1319 samples with 206, 300, 354, 459 samples at Level 0 to Level 3 respectively.

The MATH dataset proposed by Hendrycks et al. (2021) contains five difficulty levels. The majority of questions in the MATH dataset do not have a context, which makes understanding easier. The training set includes 7498 samples: 564 at Level 0, 1348 at Level 1, 1592 at Level 2, 1690 at Level 3 and 2304 at Level 4. The test set includes 5000 samples: 437 at Level 0, 894 at Level 1, 1131 at Level 2, 1214 at Level 3 and 1324 at Level 4. The training distribution is skewed toward challenging examples, which promotes the acquisition of richer representations required to handle such cases.

The cn_k12 dataset is created by selecting samples from the cn_k12 subset of NuminaMath-CoT (Li et al., 2024). To align with existing datasets, we cleaned the cn_k12 subset by removing samples with links, non-ASCII characters, inconsistent question types, omitted or incomplete answers. The cleaned data was then categorized into four difficulty levels by answer length, namely: Level 1 (1-179 characters), Level 2 (180-259 characters), Level 3 (260-369 characters), and Level 4 (370-800 characters). Samples with answers longer than 800 characters are dropped as outliers. We sampled 2000 training and 700 testing instances from each level to form the final cn_k12 dataset.



(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Level 4 (f) Overall

Figure 3: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on the MATH dataset based on Qwen2.5-3B-Instruct, evaluated on the test sets of the five curriculum tasks.

C Hyperparameter Details

In this section, we list all hyperparameter settings for methods in the experiments. Moreover, we tuned the learning rates for different datasets on diverse models, the adopted learning rates are also reported in this section.

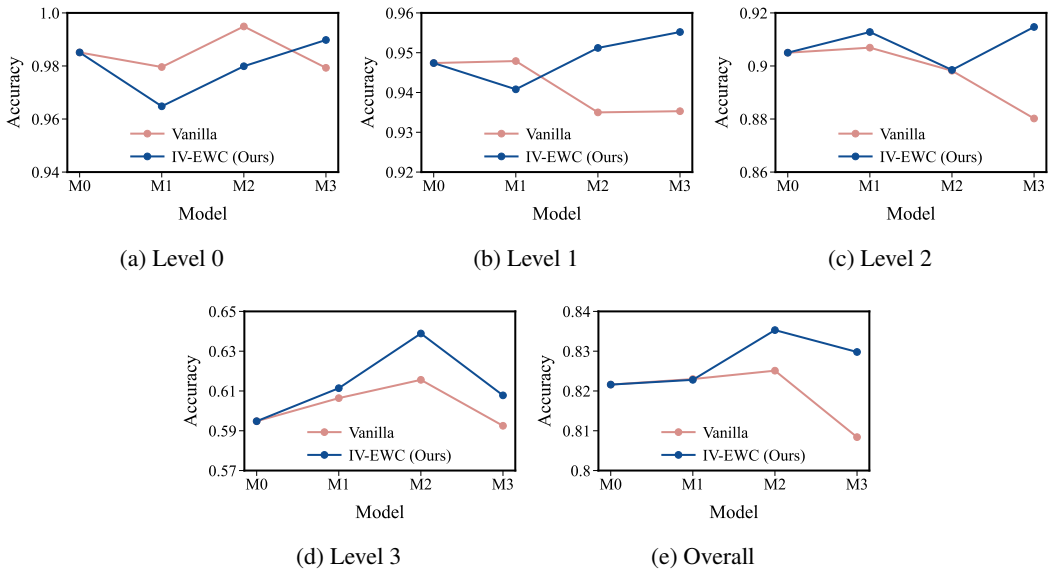
For replay-based methods (random replay and A-GEM), the number of selected replay samples or reference samples was 100, in correspondence with the validation set increment 100 of IV-EWC. For regularization-based methods (EWC and IV-EWC), we keep the same regularization weight across methods, adding dynamic regularization to EWC. The initial regularization weight γ was set to 1.2e8. For UPGD method, we set hyperparameter sigma to zero, disabling random noise to align with other methods. For IV-EWC, LiSSA recursion depth used to compute influence was fixed at 50. To find the appropriate selections of hyperparameters for IV-EWC, a parameter sensitivity analysis was carried out with details in Appendix E.

To avoid overfitting to the math reasoning tasks, which is reflected by the sharp decrease in all curriculum tasks' accuracy, we adjust the learning rate and reduce it for easier datasets and smaller models. For the GSM8K dataset, the learning rates were 1.0e-7 for Qwen2.5-3B-Instruct, 1.5e-7 for Qwen2.5-7B-Instruct, and 3.0e-8 for Llama-3.2-3B-Instruct. For the MATH dataset, the learning rates were 5.5e-8 for Qwen2.5-3B-Instruct, 8.0e-8 for Qwen2.5-7B-Instruct, and 1.0e-8 for Llama-3.2-3B-Instruct. For the cn_k12 dataset, the learning rate was set to 1.5e-7 for all base models.

D Additional Results during Curriculum Learning

In this section, we provide detailed accuracy results during the curriculum learning process of models. The results include accuracies of models trained after each curriculum task, evaluated on the test sets of all curriculum tasks. To ensure clarity, the data are sorted by the trained base model.

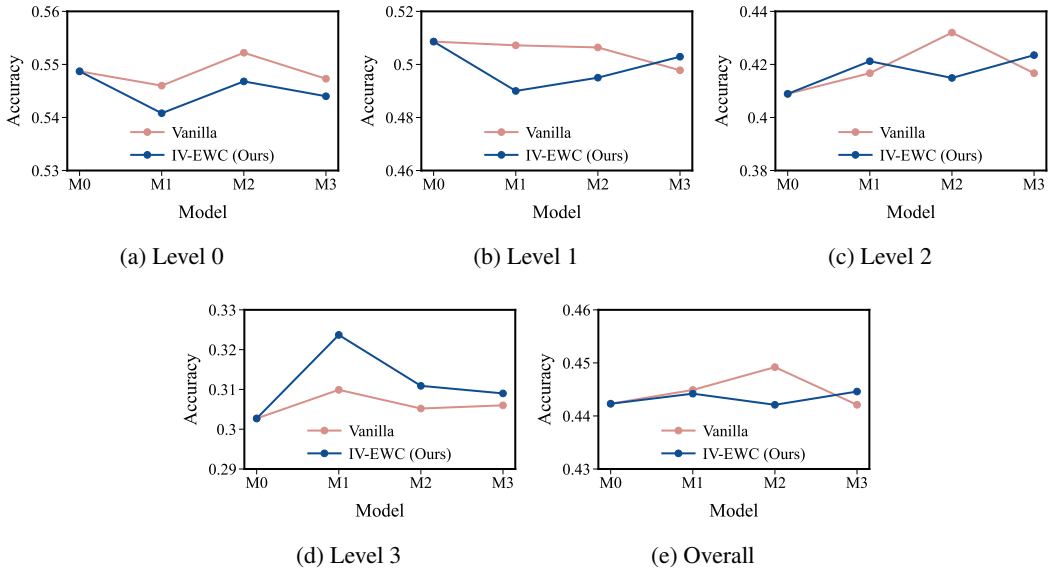
Figure 3 shows the results of vanilla curriculum learning and IV-EWC on the MATH dataset with Qwen2.5-3B-Instruct. In the later stages of training, IV-EWC surpasses the vanilla baseline in overall accuracy. During curriculum learning for mathematical reasoning, training on a single task does not necessarily translate into substantial accuracy gains for that task and can also alter performance on other tasks, which is in correspondence with the results of grouped accuracy.



(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Overall

Figure 4: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on GSM8K dataset based on Qwen2.5-3B-Instruct, evaluated on the test sets of the four curriculum tasks.



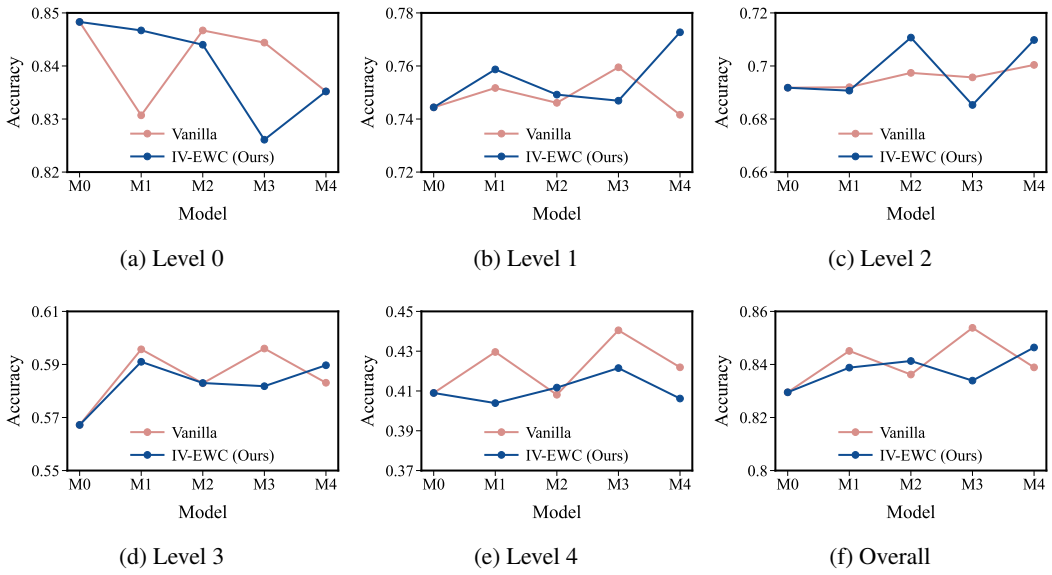
(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Overall

Figure 5: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on cn_k12 dataset based on Qwen2.5-3B-Instruct, evaluated on the test sets of the four curriculum tasks.

Figure 4 reports GSM8K results with Qwen2.5-3B-Instruct for the vanilla curriculum baseline and IV-EWC. Given the relative simplicity of GSM8K, curriculum learning on this dataset is prone to overfitting. Across the tasks of Level 1 and Level 2, the baseline accuracy declines as training proceeds, while IV-EWC keeps accuracy variations within a narrow range.

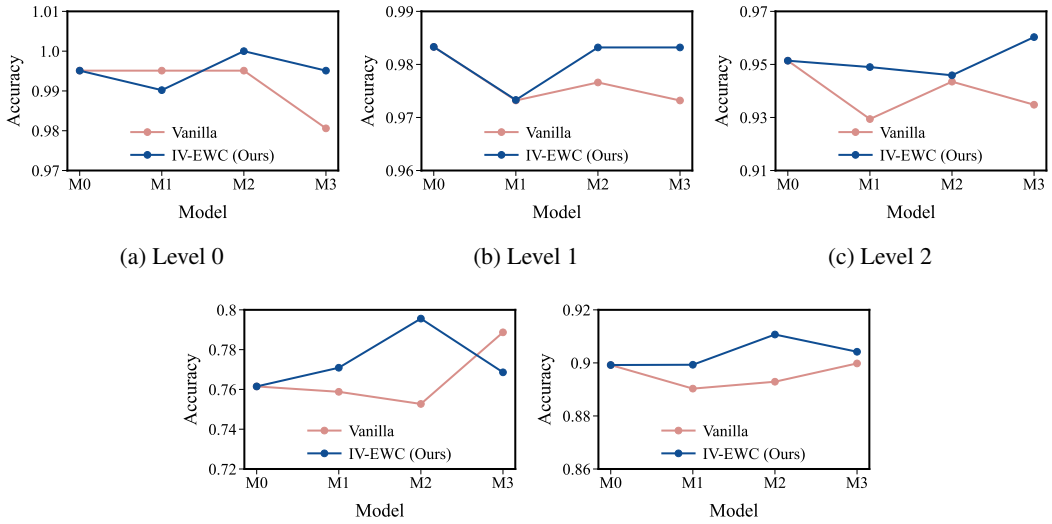
Figure 5 presents results for the vanilla curriculum and IV-EWC on the cn_k12 dataset using Qwen2.5-3B-Instruct. Both methods exhibit forgetting, evidenced by declining accuracy on the easier tasks. For the harder tasks, accuracy improves gradually as the curriculum progresses, indicating



(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Level 4 (f) Overall

Figure 6: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on the MATH dataset based on Qwen2.5-7B-Instruct, evaluated on the test sets of the five curriculum tasks.



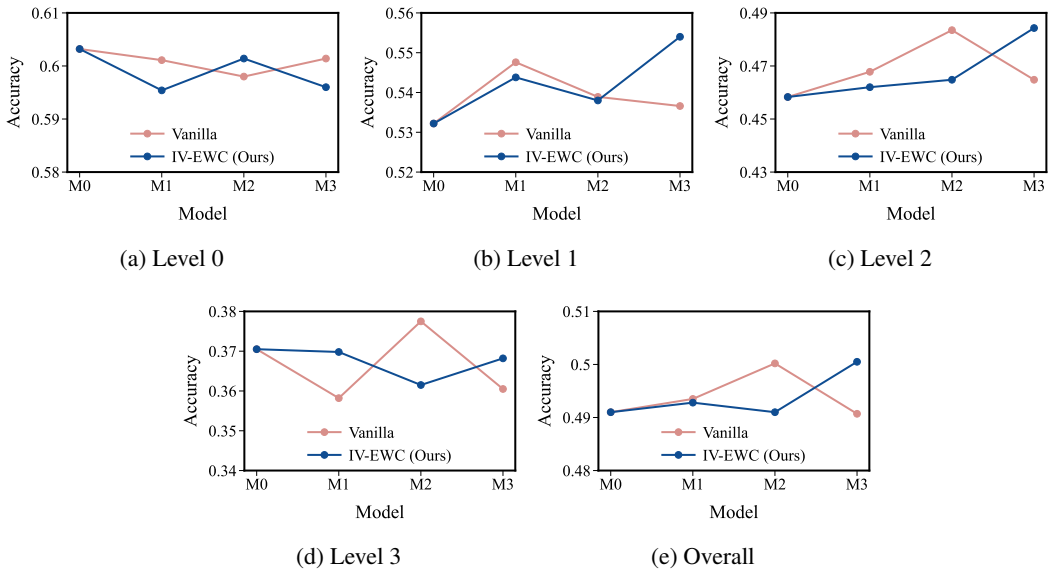
(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Overall

Figure 7: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on GSM8K dataset based on Qwen2.5-7B-Instruct, evaluated on the test sets of the four curriculum tasks.

that curriculum learning facilitates acquisition of new tasks. At the conclusion of the curriculum, IV-EWC outperforms the vanilla approach, with gains primarily on the harder tasks.

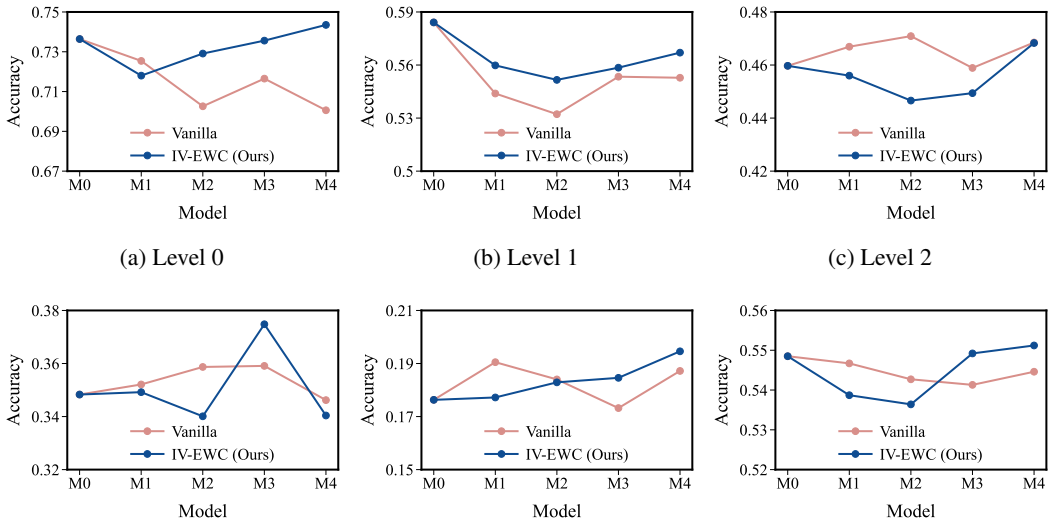
Figure 6 presents the performance of the vanilla curriculum and IV-EWC on the MATH dataset using Qwen2.5-7B-Instruct. Compared with the Qwen2.5-3B-Instruct setting, the training dynamics are less stable and the overall trend is less pronounced. Nevertheless, accuracy on previously learned tasks declines after training on each new task. This decline is smaller under IV-EWC than under the vanilla curriculum, indicating reduced forgetting for IV-EWC.



(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Overall

Figure 8: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on cn_k12 dataset based on Qwen2.5-7B-Instruct, evaluated on the test sets of the four curriculum tasks.



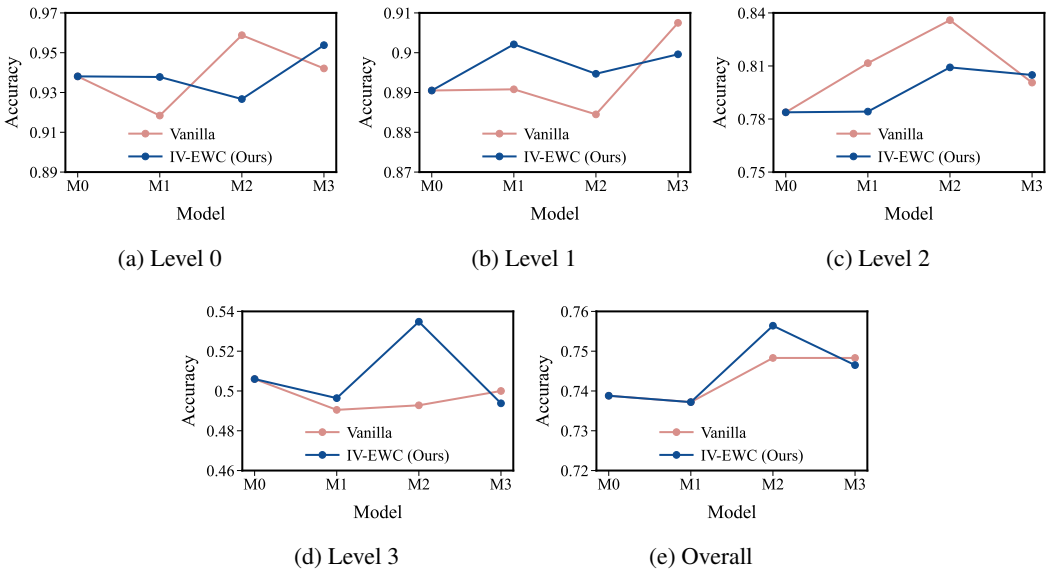
(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Level 4 (f) Overall

Figure 9: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on the MATH dataset based on Llama-3.2-3B-Instruct, evaluated on the test sets of the five curriculum tasks.

Figure 7 compares the vanilla curriculum against IV-EWC on the GSM8K dataset using Qwen2.5-7B-Instruct. By leveraging the reasoning capability of a larger model, both approaches achieve test accuracy consistently above 0.89. Given this strong baseline, incorporating regularization further improves accuracy and reduces forgetting relative to curriculum learning without regularization.

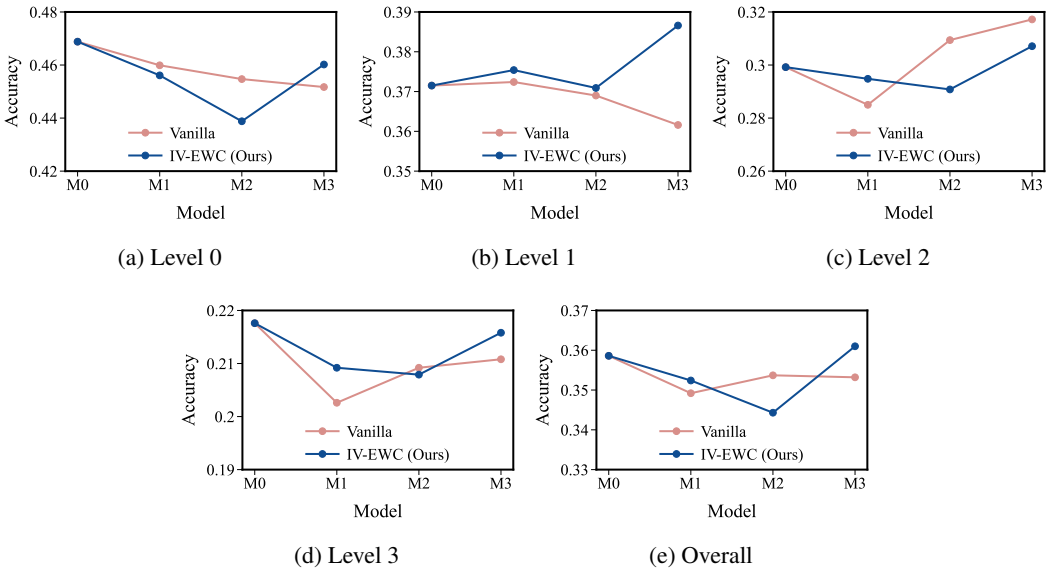
Figure 8 presents results for the vanilla curriculum baseline and IV-EWC on the cn_k12 dataset with Qwen2.5-7B-Instruct. For the baseline, test accuracy on a specific task improves during its



(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Overall

Figure 10: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on GSM8K dataset based on Llama-3.2-3B-Instruct, evaluated on the test sets of the four curriculum tasks.



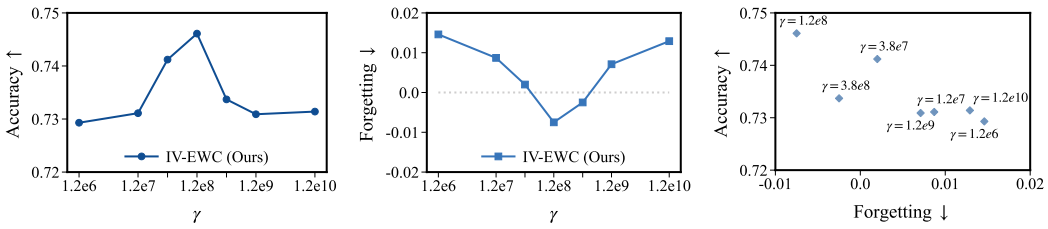
(a) Level 0 (b) Level 1 (c) Level 2

(d) Level 3 (e) Overall

Figure 11: Accuracy of models trained after each task using vanilla curriculum (dashed line) and IV-EWC (solid line) on cn_k12 dataset based on Llama-3.2-3B-Instruct, evaluated on the test sets of the four curriculum tasks.

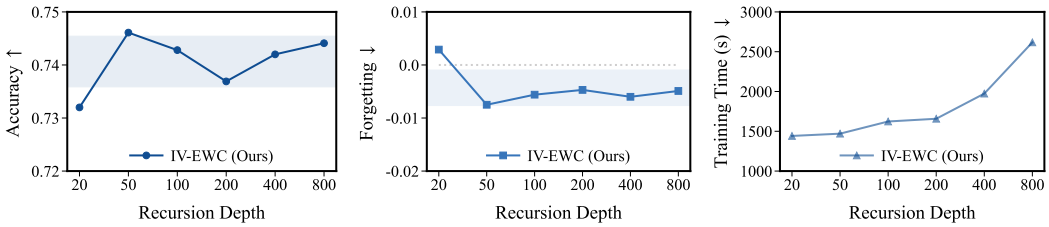
training phase but deteriorates as training proceeds to subsequent tasks. With regularization, IV-EWC achieves and sustains higher accuracy on previously learned tasks, mitigating forgetting.

Figure 9 presents results on the MATH dataset with Llama-3.2-3B-Instruct, comparing a vanilla curriculum learning baseline to IV-EWC. Relative to Qwen, Llama exhibits lower overall performance, likely due to differences in pretraining (Qwen models are trained on math-reasoning data). We further observe performance degradation after training on the current curriculum task. Notably,



(a) Accuracy (b) Forgetting (c) Trade-Off View

Figure 12: Performance of IV-EWC with different initial regularization weight γ using Qwen2.5-3B-Instruct on the MATH dataset.



(a) Accuracy (b) Forgetting (c) Training Time

Figure 13: Accuracy, forgetting and training time of IV-EWC with diverse recursive depths for LiSSA using Qwen2.5-3B-Instruct on the MATH dataset. Shaded areas of accuracy and forgetting indicate mean $\pm$ SD of all data points.

beginning at the task of Level 3, the ranking reverses and IV-EWC outperforms the baseline, suggesting that its benefits are more pronounced on higher-difficulty tasks.

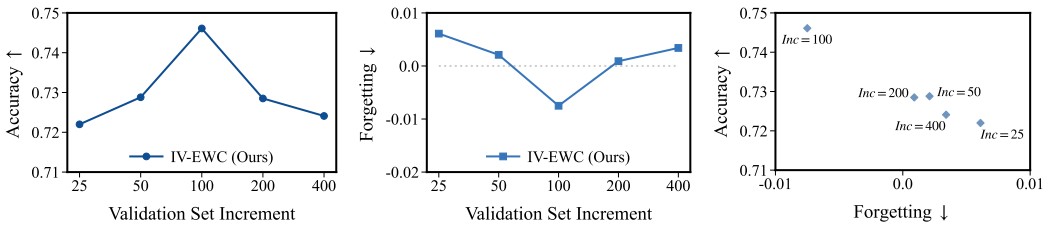
Figure 10 shows the results of vanilla curriculum and IV-EWC on GSM8K dataset with Llama-3.2-3B-Instruct. Vanilla curriculum learning exhibits limited forgetting under this setting, likely owing to the homogeneity of the levels. Nevertheless, IV-EWC can improve training stability, which is evidenced by higher mid-curriculum accuracy, stronger forward transfer to the hardest level and reduced forgetting on earlier levels.

Figure 11 reports cn_k12 results with Llama-3.2-3B-Instruct for the vanilla curriculum baseline and IV-EWC. IV-EWC yields higher end-of-curriculum accuracy. The gains are driven by better retention of earlier tasks: at the end of training, Level 0 accuracy is 0.4602 vs 0.4517 for vanilla, and Level 1 improves to 0.3866 vs 0.3616, indicating reduced forgetting and positive backward transfer.

E Parameter Sensitivity Analysis

Regularization Weight Sensitivity. The initial choice of regularization weight γ is crucial for the training dynamics. In Eq. 16, which quantifies the forgetting divergence between training all at once and EWC, $\lambda > 1$ causes the model parameters to remain close to their initial configuration. This produces a larger deviation from the global optimum and results in unchanged or increased forgetting. A sensitivity analysis of the regularization weight was conducted on this basis with the Qwen2.5-3B-Instruct model on the MATH dataset and the results are shown in Figure 12. The results show that forgetting increases at both insufficient and excessive regularization weights. Overall accuracy exhibits a negative association with forgetting. Therefore, choosing a regularization weight outside the suitable range does not achieve the intended effect and may be deleterious. Accordingly, we fixed γ at 1.2e8 to best attenuate forgetting.

LiSSA Recursion Depth Sensitivity. Calculating influences of samples using the influence function (Koh & Liang, 2017) requires the computation of Hessian-vector products (HVPs), where the



(a) Accuracy (b) Forget (c) Trade-Off View

Figure 14: Performance of IV-EWC with different validation set increment using Qwen2.5-3B-Instruct on the MATH dataset.

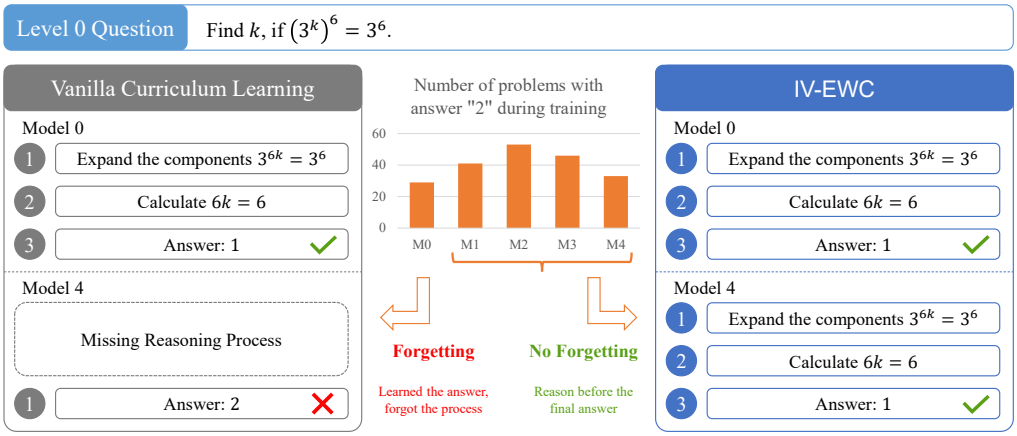


Figure 15: Case study on a MATH test example comparing Llama-3.2-3B-Instruct models trained via vanilla curriculum learning and via IV-EWC. Following training on Levels 1 to 4, the vanilla curriculum model forgets the reasoning process and output the common answer directly, while the model trained with IV-EWC avoids such forgetting.

Linear time Stochastic Second-Order Algorithm (LiSSA) (Agarwal et al., 2017) is used. The LiSSA algorithm requires a large number of iterations to converge. But a precise inverse Hessian-vector product is not necessary for the selection of a high-quality validation set. To determine the suitable recursion depth for IV-EWC, we conducted the LiSSA recursion depth sensitivity analysis with Qwen2.5-3B-Instruct on the MATH dataset, whose result is shown in Figure 13. Both forgetting and overall accuracy exhibit minimal variation even as recursion depth increases exponentially, whereas runtime scales approximately linearly with recursion depth. These findings indicate that IV-EWC is largely insensitive to the choice of recursion depth beyond 50 in LiSSA. In our experiments, we fixed the LiSSA recursion depth at 50 to save training time without impairing the effectiveness of IV-EWC.

Validation Set Increment Sensitivity. The validation set in IV-EWC is updated after each curriculum task by incorporating the top-k most influential training samples. Its size entails a bias-variance trade-off: small sets yield high-variance and potentially biased estimates of the regularizer, while excessively large sets may incorporate noisy or uninformative samples that distort the regularization direction. To identify an appropriate setting, we conducted a sensitivity analysis on the validation set's increment using Qwen2.5-3B-Instruct on the MATH dataset. Figure 14 presents the results, where accuracy declines and forgetting increases when the validation set increment is either too small or too large, revealing a U-shaped relation to this hyperparameter. Notably, even with a modest increment of 25 samples, IV-EWC outperforms vanilla curriculum learning, whose accuracy is 0.7191 and forgetting is 0.0238. Based on the results, the validation set increment was set to 100 for IV-EWC.

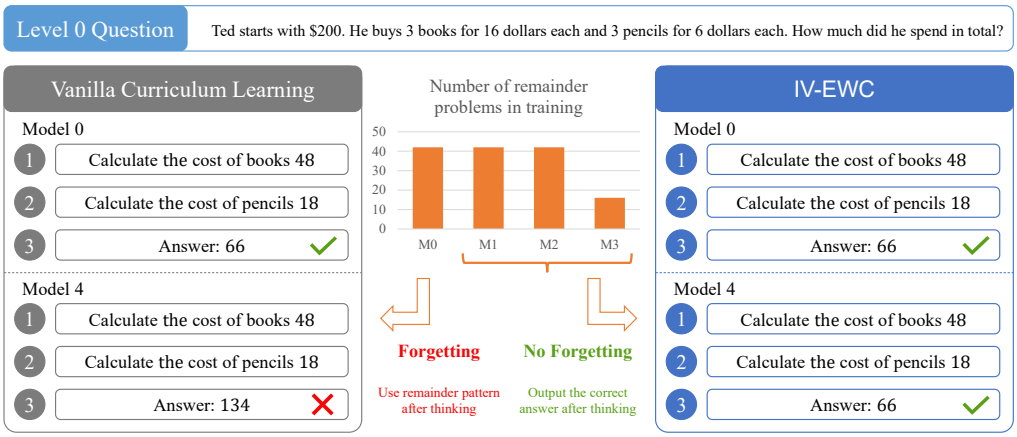


Figure 16: Case study on a GSM8K test example comparing Qwen2.5-3B-Instruct models trained via vanilla curriculum learning and via IV-EWC. Following training on Levels 1 to 3, the vanilla curriculum model outputs the remainder after correct reasoning, while the model trained with IV-EWC outputs the correct answer.

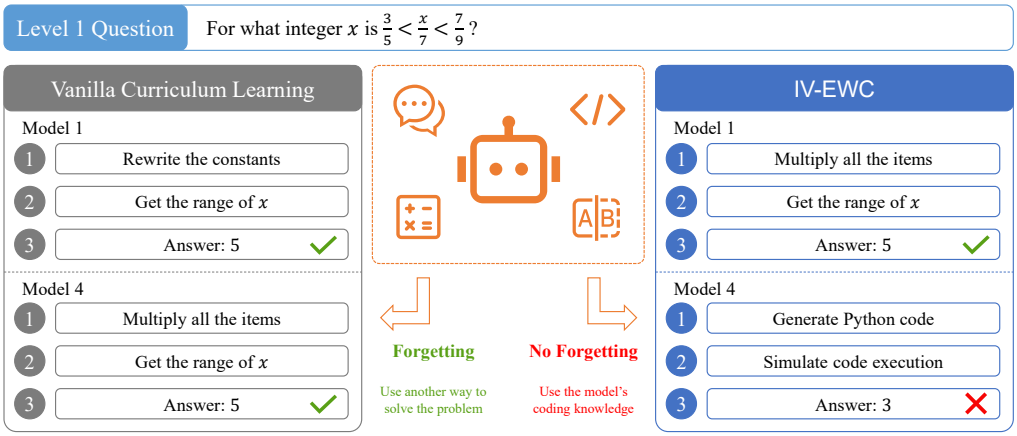


Figure 17: Case study on a MATH test example comparing Llama-3.2-3B-Instruct models trained via vanilla curriculum learning and via IV-EWC. The vanilla curriculum model appears to forget the initial strategy but nonetheless arrives at the correct answer by using another strategy. IV-EWC inadvertently encourages memorization of the model's intrinsic coding knowledge, resulting in a simulated code execution and a wrong answer.

F Additional Results of Case Study

Figure 15 presents the outputs for an example from the Llama-3.2-3B-Instruct base model fine-tuned on MATH via vanilla curriculum learning and IV-EWC. The question is straightforward and both models answer it correctly after Level 0 training. During the subsequent training stages, the training distribution contains more questions with answer 2. When the curriculum concludes, the model trained using vanilla curriculum learning reverts to a shortcut solution, producing the incorrect answer 2 rather than executing the reasoning steps. The model trained with IV-EWC does not exhibit this degradation, demonstrating improved resistance to forgetting during curriculum learning.

Figure 16 illustrates the responses for an example question given by the Qwen2.5-3B-Instruct base model trained on GSM8K via vanilla curriculum learning and IV-EWC. The question is easy with only addition and multiplication knowledge required. Following the training of the Level 0 task,

both models arrive at the correct answer via a valid reasoning process. Conversely, upon completion of the curriculum, the vanilla curriculum model outputs an incorrect answer, which is the remainder of Ted's money, despite producing a correct thinking process. Statistics of the training data show that remainder problems are consistently present during training and the model trained with vanilla curriculum learning appears influenced by this distribution. IV-EWC introduces regularization and makes the model robust to the distribution of the training data, avoiding such errors.

Figure 17 displays the outputs for a representative MATH question generated by Llama-3.2-3B-Instruct after fine-tuning on the MATH dataset with vanilla curriculum learning and IV-EWC. After the training of Level 1, both models answer the question correctly. Following Level 4 training, the vanilla curriculum model adopts a different strategy to solve the problem, indicative of partial forgetting of the original strategy. Meanwhile, the model regularized via IV-EWC leverages its pretrained code-generation capability to produce and virtually execute code, but this pipeline yields an incorrect answer. This case illustrates a trade-off introduced by IV-EWC: it promotes the consolidation of problem-solving strategies, yet it can also induce the model to rely on pretraining priors that are misaligned with the task.