A TASK-CENTRIC THEORY FOR ITERATIVE SELF-IMPROVEMENT WITH EASY-TO-HARD CURRICULA

Chenruo Liu, Yijun Dong, Yiqiu Shen, Qi Lei
New York University
{cl7758,yd1319,ys1001,ql518}@nyu.edu

ABSTRACT

Iterative self-improvement fine-tunes an autoregressive large language model (LLM) on reward-verified outputs generated by the LLM itself. In contrast to the empirical success of self-improvement, the theoretical foundation of this generative, iterative procedure in a practical, finite-sample setting remains limited. We make progress toward this goal by modeling each round of self-improvement as maximum-likelihood fine-tuning on a reward-filtered distribution and deriving finite-sample guarantees for the expected reward. Our analysis reveals a feedback loop where better models accept more data per iteration, supporting sustained self-improvement while explaining eventual saturation of such improvement. Adopting a task-centric view by considering reasoning tasks with multiple difficulty levels, we further prove quantifiable conditions on model initialization, task difficulty, and sample budget where easy-to-hard curricula provably achieve better guarantees than training on fixed mixtures of tasks. Our analyses are validated via Monte-Carlo simulations and controlled experiments on graph-based reasoning tasks.

1 INTRODUCTION

Conditioned on strong pre-training, modern large language models (LLMs) increasingly acquire their unprecedented reasoning skills during post-training not only from human-annotated supervision but also via *iterative self-improvement*—a supervision-free loop where the model iteratively generates candidate answers for questions from the downstream task and then gets fine-tuned on curated question-answer pairs that pass certain external verifications (Xin et al., 2024a;b; Zelikman et al., 2022; Lin et al., 2025a;b; Ren et al., 2025; Zhang et al., 2025; Guan et al., 2025). A closely related practice is to schedule these self-improvement iterations using an *easy-to-hard curriculum*, where gradually shifting the downstream question distribution toward more challenging instances often improves the final performance (Ren et al., 2025; Koh et al., 2025; Lee et al., 2025).

From a theoretical perspective, the empirical success of self-improvement is arguably surprising because the training data is not exogenous, i.e., the candidate solutions are generated by the model itself. Recent theories on self-improvement have made progress on clarifying why learning such endogenous data succeeds without violating the data processing inequality (Shannon, 1948), mainly from the model evolution perspective. For example, Huang et al. (2024) casts self-improvement as a form of probability-mass "sharpening"; while Sun et al. (2025) takes a solver–verifier-gap view of the learning dynamics at the continuous limit. However, as a post-training strategy, the success of self-improvement is highly dependent on the interaction between the pre-trained model and the downstream task. In addition, the discrete, multi-step iterations are critical for the appealing empirical gain of self-improvement. What remains theoretically under-specified for such self-improvement pipelines is a finite-sample, task-centric account that answers two practical questions:

With a single task, when does (multi-step) self-improvement happen?
With a mixture of tasks, when does task scheduling, like easy-to-hard curricula, provably help?

Our contributions. We provide theory-grounded answers to the above questions that closely match self-improvement in practice, which can be summarized as follows:

- **A task-centric framework unveiling when does (multi-step) self-improvement happen on a single task.** Toward the first question, we model each self-improvement iteration as maximum-likelihood fine-tuning on a reward-filtered distribution induced by the model's own generation and an external verifier (Section 3). On a single task, our expected reward lower bounds for single/multi-step self-improvement highlight a feedback loop in which the filtered distributions of better models have more data per iteration being accepted, effectively increasing the training set size. This lens further clarifies the effects of model initialization, task difficulty, and finite sample sizes on the sustained self-improvement and its eventual saturation (Section 4).
- **Across multiple tasks, moderate separation in task difficulties is essential for effective easy-to-hard curricula.** Toward the second question, we compare easy-to-hard scheduling across self-improvement iterations with training on a fixed-mixture baseline under the same budget. We derive quantifiable conditions under which easy-to-hard curricula enjoy strictly better expected reward lower bounds. The analysis unveils three levers for the effectiveness of easy-to-hard scheduling in self-improvement: moderate separation of task difficulties, a critical sample budget, and model initialization (Section 5).
- **Empirical validation on graph-based reasoning tasks.** We validate the analysis for both questions via controlled experiments on graph-based reasoning tasks (Section 6) and complement these experiments with Monte-Carlo simulations that directly visualize the evolution of the expected reward lower bound across self-improvement iterations (Figures 3 and 4).

2 Related Work

Theoretical Understanding of LLM self-improvement. In the context of LLMs, self-improvement broadly refers to a family of procedures in which an LLM produces its own supervision signal and then leverages this signal to improve its capabilities. While LLM self-improvement has shown strong empirical success (Zelikman et al., 2022; Xin et al., 2024a; Guo et al., 2025; Lin et al., 2025b; Zhang et al., 2025; Guan et al., 2025), a growing line of work seeks to understand its underlying mechanisms from a theoretical perspective. Huang et al. (2024) formalizes self-improvement as a consequence of a sharpening mechanism, which encourages the model to place larger probability mass on higher-quality sequences. From a different perspective, Mohri et al. (2025) studies self-improvement through the lens of coherence. Sun et al. (2025) models the training dynamics of self-improvement via the solver-verifier gap. In contrast to these works, our theoretical modeling of self-improvement is more closely aligned with mathematical reasoning settings, i.e., binary and verifiable rewards together with reject sampling for data collection. More importantly, our analysis covers multi-step iterative self-improvement and further incorporates its interaction with easy-to-hard curricula. Due to space constraints, we defer further discussion of empirical self-improvement pipelines, and the connections between our work and self-distillation, self-consuming loops, and model collapse to Appendix D.

3 Problem Setup and Notation

Problem setup. We introduce a theoretical formulation of a practical single-iteration self-improvement procedure for mathematical reasoning (Zelikman et al., 2022; Xin et al., 2024b; Guo et al., 2025; Lin et al., 2025b). At iteration t (with t starting from 0), we sample questions $q \sim p_0$ and generate answers $a \sim \pi_{\theta_t}(\cdot \mid q)$ using the current model parameters θ_t. Unless otherwise stated, we generate a single candidate answer per question. For each pair (q, a) we define a reward (score) function $s(q, a) \in [0, 1]$, where a larger reward indicates a better answer. We retain only samples whose reward is at least a threshold $\tau \in (0, 1]$ and discard the rest.

At the population level, this filtering induces the distribution

$$D'_{p_0,\theta_t}(q, a) = \frac{p_0(q)\, \pi_{\theta_t}(a \mid q)\, \mathbf{1}_{\{\, s(q,a) \geq \tau \,\}}}{Z_{p_0}(\theta_t)},$$

where $\alpha(\theta, q)^1 := \Pr_{a \sim \pi_\theta(\cdot|q)}\big[s(q, a) \geq \tau\big]$ and $Z_p(\theta) := \mathbb{E}_{q \sim p}\big[\alpha(\theta, q)\big]$ denote the per-question and global acceptance rates, respectively. The idealized model update is then $\theta_{t+1} =$

[1] Since an LLM can, in principle, assign nonzero probability to any reasonable text continuation, including a correct solution, we treat $\alpha(\theta, q) > 0$ throughout.

$\arg\max_\theta \ \mathbb{E}_{(q,a)\sim D'_{p_0,\theta_t}}\big[\log \pi_\theta(a \mid q)\big]$. In practice, with finite samples, given the current model $\hat{\theta}_t$ and a dataset of n sampled questions, we obtain a random number n_t (with $n_t \leq n$) of accepted samples $\{(q_i, a_i)\}_{i=1}^{n_t} \sim D'_{p_0,\hat{\theta}_t}$, and perform empirical maximum likelihood estimation:

$$\hat{\theta}_{t+1} = \arg\max_\theta \ \frac{1}{n_t}\sum_{i=1}^{n_t} \log \pi_\theta(a_i \mid q_i).$$

Our goal is to relate this empirical self-improvement objective to the evaluation metric, namely the expected reward $V_{p_0}(\hat{\theta}_{t+1}) := \mathbb{E}_{(q,a)\sim D_{p_0,\hat{\theta}_{t+1}}}\big[s(q,a)\big]$, where $D_{p,\theta}(q,a) := p(q)\pi_\theta(a \mid q)$.

Notation. Due to space constraints, Appendix Table 1 summarizes the notation used throughout the paper, together with brief descriptions and definitions.

4 Iterative Self-Improvement

This section develops basic theoretical tools for iterative self-improvement. We start with a single-step analysis under a general reward function $s(q,a)$ and threshold τ. We then specialize to mathematical reasoning and study the resulting multi-step self-improvement, which will serve as a foundation for our analysis of the easy-to-hard curriculum in Section 5.

4.1 Single-Step Self-Improvement

We begin by analyzing a single round of self-improvement in the finite-sample regime.

Theorem 4.1 (Informal result (formally in Theorem B.1))**.** *Fix an iteration t with current model $\hat{\theta}_t$. Let Π be a finite model class that maps questions to distributions over answers. Suppose that for each $q \sim p_0$ we draw m ($m \geq 1$) i.i.d. candidate answers, keep the first candidate whose reward is at least τ, and discard q if no such candidate exists. Let $n_t^{(m)}$ be the resulting number of accepted training pairs. Then, with probability at least $1-\delta$,*

$$V_{p_0}(\hat{\theta}_{t+1}) \ \geq \ \tau\left(1 - \frac{Z_{p_0}^{(m)}(\hat{\theta}_t)}{\alpha^{(m)}(\hat{\theta}_t)}\sqrt{\frac{2\log\big(|\Pi|\,\delta^{-1}\big)}{n_t^{(m)}}}\right),$$

where $\alpha^{(m)}(\hat{\theta}_t, q) := 1 - \big(1-\alpha(\hat{\theta}_t,q)\big)^m$, $Z_{p_0}^{(m)}(\hat{\theta}_t) := \mathbb{E}_{q\sim p_0}\big[\alpha^{(m)}(\hat{\theta}_t,q)\big]$, and $\alpha^{(m)}(\hat{\theta}_t) := \operatorname{ess\,inf}_q \ \alpha^{(m)}(\hat{\theta}_t,q)$. Moreover, the ratio $Z_{p_0}^{(m)}(\hat{\theta}_t)/\alpha^{(m)}(\hat{\theta}_t)$ is non-increasing in m and satisfies $\lim_{m\to\infty} Z_{p_0}^{(m)}(\hat{\theta}_t)/\alpha^{(m)}(\hat{\theta}_t) = 1$.

Theorem 4.1 extends our setup in Section 3; it reduces to the setting in Section 3 by taking $m=1$, in which case $Z_{p_0}^{(m)}(\hat{\theta}_t) = Z_{p_0}(\hat{\theta}_t)$ and $n_t^{(m)} = n_t$. Theorem 4.1 highlight the importance of finite-sample effects for characterizing self-improvement. Concretely, with infinite samples, the idealized update yields $\hat{\theta}_{t+1}$ satisfying

$$\pi_{\hat{\theta}_{t+1}}(a \mid q) = \frac{\pi_{\hat{\theta}_t}(a \mid q)\,\mathbf{1}_{\{\, s(q,a)\geq\tau\,\}}}{\alpha(\hat{\theta}_t, q)},$$

and hence $V_{p_0}(\hat{\theta}_{t+1}) \geq \tau$. In other words, an infinite-sample (population) update would suggest that a single iteration already guarantees performance above τ and that this guarantee is independent of $\hat{\theta}_t$, both of which are inconsistent with practice. This motivates our finite-sample regime analysis, which further shows that (i) the ratio $Z_{p_0}^{(m)}(\hat{\theta}_t)/\alpha^{(m)}(\hat{\theta}_t)$ decreases with m, and (ii) the effective sample size $n_t^{(m)}$ increases (in expectation) with both n and m. Consequently, to obtain a stronger guarantee on self-improvement (i.e., a larger lower bound on $V_{p_0}(\hat{\theta}_{t+1})$), it is beneficial to increase both the question budget n and the per-question answer budget m.

Remark 4.2 (*On the model class* Π). Our assumption of a finite model class Π is consistent with prior theoretical treatments of self-improvement (Huang et al., 2024). More importantly, the effectiveness of recent work in formalizing self-improvement as tree search over a finite archive of candidate agents (Wang et al., 2025) suggests that the candidate set $|\Pi|$ is typically not very large in practice. Moreover, evidence that stronger language models admit a smaller intrinsic dimension during fine-tuning (Aghajanyan et al., 2021) indicates that $|\Pi|$ tends to be smaller for stronger base models.

4.2 MULTI-STEP SELF-IMPROVEMENT

We specialize to mathematical reasoning by adopting a binary reward $s(q, a) \in \{0, 1\}$, where $s(q, a) = 1$ indicates a correct (verifiable) solution and $s(q, a) = 0$ otherwise. In this regime, combined with Assumption 4.3 which posits a positive relationship between the per-question acceptance rate and the expected reward for all but a γ fraction of questions, Corollary 4.4 relates $V_{p_0}(\hat{\theta}_{t+1})$ directly to the pretrained initialization performance $V_{p_0}(\hat{\theta}_0)$ through an iterated map.

Assumption 4.3. Let Θ be a small neighborhood of the pretrained initialization in which post-training is performed. Then there exist a constant $c \in (0, 1)$ and a small constant $\gamma \geq 0$ such that for any question distribution p and model $\theta \in \Theta$, $\Pr_{q \sim p}\big[\alpha(\theta, q) < c\, V_p(\theta)\big] \leq \gamma$.

Corollary 4.4. *Consider the binary reward setting $s(q, a) \in \{0, 1\}$, where each iteration t uses the same question budget n with $n_t \leq n$ accepted samples. Under Assumption 4.3, define*

$$F(x) \;:=\; 1 - \gamma - \frac{c_\delta \nu}{c\sqrt{x - c_{\delta'}\nu}}$$

with $x > c_{\delta'}\nu$, where $\nu := \sqrt{1/n}$, $c_\delta := \sqrt{2\log(|\Pi|\,\delta^{-1})}$, and $c_{\delta'} := \sqrt{\log(\delta'^{-1})/2}$. Then, with probability at least $1 - \delta - \delta'$, $V_{p_0}(\hat{\theta}_{t+1}) \geq F\big(V_{p_0}(\hat{\theta}_t)\big)$. Moreover, with probability at least $1 - t(\delta + \delta')$, $V_{p_0}(\hat{\theta}_t) \geq F^{\circ t}\big(V_{p_0}(\hat{\theta}_0)\big)$, where $F^{\circ t}$ denotes the t-fold composition of F.

Proposition 4.5. *Under the setting of Corollary 4.4, let ν be sufficiently small such that $0 < \frac{c_\delta \nu}{c\,(1-\gamma-c_{\delta'}\nu)^{3/2}} < \frac{2}{3\sqrt{3}}$. Let $\mathcal{I}(1, \nu) = (x_-(1, \nu), x_+(1, \nu)) \subset (c_{\delta'}\nu,\, 1-\gamma)$ be the interval defined in Definition A.1 with $a = 1$. Then, for any non-negative integer t, $F^{\circ(t+1)}\big(V_{p_0}(\hat{\theta}_0)\big) > F^{\circ t}\big(V_{p_0}(\hat{\theta}_0)\big)$ and $F^{\circ t}\big(V_{p_0}(\hat{\theta}_0)\big) \in \mathcal{I}(1, \nu)$ hold if and only if $V_{p_0}(\hat{\theta}_0) \in \mathcal{I}(1, \nu)$. Moreover, $x_-(1, \nu)$ is increasing in ν, $x_+(1, \nu)$ is decreasing in ν, and the interval length $|\mathcal{I}(1, \nu)| = x_+(1, \nu) - x_-(1, \nu)$ is decreasing in ν and satisfies*

$$|\mathcal{I}(1, \nu)| \;\geq\; (1 - \gamma - c_{\delta'}\nu) - \frac{3\sqrt{3}}{2} \cdot \frac{c_\delta \nu}{c\sqrt{1 - \gamma - c_{\delta'}\nu}}.$$

Remark 4.6.1 (*Moderate task difficulty benefits iterative self-improvement*). Corollary 4.4 and Proposition 4.5 suggest that iterative self-improvement admits monotonic lower-bound guarantees only when the task difficulty is neither too hard nor too easy for the pretrained initialization such that $V_{p_0}(\hat{\theta}_0) \in \mathcal{I}(1, \nu)$. Within $\mathcal{I}(1, \nu)$, better models admit more data per iteration being accepted, thereby sustaining self-improvement over successive iterations.

Remark 4.6.2 (*Benefits of larger budgets*). Increasing the question budget n (i.e., decreasing ν) enlarges the interval $\mathcal{I}(1, \nu)$, and hence enlarges the set of initial performances for which the bound sequence $\{F^{\circ t}(V_{p_0}(\hat{\theta}_0))\}_{t \geq 0}$ is guaranteed to be strictly increasing.

Remark 4.6.3 (*Inherent upper bound*). Iterative self-improvement is inherently bounded: for $V_{p_0}(\hat{\theta}_0) \in \mathcal{I}(1, \nu)$, the lower bound cannot exceed $x_+(1, \nu)$, which is strictly below $1 - \gamma$. This provides a rationale for practical mathematical reasoning pipelines to incorporate additional optimization phases (e.g., reinforcement learning (Guo et al., 2025)) to push performance further.

5 ITERATIVE EASY-TO-HARD CURRICULUM FOR SELF-IMPROVEMENT

Combining self-improvement with an easy-to-hard curriculum across iterations has emerged as a promising approach for further improving model performance by progressively increasing the difficulty of questions encountered in different rounds (Ren et al., 2025; Koh et al., 2025; Lee et al., 2025). Despite its empirical appeal, a principled theoretical understanding of such curriculum-guided self-improvement remains limited. In this section, we extend the tools developed in Section 4 to study when and why integrating iterative self-improvement with an easy-to-hard curriculum can yield stronger self-improvement guarantees.

5.1 EASY-TO-HARD CURRICULUM AND BASELINE

Difficulty levels. We assume there exist $L \geq 2$ task distributions $p_1, \ldots, p_L$, where each p_i is a valid question distribution, and the difficulty increases progressively from p_1 to p_L. Assumption 5.1

formalizes this notion via a power-law separation between adjacent tasks (in difficulty) p_i and p_{i+1}. Concretely, under Assumption 5.1, for every $i \in [L-1]$ and every $\theta \in \Theta$, we have

$$1 \;<\; \frac{i^{-\beta'}}{(i+1)^{-\beta'}} \;\leq\; \frac{V_{p_i}(\theta)}{V_{p_{i+1}}(\theta)} \;\leq\; \frac{i^{-\beta}}{(i+1)^{-\beta}}.$$

This reflects the view that expected reward is a natural measure of task difficulty, and that the relative ordering of difficulty levels should be model-invariant for $\theta \in \Theta$. Moreover, a larger β' corresponds to a larger difficulty ratio between adjacent tasks, while a larger uncertainty width $\Delta := \beta - \beta'$ indicates greater ambiguity in this difficulty ratio.

Assumption 5.1. Let Θ be a small neighborhood of the pretrained initialization in which post-training is performed. Consider L question distributions $\{p_1, p_2, \dots, p_L\}$. For $\theta \in \Theta$ and $i \in [L]$, define the expected reward $V_{p_i}(\theta) := \mathbb{E}_{(q,a)\sim D_{p_i,\theta}}\big[s(q,a)\big]$. Define

$$\beta' := \min_{i\in[L-1]} \inf_{\theta\in\Theta} \frac{\log\big(V_{p_i}(\theta)/V_{p_{i+1}}(\theta)\big)}{\log(1+1/i)} \quad \text{and} \quad \beta := \max_{i\in[L-1]} \sup_{\theta\in\Theta} \frac{\log\big(V_{p_i}(\theta)/V_{p_{i+1}}(\theta)\big)}{\log(1+1/i)}.$$

We assume that $0 < \beta' < \beta$.

Easy-to-hard. For the easy-to-hard curriculum, we consider L iterations of self-improvement. At each iteration $t \in \{0, 1, \dots, L-1\}$, we sample n questions from p_{t+1} to reflect progressively increasing difficulty, and perform one round of self-improvement using the current model. We initialize the curriculum with $\hat{\theta}_0^{\mathrm{E2H}} = \hat{\theta}_0$, and denote the model after iteration t by $\hat{\theta}_{t+1}^{\mathrm{E2H}}$.

Baseline. The baseline we compare against trains for L iterations using a fixed and uniform mixture over all difficulty levels. At each iteration $t \in \{0, 1, \dots, L-1\}$, we always sample n training questions from $p_0 := \frac{1}{L}\sum_{i=1}^{L} p_i$. We use the same initialization $\hat{\theta}_0^{\mathrm{B}} = \hat{\theta}_0$, and denote the model after iteration t by $\hat{\theta}_{t+1}^{\mathrm{B}}$.

5.2 Main Results

We now present our core comparison between the final self-improvement performance under the baseline, $V_{p_0}(\hat{\theta}_L^{\mathrm{B}})$, and under the easy-to-hard curriculum, $V_{p_0}(\hat{\theta}_L^{\mathrm{E2H}})$. Theorem 5.2 provides feasibility conditions that characterize when the lower-bound sequences for both training schemes are monotone across iterations, and an improvement condition under which the easy-to-hard curriculum yields a strictly tighter lower bound than the baseline. These sufficient conditions are highly predictive in practice: they closely track the trends observed in our Monte-Carlo simulations (Figures 3 and 4), and are consistent with the empirical gains on mathematical reasoning tasks reported in Section 6.

Theorem 5.2. *Follow the notation of Corollary 4.4. Fix all parameters except β', β, ν and $V_{p_0}(\hat{\theta}_0)$.*

(i) *Suppose feasibility conditions $\mathcal{M}_i\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0$ holds[2] for all $i \in [4]$. Then, w.h.p., the following statements hold. For the baseline, $V_{p_0}\big(\hat{\theta}_L^{\mathrm{B}}\big) \geq F^{\circ L}\big(V_{p_0}(\hat{\theta}_0)\big)$, and the sequence $\{F^{\circ t}(V_{p_0}(\hat{\theta}_0))\}_{t\geq 0}$ is monotonically increasing in t. For the easy-to-hard curriculum, $V_{p_0}\big(\hat{\theta}_L^{\mathrm{E2H}}\big) \geq (G \circ H_{L-1} \circ H_{L-2} \circ \cdots \circ H_0)\big(V_{p_0}(\hat{\theta}_0)\big)$, where for each $t \in \{0, 1, \dots, L-1\}$,*

$$H_t(x) \;:=\; 1 - \gamma - \frac{c_\delta \nu}{c\sqrt{a_t x - c_{\delta'}\nu}}, \qquad G(x) := a_L x,$$

and $a_0 = L/\sum_{i=1}^{L} i^{-\beta'}$, $a_L = \sum_{i=1}^{L} i^{-\beta'}/L^{1-\beta'}$, $a_t = (t+1)^{-\beta}/t^{-\beta}$ for $t \in [L-1]$. Also, the sequence $\{(H_t \circ \cdots \circ H_0)\big(V_{p_0}(\hat{\theta}_0)\big)\}_{t\geq 0}$ is monotonically increasing in t.

(ii) *If the improvement condition $\mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0$ further holds[2], then the easy-to-hard lower bound is strictly larger than the baseline lower bound: $(G \circ H_{L-1} \circ H_{L-2} \circ \cdots \circ H_0)\big(V_{p_0}(\hat{\theta}_0)\big) > F^{\circ L}\big(V_{p_0}(\hat{\theta}_0)\big)$.*

Interpreting $\{\mathcal{M}_i < 0\}$. To enable meaningful comparisons across iterations, feasibility conditions $\{\mathcal{M}_i < 0\}_{i=1}^{4}$ in Theorem 5.2 rule out degenerate regimes in which the evolution of the expected

[2] Due to space constraints, we defer the explicit forms of $\{\mathcal{M}_i\}_{i=1}^{4}$ and $\mathcal{N}$ to Definition A.2.

reward lower bound becomes ill-defined or fail to be monotonically increasing. We provide a concrete interpretation of the region $\{\mathcal{M}_i < 0\}_{i=1}^4$ in Remark 5.4.

Corollary 5.3. *Follow the setting of Theorem 5.2. Let* $\mathcal{I}(2^{-\beta}, \nu) = (x_-(2^{-\beta}, \nu), x_+(2^{-\beta}, \nu))$ *be the interval defined in Definition A.1 with* $a = 2^{-\beta}$. *Then the feasibility conditions* $\mathcal{M}_i\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0$*,* $i \in [4]$ *are equivalent to* $V_{p_0}(\hat{\theta}_0) \in \mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu)$*, where* $\mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu) := \big(x_-(2^{-\beta}, \nu),\ \frac{2^{-\beta}}{a_0} x_+(2^{-\beta}, \nu)\big)$. *Moreover, the interval length* $|\mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu)|$ *satisfies*

$$2^\beta c_{\delta'}\, \nu \ \le\ \big|\mathcal{I}_{\mathcal{M}}(\beta', \beta, 0)\big| - \big|\mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu)\big| \ \le\ 2^\beta c_{\delta'}\, \nu + \frac{3\sqrt{3}}{2} \cdot \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'} \nu}}.$$

Remark 5.4 (*Feasibility disfavors small budgets and large adjacent difficulty ratios*). Corollaries 5.3 and B.4 together imply that $|\mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu)|$ decreases in ν, β', and β, and that its shrinkage rate in ν is $\Theta(\nu)$ as $\nu \to 0$. Although $\{\mathcal{M}_i < 0\}_{i=1}^4$ only enforces feasibility (rather than directly characterizing when the easy-to-hard curriculum improves over the baseline), we generally prefer $|\mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu)|$ not to be too small. Consequently, (i) an overly small question budget n and (ii) overly large difficulty ratios between adjacent tasks are both undesirable from the standpoint of feasibility. Finally, Figure 3 shows that the condition $V_{p_0}(\hat{\theta}_0) \in \mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu)$ closely matches the behavior observed in direct Monte-Carlo simulations, making $\{\mathcal{M}_i < 0\}_{i=1}^4$ a useful proxy.

Interpreting $\mathcal{N} < 0$. $\mathcal{N} < 0$ serves as the key criterion for improvement: it guarantees that the easy-to-hard curriculum attains a strictly larger final lower bound than the baseline. A concrete interpretation of $\mathcal{N} < 0$ is provided in Remarks 5.7.1–5.7.4. Notably, all the resulting predictions based on the improvement condition $\mathcal{N} < 0$ closely match the trends observed in direct Monte-Carlo simulations in Figure 4.

Proposition 5.5. *Follow the setting of Theorem 5.2. Then the improvement condition* $\mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0$ *is equivalent to* $V_{p_0}(\hat{\theta}_0) \in \mathcal{I}_{\mathcal{N}}(\beta', \beta, \nu)$*, where* $\mathcal{I}_{\mathcal{N}}(\beta', \beta, \nu) := \big(x(\beta', \beta, \nu),\ 1-\gamma\big)$. *For fixed* (β', β)*, we write* $x(\nu) := x(\beta', \beta, \nu)$ *for brevity. Then,* $x(\nu)$ *is monotonically increasing in* ν *and satisfies* $x(0) = 0$*,*

$$x'(\nu) = \frac{c_{\delta'}}{a_0} + \frac{2}{a_0}\Big(\frac{c_\delta}{c(1-\gamma)}\Big)^2 \nu + O\big(\nu^{5/3}\big) \quad \textit{as } \nu \to 0.$$

Let $\mathcal{N}_\infty(\nu) := \lim_{V_{p_0}(\hat{\theta}_0) \to \infty} \mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big)$. *Then there exists a unique* $\nu_c > 0$ *such that* $\mathcal{N}_\infty(\nu_c) = 0$*, and there exists a constant* $C(\nu_c) > 0$ *such that*

$$x'(\nu) \ = \ \frac{C(\nu_c)}{(\nu_c - \nu)^3}\,\big(1 + O(\nu_c - \nu)\big) \quad \textit{as } \nu \uparrow \nu_c.$$

Corollary 5.6. *In Proposition 5.5, fix any initialization* $V_{p_0}(\hat{\theta}_0) \in (0, 1-\gamma)$*, and let* $\nu^\star(\beta', \beta)$ *be defined by the threshold equation* $x\big(\beta', \beta, \nu^\star(\beta', \beta)\big) = V_{p_0}(\hat{\theta}_0)$. *Then* $\nu^\star(\beta', \beta) \ = \ \sup\big\{\, \nu > 0 : \mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0 \,\big\}$. *Moreover,* (i) *fixing* β'*, for* $\beta > \beta'$*,* $\nu^\star(\beta', \beta)$ *is decreasing in* β*;* (ii) *fixing* β*, for* $\beta' \in (0, \beta)$*,* $\nu^\star(\beta', \beta)$ *is increasing in* β'*;* (iii) *fixing* $\Delta = \beta - \beta'$ *and writing* $\nu^\star(\beta', \beta)$ *as* $\nu^\star(\beta', \beta' + \Delta)$*, when* β' *is small,* $\nu^\star(\beta', \beta' + \Delta)$ *is increasing in* β' *and*

$$\nu^\star(\beta', \beta' + \Delta) = \frac{c(1-\gamma)^{3/2} \log\Big(\frac{L}{(L!)^{1/L}}\Big)}{2c_\delta\big(2^{\Delta/2} - 1\big)}\, \beta' \ + \ o(\beta')$$

as $\beta' \to 0$. *Moreover, let* $\nu_T > 0$ *be the (unique) solution to* $\frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} \cdot \frac{1-\Big(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\Big)^{L-1}}{1-\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}} = \frac{1-\gamma}{2}$. *Then there exists a constant* $\nu_0 > 0$ *such that whenever* $\nu_T < \nu_0$*, we have* $\nu^\star(\beta', \beta' + \Delta) < \nu_T$*, and* $\nu^\star(\beta', \beta' + \Delta)$ *is first increasing and then decreasing in* β'*, with a unique maximizer. Moreover, for sufficiently large* β'*, the tail scaling satisfies* $\nu^\star(\beta', \beta' + \Delta) = \Theta\big(2^{-\beta'/2}\big)$.

Remark 5.7.1 (*Phase transition with respect to the question budget*). Proposition 5.5 shows that the interval length $|\mathcal{I}_{\mathcal{N}}(\beta', \beta, \nu)|$ decreases as ν increases. Moreover, the shrinkage rate is mild when ν is small (since $c_{\delta'}$ is typically small and $a_0 > 1$), but becomes steep as ν approaches the critical value

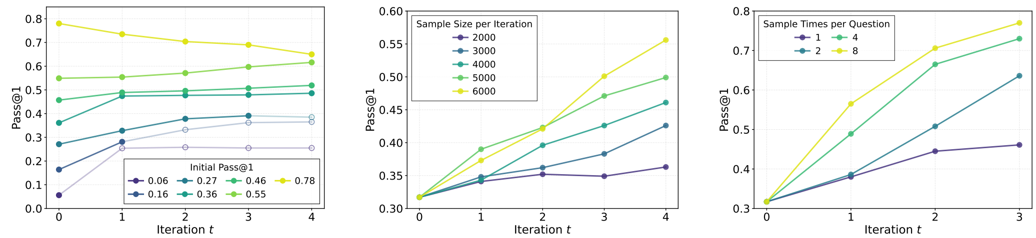


Figure 1: **Iterative self-improvement.** Left panel shows the self-improvement trajectories of a fixed $\hat{\theta}_0$ across tasks with different initial Pass@1 accuracies; hollow markers and faded line segments indicate model collapse (Pass@1$=0$ for at least one target distance l). Middle panel shows the performance under different question budgets n, with $\hat{\theta}_0$ and the initial Pass@1 fixed. Right panel shows the performance under different per-question answer budgets m, with $\hat{\theta}_0$ and the initial Pass@1 fixed.

ν_c ($x'(\nu)$ blows up on the order of $\Theta((\nu_c - \nu)^{-3})$. Equivalently, decreasing the question budget n makes it harder for the easy-to-hard curriculum to provably outperform the baseline, and there is a critical sample size such that as n decreases toward this threshold, the range of initializations $V_{p_0}(\hat{\theta}_0)$ for which easy-to-hard is provably advantageous collapses sharply.

Remark 5.7.2 (*Smaller uncertainty in the adjacent difficulty ratio is better*). Parts (i)-(ii) of Corollary 5.6 imply that, whether we fix β' or β, as the uncertainty width $\Delta = \beta - \beta'$ increases, the maximal admissible ν (and hence the minimal question budget n) for $\mathcal{N} < 0$ becomes more stringent.

Remark 5.7.3 (*Moderate difficulty ratios between adjacent tasks are most favorable*). Part (iii) of Corollary 5.6 further shows that, when Δ is fixed, as the difficulty ratios between adjacent tasks (captured by β') increase, the minimal admissible sample budget n required for the easy-to-hard curriculum to be provably better than the baseline first decreases and then increases. Moreover, in the large-sample regime, there exists a unique optimal β' that minimizes the required sample budget.

Remark 5.7.4 (*Improvement dominates for small budgets, while feasibility dominates for large budgets*). By Corollary 5.3 and Proposition 5.5, for any fixed (β', β), both $|\mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu)|$ and $|\mathcal{I}_{\mathcal{N}}(\beta', \beta, \nu)|$ decrease as ν increases. Moreover, since $c_{\delta'}/a_0 < 2^{\beta} c_{\delta'}$, we have the following dichotomy: as $\nu \to 0$, the shrinkage of the admissible range of $V_{p_0}(\hat{\theta}_0)$ is dominated by the feasibility conditions $\{\mathcal{M}_i < 0\}_{i=1}^{4}$; whereas as ν approaches ν_c, the shrinkage is dominated by the improvement condition $\mathcal{N} < 0$.

6 Experiment

6.1 Task and Training Setup

Shortest path. Given the capability of LLMs to solve graph problems in natural language (Wang et al., 2023), we study self-improvement on a shortest path task using synthetically generated graphs. We consider a directed unweighted graph $\mathcal{G}$. Our task is: given $\mathcal{G}$ and two distinct vertices $v_s \neq v_t$ in $\mathcal{G}$, predict the shortest path length l, i.e., the minimum number of edges among all directed paths from v_s to v_t; if no such path exists, we set $l = -1$.

Model and training. We use Llama-3.2-1B-Instruct as our base LLM (Grattafiori et al., 2024). For each choice of the number of nodes N, expected out-degree $\bar{d}$, and target distance l, we generate a large collection of distinct graphs $\mathcal{G}$ together with vertex pairs (v_s, v_t) whose shortest path length equals l. Each instance is rendered into natural language using a unified prompt template, forming a single shared sample pool. Across experiments, to obtain initialization models with diverse expected rewards $V_{p_0}(\hat{\theta}_0)$, we warm up the base LLM by finetuning it on different subsets of the sample pool with varied finetuning hyperparameters, and draw task sets of varying difficulty from a disjoint portion of the pool. We then follow the procedure described in Section 3 and the easy-to-hard/baseline setup in Section 5.1 to run iterative self-improvement. Note that in the binary reward setting, the expected reward is equivalent to the population Pass@1. Therefore, we report the Pass@1 accuracy on a held-out test set sampled from p_0 as our evaluation metric. More details of dataset construction, warm-up, and self-improvement finetuning are provided in Appendix F.

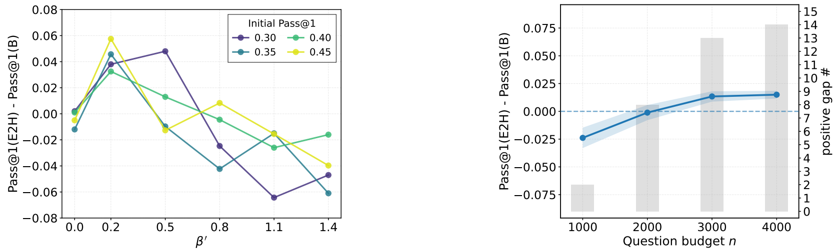


Figure 2: **Iterative self-improvement with an easy-to-hard curriculum.** Left panel fixes $\Delta = 0.04$ and $\hat{\theta}_0$, and shows for different initial Pass@1 accuracies, the final Pass@1 gap between easy-to-hard and the baseline (i.e., $V_{p_0}(\hat{\theta}_L^{\mathrm{E2H}}) - V_{p_0}(\hat{\theta}_L^{\mathrm{B}})$) as a function of the adjacent task difficulty ratio (captured by β'). Right panel fixes $\Delta = 0.04$ and $\beta' = 0.25$, and shows for different initial Pass@1 accuracies (spanning 35%–55%), how the final Pass@1 gap varies with the question budget n. The solid line reports the mean gap across 15 initializations and the shaded region indicates ± 1 standard error; the gray bars (right axis) show the number of initializations with a positive gap at each n.

6.2 Experimental Results

Iterative self-improvement. Figure 1(a) shows the iterative self-improvement performance under tasks of varying difficulty, where we construct task sets such that the initial Pass@1 accuracy (corresponding to $V_{p_0}(\hat{\theta}_0)$) ranges from 6% to 78%. When the task is overly easy, we observe a decreasing trend in the test Pass@1 across iterations, whereas when the task is overly hard, the Pass@1 becomes unstable and may collapse. These trends are consistent with the analysis in Remark 4.6.1, which predicts that effective iterative self-improvement is only guaranteed to occur in a moderate difficulty regime. We also note that the improvement in Pass@1 often slows down after $t = 2$ and the curves begin to plateau at values clearly below 1, which is consistent with Remark 4.6.3.

Moreover, Figure 1(b)-(c) demonstrate that increasing either the question budget n or the per-question answer budget m consistently improves self-improvement performance. This aligns with the finite-sample interpretation in Section 4.1. It is worth mentioning that, despite minor differences in self-improvement setups, a range of empirical studies echo our findings on a broader set of real-world benchmarks that self-improvement tends to favor a moderate task difficulty regime (Singh et al., 2023), benefits from larger budgets (n (Singh et al., 2023; Wilf et al., 2025) and m (Zeng et al., 2024; Bansal et al., 2024; Yao et al., 2025)), and often exhibits a clear saturation limit (Song et al., 2024).

Iterative self-improvement with easy-to-hard curriculum. In Figure 2(a), each curve varies the difficulty ratios between adjacent tasks (controlled by β') while keeping the initial Pass@1 (corresponding to $V_{p_0}(\hat{\theta}_0)$) fixed. Across different values of the initial Pass@1, the final Pass@1 gap between easy-to-hard and the baseline exhibits an overall trend of first increasing and then decreasing as β' grows, with the largest gaps typically attained around $\beta' \in [0.2, 0.5]$. This aligns with Remark 5.7.3, which suggests that moderate difficulty ratios between adjacent tasks are desirable.

Figure 2(b) reports results under different initial Pass@1 accuracies, with a fixed relative difficulty across tasks (i.e., fixed β' and Δ). First, we observe that larger question budgets n lead to a larger final Pass@1 gap of easy-to-hard over the baseline on average. Second, for a diverse set of initial Pass@1 accuracies in the range of 35%–55%, the question budget n at which easy-to-hard starts to outperform the baseline mostly clusters in a relatively narrow range, roughly between 2000 and 3000. This agrees with the phase transition behavior predicted in Remark 5.7.1.

7 Conclusion

In this work, we developed a task-centric framework for understanding LLM self-improvement. For a single task, our finite-sample analysis characterizes key factors (e.g., task difficulty and sampling budget) that determine when (multi-step) self-improvement happens, and explain eventual saturation of such improvement. Beyond single-task training, we further provide theoretical guidance to identify regimes where easy-to-hard scheduling yields a stronger lower bound guarantee than fixed-mixture training, and highlight the role of appropriate adjacent task difficulty ratios and a critical sample size. Our predictions are further supported by synthetic shortest path experiments.

ACKNOWLEDGMENTS

QL acknowledges support of NSF DMS-2523382 and DOE Office of Science under Award #DE-SC0024721. YS was supported in part by the National Institutes of Health (grant no. 1R01EB036530-01A1).

REFERENCES

Armen Aghajanyan, Sonal Gupta, and Luke Zettlemoyer. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. In *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. 7319–7328, 2021.

Hritik Bansal, Arian Hosseini, Rishabh Agarwal, Vinh Q Tran, and Mehran Kazemi. Smaller, weaker, yet better: Training llm reasoners via compute-optimal sampling. *arXiv preprint arXiv:2408.16737*, 2024.

Rudrajit Das and Sujay Sanghavi. Understanding self-distillation in the presence of label noise. In *International Conference on Machine Learning*, pp. 7102–7140. PMLR, 2023.

Yunzhen Feng, Elvis Dohmatob, Pu Yang, Francois Charton, and Julia Kempe. Beyond model collapse: Scaling up with synthesized data requires verification. *arXiv preprint arXiv:2406.07515*, 2024.

Damien Ferbach, Quentin Bertrand, Avishek Joey Bose, and Gauthier Gidel. Self-consuming generative models with curated data provably optimize human preferences. *arXiv preprint arXiv:2407.09499*, 2024.

Shi Fu, Sen Zhang, Yingjie Wang, Xinmei Tian, and Dacheng Tao. Towards theoretical understandings of self-consuming generative models. *arXiv preprint arXiv:2402.11778*, 2024.

Shi Fu, Yingjie Wang, Yuzhu Chen, Li Shen, and Dacheng Tao. Self-verification provably prevents model collapse in recursive synthetic training. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a.

Shi Fu, Yingjie Wang, Yuzhu Chen, Xinmei Tian, and Dacheng Tao. A theoretical perspective: How to prevent model collapse in self-consuming training loops. *arXiv preprint arXiv:2502.18865*, 2025b.

Sara A Geer. *Empirical Processes in M-estimation*, volume 6. Cambridge university press, 2000.

Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, et al. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. *arXiv preprint arXiv:2404.01413*, 2024.

Nate Gillman, Michael Freeman, Daksh Aggarwal, Chia-Hong Hsu, Calvin Luo, Yonglong Tian, and Chen Sun. Self-correcting self-consuming loops for generative model training. *arXiv preprint arXiv:2402.07087*, 2024.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Xinyu Guan, Li Lyna Zhang, Yifei Liu, Ning Shang, Youran Sun, Yi Zhu, Fan Yang, and Mao Yang. rstar-math: Small llms can master math reasoning with self-evolved deep thinking. *arXiv preprint arXiv:2501.04519*, 2025.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Audrey Huang, Adam Block, Dylan J Foster, Dhruv Rohatgi, Cyril Zhang, Max Simchowitz, Jordan T Ash, and Akshay Krishnamurthy. Self-improvement in language models: The sharpening mechanism. *arXiv preprint arXiv:2412.01951*, 2024.

Woosung Koh, Wonbeen Oh, Jaein Jang, MinHyung Lee, Hyeongjin Kim, Ah Yeon Kim, Joonkee Kim, Junghyun Lee, Taehyeon Kim, and Se-Young Yun. Adastar: Adaptive data sampling for training self-taught reasoners. *arXiv preprint arXiv:2505.16322*, 2025.

Nayoung Lee, Ziyang Cai, Avi Schwarzschild, Kangwook Lee, and Dimitris Papailiopoulos. Self-improving transformers overcome easy-to-hard and length generalization challenges. *arXiv preprint arXiv:2502.01612*, 2025.

Yong Lin, Shange Tang, Bohan Lyu, Jiayun Wu, Hongzhou Lin, Kaiyu Yang, Jia Li, Mengzhou Xia, Danqi Chen, Sanjeev Arora, et al. Goedel-prover: A frontier model for open-source automated theorem proving. *arXiv preprint arXiv:2502.07640*, 2025a.

Yong Lin, Shange Tang, Bohan Lyu, Ziran Yang, Jui-Hui Chung, Haoyu Zhao, Lai Jiang, Yihan Geng, Jiawei Ge, Jingruo Sun, et al. Goedel-prover-v2: Scaling formal theorem proving with scaffolded data synthesis and self-correction. *arXiv preprint arXiv:2508.03613*, 2025b.

Hossein Mobahi, Mehrdad Farajtabar, and Peter Bartlett. Self-distillation amplifies regularization in hilbert space. *Advances in Neural Information Processing Systems*, 33:3351–3361, 2020.

Mehryar Mohri, Jon Schneider, and Yifan Wu. Coherence mechanisms for provable self-improvement. *arXiv preprint arXiv:2511.08440*, 2025.

Divyansh Pareek, Simon S Du, and Sewoong Oh. Understanding the gains from repeated self-distillation. *Advances in Neural Information Processing Systems*, 37:7759–7796, 2024.

ZZ Ren, Zhihong Shao, Junxiao Song, Huajian Xin, Haocheng Wang, Wanjia Zhao, Liyue Zhang, Zhe Fu, Qihao Zhu, Dejian Yang, et al. Deepseek-prover-v2: Advancing formal mathematical reasoning via reinforcement learning for subgoal decomposition. *arXiv preprint arXiv:2504.21801*, 2025.

Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27 (3):379–423, 1948.

Avi Singh, John D Co-Reyes, Rishabh Agarwal, Ankesh Anand, Piyush Patil, Xavier Garcia, Peter J Liu, James Harrison, Jaehoon Lee, Kelvin Xu, et al. Beyond human data: Scaling self-training for problem-solving with language models. *arXiv preprint arXiv:2312.06585*, 2023.

Yuda Song, Hanlin Zhang, Carson Eisenach, Sham Kakade, Dean Foster, and Udaya Ghai. Mind the gap: Examining the self-improvement capabilities of large language models. *arXiv preprint arXiv:2412.02674*, 2024.

Yifan Sun, Yushan Liang, Zhen Zhang, and Jiaye Teng. Theoretical modeling of llm self-improvement training dynamics through solver-verifier gap. *arXiv preprint arXiv:2507.00075*, 2025.

Heng Wang, Shangbin Feng, Tianxing He, Zhaoxuan Tan, Xiaochuang Han, and Yulia Tsvetkov. Can language models solve graph problems in natural language? *Advances in Neural Information Processing Systems*, 36:30840–30861, 2023.

Wenyi Wang, Piotr Piękos, Li Nanbo, Firas Laakom, Yimeng Chen, Mateusz Ostaszewski, Mingchen Zhuge, and Jürgen Schmidhuber. Huxley-gödel machine: Human-level coding agent development by an approximation of the optimal self-improving machine, 2025.

Alex Wilf, Pranjal Aggarwal, Bryan Parno, Daniel Fried, Louis-Philippe Morency, Paul Pu Liang, and Sean Welleck. Propose, solve, verify: Self-play through formal verification. *arXiv preprint arXiv:2512.18160*, 2025.

Wing Hung Wong and Xiaotong Shen. Probability inequalities for likelihood ratios and convergence rates of sieve mles. *The Annals of Statistics*, pp. 339–362, 1995.

Huajian Xin, Daya Guo, Zhihong Shao, Zhizhou Ren, Qihao Zhu, Bo Liu, Chong Ruan, Wenda Li, and Xiaodan Liang. Deepseek-prover: Advancing theorem proving in llms through large-scale synthetic data. *arXiv preprint arXiv:2405.14333*, 2024a.

Huajian Xin, ZZ Ren, Junxiao Song, Zhihong Shao, Wanjia Zhao, Haocheng Wang, Bo Liu, Liyue Zhang, Xuan Lu, Qiushi Du, et al. Deepseek-prover-v1. 5: Harnessing proof assistant feedback for reinforcement learning and monte-carlo tree search. *arXiv preprint arXiv:2408.08152*, 2024b.

Jiarui Yao, Yifan Hao, Hanning Zhang, Hanze Dong, Wei Xiong, Nan Jiang, and Tong Zhang. Optimizing chain-of-thought reasoners via gradient variance minimization in rejection sampling and rl. *arXiv preprint arXiv:2505.02391*, 2025.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. Star: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems*, 35:15476–15488, 2022.

Weihao Zeng, Yuzhen Huang, Lulu Zhao, Yijun Wang, Zifei Shan, and Junxian He. B-star: Monitoring and balancing exploration and exploitation in self-taught reasoners. *arXiv preprint arXiv:2412.17256*, 2024.

Jingyuan Zhang, Qi Wang, Xingguang Ji, Yahui Liu, Yang Yue, Fuzheng Zhang, Di Zhang, Guorui Zhou, and Kun Gai. Leanabell-prover: Posttraining scaling in formal reasoning. *arXiv preprint arXiv:2504.06122*, 2025.

Tong Zhang. From ε-entropy to kl-entropy: Analysis of minimum information complexity density estimation. *The Annals of Statistics*, pp. 2180–2210, 2006.

APPENDIX: PROOFS

A. NOTATION AND DEFERRED DEFINITIONS

A.1 NOTATION SUMMARY

Symbol	Meaning
q	question
a	answer
$s(q,a)$	reward; $s(q,a) \in [0,1]$, where larger values indicate a better answer a to the question q
τ	acceptance threshold; $\tau \in (0,1]$ for filtering $s(q,a) \geq \tau$
$\theta, \hat{\theta}$	model parameters; $\hat{\theta}$ denotes the empirical model; includes variants with superscripts or subscripts[3]
$p(\cdot)$	question distribution; includes variants with subscripts
$\pi_\theta(\cdot \mid q)$	answer distribution; conditional distribution induced by θ given q
$D_{p,\theta}(q,a)$	$D_{p,\theta}(q,a) = p(q)\pi_\theta(a \mid q)$
$V_p(\theta)$	expected reward; $V_p(\theta) = \mathbb{E}_{q\sim p(\cdot)}\, \mathbb{E}_{a\sim\pi_\theta(\cdot\mid q)}\big[\, s(q,a)\,\big]$
$\alpha(\theta,q)$	per-question acceptance rate; $\alpha(\theta,q) = \Pr_{a\sim\pi_\theta(\cdot\mid q)}[s(q,a) \geq \tau]$
$Z_p(\theta)$	global acceptance rate; $Z_p(\theta) = \mathbb{E}_{q\sim p(\cdot)}[\alpha(\theta,q)]$
n	question sampling budget; total number of sampled questions per iteration
m	per-question answer budget; number of sampled answers per question (unless otherwise stated, $m=1$)
ν	$\nu = \sqrt{1/n}$
Π	model class; see Thm. B.1 for details
c, γ	reward-acceptance coupling constants; see Assump. 4.3 for details
δ, δ'	failure probabilities; see Thm. B.1 and Cor. 4.4 for details
$c_\delta, c_{\delta'}$	$c_\delta = \sqrt{2\log(\lvert\Pi\rvert\,\delta^{-1})}$, $c_{\delta'} = \sqrt{\log(\delta'^{-1})/2}$
L	total number of iterations
β', β	difficulty separation exponents; controls the difficulty ratios between adjacent tasks; see Assump. 5.1 for details
Δ	difficulty uncertainty width; $\Delta = \beta - \beta'$

Table 1: Common notation used throughout the paper.

A.2 DEFERRED DEFINITIONS AND EXPLICIT EXPRESSIONS

For readability, several auxiliary quantities are only referenced in the main text. Here we collect their explicit definitions and expressions.

Definition A.1 ($\mathcal{I}(a,\nu)$)**.** Suppose (a,ν) are chosen such that

$$1-\gamma-\frac{c_{\delta'}\nu}{a} > 0 \qquad \text{and} \qquad 0 < \frac{a\,c_\delta \nu}{c\,(a(1-\gamma)-c_{\delta'}\nu)^{3/2}} < \sqrt{\frac{4}{27}}.$$

Then the equation

$$y(1-y)^2 = \left(\frac{a\,c_\delta\nu}{c\,(a(1-\gamma)-c_{\delta'}\nu)^{3/2}}\right)^2$$

admits two solutions in $(0,1)$, denoted by $y_-(a,\nu) < y_+(a,\nu)$. Set

$$x_-(a,\nu) = \frac{c_{\delta'}\nu}{a} + \left(1-\gamma-\frac{c_{\delta'}\nu}{a}\right) y_-(a,\nu), \qquad x_+(a,\nu) = \frac{c_{\delta'}\nu}{a} + \left(1-\gamma-\frac{c_{\delta'}\nu}{a}\right) y_+(a,\nu).$$

We define

$$\mathcal{I}(a,\nu) \;:=\; \big(x_-(a,\nu),\, x_+(a,\nu)\big) \;\subset\; \left(\frac{c_{\delta'}\nu}{a},\, 1-\gamma\right).$$

[3] A subscript t ($0 \leq t \leq L$) denotes the model before iteration t (or after iteration $t-1$). Superscripts B and E2H refer to the baseline and the easy-to-hard curriculum in Section 5.1, respectively.

Definition A.2 ($\{\mathcal{M}_i\}_{i=1}^4$ **and** $\mathcal{N}$)**.** Let $x_-(2^{-\beta},\nu)$ and $x_+(2^{-\beta},\nu)$ be the endpoints of the interval $\mathcal{I}(2^{-\beta},\nu)$ defined in Definition A.1 with $a = 2^{-\beta}$. Let, $a_0 = L/\sum_{i=1}^L i^{-\beta'}$ and $a_L = \sum_{i=1}^L i^{-\beta'}/L^{1-\beta'}$. Then, we define

$$\mathcal{M}_1\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big) := x_-(2^{-\beta},\nu) - V_{p_0}(\hat\theta_0),$$

$$\mathcal{M}_2\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big) := V_{p_0}(\hat\theta_0) - x_+(2^{-\beta},\nu),$$

$$\mathcal{M}_3\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big) := x_-(2^{-\beta},\nu) - \left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0V_{p_0}(\hat\theta_0)-c_{\delta'}\nu}}\right),$$

$$\mathcal{M}_4\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big) := \left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0V_{p_0}(\hat\theta_0)-c_{\delta'}\nu}}\right) - x_+(2^{-\beta},\nu),$$

Next, define

$$\begin{aligned}\mathcal{N}\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big) := & -\frac{1}{2}(a_L-1)(1-\gamma) - \frac{c_\delta\nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}}\cdot\frac{1-\left(\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1-\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}}\\
&+a_L\left[\frac{c_\delta\nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}}\cdot\frac{1}{1-\frac{c_\delta\nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0V_{p_0}(\hat\theta_0)-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\cdot e^{-\beta/L}}\right.\\
&\left.+\left(\frac{c_\delta\nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0V_{p_0}(\hat\theta_0)-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\right)^{L-1} L^{-\beta}\cdot\frac{c_\delta\nu}{c\sqrt{a_0V_{p_0}(\hat\theta_0)-c_{\delta'}\nu}}\right].\end{aligned}$$

B. Proofs for Section 4

B.1 Proof of Theorem B.1

Theorem B.1. *Fix an iteration t with current model $\hat\theta_t$. Let $\mathcal{Q}$ denote the question space and let $\Delta(\mathcal{A})$ be the set of probability measures on the answer space $\mathcal{A}$. Let $\Pi\subset(\mathcal{Q}\to\Delta(\mathcal{A}))$ be a finite model class, and suppose that the conditional distribution over answers induced by $D'_{p_0,\hat\theta_t}$ belongs to Π. Suppose that for each $q\sim p_0$ we draw m ($m\ge 1$) i.i.d. candidates $a_1,a_2,\dots,a_m\sim\pi_{\hat\theta_t}(\cdot\mid q)$ sequentially and keep the first accepted candidate: we include (q,a_j) in the training data where $j:=\min\{i\in[m]: s(q,a_i)\ge\tau\}$, and discard q if no such j exists. Let $n_t^{(m)}$ be the resulting number of accepted training pairs. Assume that $\operatorname{ess\,inf}_q\ \alpha(\hat\theta_t,q)>0$. Then, with probability at least $1-\delta$,*

$$V_{p_0}(\hat\theta_{t+1}) \ \ge\ \tau\left(1-\frac{Z_{p_0}^{(m)}(\hat\theta_t)}{\alpha^{(m)}(\hat\theta_t)}\sqrt{\frac{2\log\big(|\Pi|\,\delta^{-1}\big)}{n_t^{(m)}}}\right),$$

where $\alpha^{(m)}(\hat\theta_t,q) := 1-\big(1-\alpha(\hat\theta_t,q)\big)^m$, $Z_{p_0}^{(m)}(\hat\theta_t) := \mathbb{E}_{q\sim p_0}\big[\alpha^{(m)}(\hat\theta_t,q)\big]$, and $\alpha^{(m)}(\hat\theta_t) := \operatorname{ess\,inf}_q\ \alpha^{(m)}(\hat\theta_t,q)$. Moreover, the ratio $Z_{p_0}^{(m)}(\hat\theta_t)/\alpha^{(m)}(\hat\theta_t)$ is non-increasing in m and satisfies $\lim_{m\to\infty} Z_{p_0}^{(m)}(\hat\theta_t)/\alpha^{(m)}(\hat\theta_t)=1$.

Proof. At iteration t we obtain $n_t^{(m)}$ accepted samples $\{(q_i,a_i)\}_{i=1}^{n_t^{(m)}}\sim D'^{(m)}_{p_0,\hat\theta_t}$. Under this scheme, the population joint distribution of accepted pairs can be written as

$$D'^{(m)}_{p_0,\hat\theta_t}(q,a) = p'^{(m)}_{p_0,\hat\theta_t}(q)\ p'_{\hat\theta_t}(a\mid q),$$

where the marginal over questions and the conditional over answers are, respectively,

$$p'^{(m)}_{p_0,\hat{\theta}_t}(q) = \frac{p_0(q)\,\alpha^{(m)}(\hat{\theta}_t, q)}{Z^{(m)}_{p_0}(\hat{\theta}_t)}, \qquad p'_{\hat{\theta}_t}(a \mid q) = \frac{\pi_{\hat{\theta}_t}(a \mid q)\,\mathbf{1}_{\{\, s(q,a)\geq\tau \,\}}}{\alpha(\hat{\theta}_t, q)},$$

where

$$\alpha(\hat{\theta}_t, q) := \Pr_{a\sim\pi_{\hat{\theta}_t}(\cdot|q)}\big[s(q,a) \geq \tau \,\big|\, q\big] = \sum_a \pi_{\hat{\theta}_t}(a \mid q)\,\mathbf{1}_{\{\, s(q,a)\geq\tau \,\}},$$

$$\alpha^{(m)}(\hat{\theta}_t, q) := 1 - \big(1 - \alpha(\hat{\theta}_t, q)\big)^m, \qquad Z^{(m)}_{p_0}(\hat{\theta}_t) := \mathbb{E}_{q\sim p_0}\big[\alpha^{(m)}(\hat{\theta}_t, q)\big].$$

For any $m \geq 1$, the population MLE objective $\mathbb{E}_{(q,a)\sim D'^{(m)}_{p_0,\hat{\theta}_t}}\big[\log \pi_\theta(a \mid q)\big]$ at iteration t achieves its maximum at $\theta = \theta^\star_{t+1}$ where

$$\pi_{\theta^\star_{t+1}}(\cdot \mid q) = p'_{\hat{\theta}_t}(\cdot \mid q)$$

for almost every q. Note that $p'_{\hat{\theta}_t}(\cdot \mid q)$ is independent of m and coincides with the conditional distribution over answers induced by $D'_{p_0,\hat{\theta}_t}$; hence, by assumption, $p'_{\hat{\theta}_t}(\cdot \mid q) \in \Pi$.

Define $A_q := \{a : \ s(q,a) \geq \tau\}$. By construction, $\pi_{\theta^\star_{t+1}}(A_q \mid q) = 1$ and hence $\pi_{\theta^\star_{t+1}}(A_q^{\mathrm{c}} \mid q) = 0$. Further define

$$\delta_t(q) := \Pr_{a\sim\pi_{\hat{\theta}_{t+1}}(\cdot|q)}[s(q,a) < \tau] = \pi_{\hat{\theta}_{t+1}}(A_q^{\mathrm{c}} \mid q).$$

By the definition of total variation distance,

$$\delta_t(q) = \big|\pi_{\theta^\star_{t+1}}(A_q^{\mathrm{c}} \mid q) - \pi_{\hat{\theta}_{t+1}}(A_q^{\mathrm{c}} \mid q)\big| \ \leq \ \mathrm{TV}\Big(\pi_{\theta^\star_{t+1}}(\cdot \mid q),\, \pi_{\hat{\theta}_{t+1}}(\cdot \mid q)\Big).$$

For any dominating measure ω, it follows that

$$\begin{aligned}
\mathrm{TV}\Big(\pi_{\theta^\star_{t+1}}(\cdot \mid q),\, \pi_{\hat{\theta}_{t+1}}(\cdot \mid q)\Big) &= \frac{1}{2}\int \left| \frac{\mathrm{d}\,\pi_{\theta^\star_{t+1}}(\cdot \mid q)}{\mathrm{d}\omega} - \frac{\mathrm{d}\,\pi_{\hat{\theta}_{t+1}}(\cdot \mid q)}{\mathrm{d}\omega} \right| \mathrm{d}\omega \\
&\leq \left(\int \left(\sqrt{\frac{\mathrm{d}\,\pi_{\theta^\star_{t+1}}(\cdot \mid q)}{\mathrm{d}\omega}} - \sqrt{\frac{\mathrm{d}\,\pi_{\hat{\theta}_{t+1}}(\cdot \mid q)}{\mathrm{d}\omega}} \right)^2 \mathrm{d}\omega \right)^{1/2} \\
&=: \sqrt{D^2_{\mathrm{H}}\Big(\pi_{\theta^\star_{t+1}}(\cdot \mid q),\, \pi_{\hat{\theta}_{t+1}}(\cdot \mid q)\Big)}.
\end{aligned}$$

Here $D^2_{\mathrm{H}}(\cdot,\cdot)$ denotes the Hellinger distance. Taking expectation over $q \sim p'^{(m)}_{p_0,\hat{\theta}_t}$ and using the bound above, we obtain

$$\mathbb{E}_{q\sim p'^{(m)}_{p_0,\hat{\theta}_t}}\big[\delta_t(q)\big] \leq \left(\mathbb{E}_{q\sim p'^{(m)}_{p_0,\hat{\theta}_t}}\Big[D^2_{\mathrm{H}}\big(\pi_{\theta^\star_{t+1}}(\cdot \mid q),\, \pi_{\hat{\theta}_{t+1}}(\cdot \mid q)\big)\Big] \right)^{1/2}.$$

Based on Lemma B.2, with probability at least $1-\delta$,

$$\mathbb{E}_{q\sim p'^{(m)}_{p_0,\hat{\theta}_t}}\big[\delta_t(q)\big] \ \leq \ \sqrt{\frac{2\log(|\Pi|\,\delta^{-1})}{n_t^{(m)}}}.$$

Therefore, w.h.p.,

$$\bar{\delta}_t^{(p_0)} \ := \ \mathbb{E}_{q\sim p_0}\big[\delta_t(q)\big] = \mathbb{E}_{q\sim p'^{(m)}_{p_0,\hat{\theta}_t}}\left[\frac{Z^{(m)}_{p_0}(\hat{\theta}_t)}{\alpha^{(m)}(\hat{\theta}_t, q)}\,\delta_t(q) \right] \ \leq \ \frac{Z^{(m)}_{p_0}(\hat{\theta}_t)}{\alpha^{(m)}(\hat{\theta}_t)} \sqrt{\frac{2\log\big(|\Pi|\,\delta^{-1}\big)}{n_t^{(m)}}},$$

where $\alpha^{(m)}(\hat{\theta}_t) := \operatorname{ess\,inf}_q \alpha^{(m)}(\hat{\theta}_t, q)$. For any fixed q,

$$\begin{aligned}
\mathbb{E}_{a\sim\pi_{\hat{\theta}_{t+1}}(\cdot|q)}[s(q,a)] &= \mathbb{E}\big[s(q,a)\,\mathbf{1}_{A_q} \,\big|\, q\big] \ + \ \mathbb{E}\big[s(q,a)\,\mathbf{1}_{A_q^{\mathrm{c}}} \,\big|\, q\big] \\
&\geq \tau\,\pi_{\hat{\theta}_{t+1}}(A_q \mid q) \\
&= \tau\big(1 - \delta_t(q)\big).
\end{aligned}$$

Taking expectation over $q \sim p_0$, with probability at least $1-\delta$, we have

$$V_{p_0}(\hat{\theta}_{t+1}) = \mathbb{E}_{q\sim p_0}\mathbb{E}_{a\sim\pi_{\hat{\theta}_{t+1}}(\cdot|q)}[s(q,a)] \;\geq\; \tau\big(1-\bar{\delta}_t^{(p_0)}\big) \;\geq\; \tau\left(1-\frac{Z_{p_0}^{(m)}(\hat{\theta}_t)}{\alpha^{(m)}(\hat{\theta}_t)}\sqrt{\frac{2\log\big(|\Pi|\,\delta^{-1}\big)}{n_t^{(m)}}}\right).$$

Recall the notation $\alpha(\hat{\theta}_t, q) \in [0,1]$ and $\alpha(\hat{\theta}_t) := \operatorname{ess\,inf}_q \alpha(\hat{\theta}_t, q) > 0$. For $m \in \mathbb{N}$ define

$$f_m(x) \;:=\; 1-(1-x)^m, \qquad x \in [0,1].$$

Then

$$\alpha^{(m)}(\hat{\theta}_t, q) = f_m(\alpha(\hat{\theta}_t, q)), \quad \alpha^{(m)}(\hat{\theta}_t) = \operatorname*{ess\,inf}_q f_m\big(\alpha(\hat{\theta}_t, q)\big) = f_m(\alpha(\hat{\theta}_t)), \quad Z_{p_0}^{(m)}(\hat{\theta}_t) = \mathbb{E}_{q\sim p_0}\big[f_m(\alpha(\hat{\theta}_t, q))\big].$$

Define, for $y \in [0,1)$,

$$h_m(y) \;:=\; \frac{1-y^{m+1}}{1-y^m}.$$

Using $f_{m+1}(x)/f_m(x) = h_m(1-x)$, we have

$$\frac{Z_{p_0}^{(m+1)}(\hat{\theta}_t)}{\alpha^{(m+1)}(\hat{\theta}_t)} = \mathbb{E}\left[\frac{f_{m+1}(\alpha(\hat{\theta}_t,q))}{f_m(\alpha(\hat{\theta}_t,q))}\cdot\frac{f_m(\alpha(\hat{\theta}_t,q))}{f_m(\alpha(\hat{\theta}_t))}\cdot\frac{f_m(\alpha(\hat{\theta}_t))}{f_{m+1}(\alpha(\hat{\theta}_t))}\right] = \mathbb{E}_{q\sim p_0}\left[\frac{h_m(1-\alpha(\hat{\theta}_t,q))}{h_m(1-\alpha(\hat{\theta}_t))}\cdot\frac{f_m(\alpha(\hat{\theta}_t,q))}{f_m(\alpha(\hat{\theta}_t))}\right].$$

We claim h_m is increasing on $[0,1)$. Indeed,

$$\frac{\mathrm{d}}{\mathrm{d}y}\log h_m(y) = \frac{y^{m-1}(m-(m+1)y+y^{m+1})}{(1-y^m)(1-y^{m+1})} \;\geq\; 0,$$

since the $m-(m+1)y+y^{m+1}$ is decreasing in y with value m at $y=0$ and 0 at $y=1$. Because $\alpha(\hat{\theta}_t,q) \geq \alpha(\hat{\theta}_t)$ a.s., we have $1-\alpha(\hat{\theta}_t,q) \leq 1-\alpha(\hat{\theta}_t)$ and hence $h_m(1-\alpha(\hat{\theta}_t,q)) \leq h_m(1-\alpha(\hat{\theta}_t))$. Therefore,

$$\frac{Z_{p_0}^{(m+1)}(\hat{\theta}_t)}{\alpha^{(m+1)}(\hat{\theta}_t)} \;\leq\; \frac{Z_{p_0}^{(m)}(\hat{\theta}_t)}{\alpha^{(m)}(\hat{\theta}_t)}.$$

Since $1 > \alpha(\hat{\theta}_t) > 0$, we have $f_m(\alpha(\hat{\theta}_t,q)) \uparrow 1$ for every q and $f_m(\alpha(\hat{\theta}_t)) \uparrow 1$ as $m \to \infty$. By the monotone convergence theorem, $Z_{p_0}^{(m)}(\hat{\theta}_t) = \mathbb{E}[f_m(\alpha(\hat{\theta}_t,q))] \to 1$ and $\alpha^{(m)}(\hat{\theta}_t) = f_m(\alpha(\hat{\theta}_t)) \to 1$, hence

$$\lim_{m\to\infty}\frac{Z_{p_0}^{(m)}(\hat{\theta}_t)}{\alpha^{(m)}(\hat{\theta}_t)} = 1.$$

□

Lemma B.2 (Wong & Shen (1995); Geer (2000); Zhang (2006); Huang et al. (2024))**.** *Fix iteration t and $m \geq 1$. Let $\mathcal{Q}$ be the question space and $\Delta(\mathcal{A})$ the set of probability measures on the answer space $\mathcal{A}$. Let $\Pi \subset (\mathcal{Q} \to \Delta(\mathcal{A}))$ be a finite model class and suppose the population optimizer $\pi_{\theta_{t+1}^\star}(\cdot \mid q) = p'_{\hat{\theta}_t}(\cdot \mid q)$ belongs to Π. Draw $n_t^{(m)}$ accepted samples i.i.d.*

$$(q_i, a_i) \;\sim\; D'^{(m)}_{p_0,\hat{\theta}_t}(q,a) \;=\; p'^{(m)}_{p_0,\hat{\theta}_t}(q)\,\pi_{\theta_{t+1}^\star}(a \mid q), \qquad i = 1,\ldots,n_t^{(m)},$$

and define the empirical MLE

$$\hat{\theta}_{t+1} \;\in\; \arg\max_{\theta:\,\pi_\theta\in\Pi}\sum_{i=1}^{n_t^{(m)}}\log\pi_\theta(a_i \mid q_i).$$

Then for any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\mathbb{E}_{q\sim p'^{(m)}_{p_0,\hat{\theta}_t}}\Big[D_{\mathrm{H}}^2\big(\pi_{\hat{\theta}_{t+1}}(\cdot \mid q),\, \pi_{\theta_{t+1}^\star}(\cdot \mid q)\big)\Big] \;\leq\; \frac{2\log\big(|\Pi|\,\delta^{-1}\big)}{n_t^{(m)}}.$$

B.2 Proof of Corollary 4.4

Proof. Fix iteration t, recall

$$Z_{p_0}(\hat{\theta}_t) = \mathbb{E}_{q\sim p_0}\big[\alpha(\hat{\theta}_t, q)\big], \qquad \alpha(\hat{\theta}_t, q) = \Pr_{a\sim\pi_{\hat{\theta}_t}(\cdot|q)}\big[s(q,a) \geq \tau \,\big|\, q\big].$$

When $s \in \{0,1\}$ and $\tau \in (0,1]$, we have $\mathbf{1}_{\{s(q,a)\geq\tau\}} = s(q,a)$, hence

$$\alpha(\hat{\theta}_t, q) = \mathbb{E}_{a\sim\pi_{\hat{\theta}_t}(\cdot|q)}\big[s(q,a)\big].$$

Averaging over $q \sim p_0$ yields

$$Z_{p_0}(\hat{\theta}_t) = \mathbb{E}_{q\sim p_0}\mathbb{E}_{a\sim\pi_{\hat{\theta}_t}(\cdot|q)}\big[s(q,a)\big] = V_{p_0}(\hat{\theta}_t).$$

Moreover, since $s \in \{0,1\}$, for any fixed q,

$$\mathbb{E}_{a\sim\pi_{\hat{\theta}_{t+1}}(\cdot|q)}[s(q,a)] = \Pr_{a\sim\pi_{\hat{\theta}_{t+1}}(\cdot|q)}[s(q,a)=1] = 1 - \delta_t(q),$$

where $\delta_t(q) := \Pr_{a\sim\pi_{\hat{\theta}_{t+1}}(\cdot|q)}[s(q,a) < \tau]$. Therefore,

$$V_{p_0}(\hat{\theta}_{t+1}) = \mathbb{E}_{q\sim p_0}\mathbb{E}_{a\sim\pi_{\hat{\theta}_{t+1}}(\cdot|q)}[s(q,a)] = 1 - \bar{\delta}_t^{(p_0)}, \qquad \bar{\delta}_t^{(p_0)} := \mathbb{E}_{q\sim p_0}[\delta_t(q)].$$

At iteration t, we propose n i.i.d. pairs $(q_i, a_i) \sim p_0(q)\,\pi_{\hat{\theta}_t}(a \mid q)$, and accept those with $s(q_i, a_i) \geq \tau$. Let $X_i := \mathbf{1}_{\{s(q_i,a_i)\geq\tau\}} \in \{0,1\}$ and $n_t := \sum_{i=1}^n X_i$. Then, by the binary reward identity above,

$$\mathbb{E}[X_i] = \mathbb{E}_{q\sim p_0} \Pr_{a\sim\pi_{\hat{\theta}_t}(\cdot|q)}[s(q,a) \geq \tau \mid q] = Z_{p_0}(\hat{\theta}_t) = V_{p_0}(\hat{\theta}_t).$$

Hoeffding's inequality gives, for any $\epsilon > 0$,

$$\Pr\Big[\frac{n_t}{n} \leq V_{p_0}(\hat{\theta}_t) - \epsilon\Big] \leq \exp(-2n\epsilon^2).$$

Choosing $\epsilon = \sqrt{\log(\delta'^{-1})/(2n)}$ implies that with probability at least $1-\delta'$,

$$\frac{n_t}{n} \;\geq\; V_{p_0}(\hat{\theta}_t) - \sqrt{\frac{\log(\delta'^{-1})}{2n}}.$$

Define

$$\mathcal{G}_t := \Big\{q : \; \alpha(\hat{\theta}_t, q) \;\geq\; c\,V_{p_0}(\hat{\theta}_t)\Big\}.$$

By Assumption 4.3, $\Pr_{q\sim p_0}\big[q \notin \mathcal{G}_t\big] \;\leq\; \gamma$. Since $\delta_t(q) \in [0,1]$, we have

$$\bar{\delta}_t^{(p_0)} = \mathbb{E}_{q\sim p_0}\big[\delta_t(q)\mathbf{1}_{\mathcal{G}_t}\big] + \mathbb{E}_{q\sim p_0}\big[\delta_t(q)\mathbf{1}_{\mathcal{G}_t^c}\big] \;\leq\; \mathbb{E}_{q\sim p_0}\big[\delta_t(q)\mathbf{1}_{\mathcal{G}_t}\big] + \Pr_{q\sim p_0}\big[q \notin \mathcal{G}_t\big] \;\leq\; \mathbb{E}_{q\sim p_0}\big[\delta_t(q)\mathbf{1}_{\mathcal{G}_t}\big] + \gamma.$$

Furthermore,

$$\mathbb{E}_{q\sim p_0}\big[\delta_t(q)\mathbf{1}_{\mathcal{G}_t}\big] = \mathbb{E}_{q\sim p'_{p_0,\hat{\theta}_t}}\left[\frac{Z_{p_0}(\hat{\theta}_t)}{\alpha(\hat{\theta}_t, q)}\,\delta_t(q)\mathbf{1}_{\mathcal{G}_t}\right] \;\leq\; \frac{1}{c}\,\mathbb{E}_{q\sim p'_{p_0,\hat{\theta}_t}}\big[\delta_t(q)\big],$$

where we use the fact that On $\mathcal{G}_t$, we have $\alpha(\hat{\theta}_t, q) \geq cV_{p_0}(\hat{\theta}_t) = cZ_{p_0}(\hat{\theta}_t)$. Thus

$$\bar{\delta}_t^{(p_0)} \;\leq\; \gamma + \frac{1}{c}\,\mathbb{E}_{q\sim p'_{p_0,\hat{\theta}_t}}\big[\delta_t(q)\big].$$

By repeating the identical argument in the proof of Theorem B.1 with $m = 1$, we obtain that with probability at least $1-\delta$,

$$\mathbb{E}_{q\sim p'_{p_0,\hat{\theta}_t}}\big[\delta_t(q)\big] \;\leq\; \sqrt{\frac{2\log\big(|\Pi|\,\delta^{-1}\big)}{n_t}}.$$

Then, we get (w.p. $\geq 1-\delta$):

$$V_{p_0}(\hat{\theta}_{t+1}) = 1 - \bar{\delta}_t^{(p_0)} \;\geq\; 1 - \gamma - \frac{1}{c}\sqrt{\frac{2\log\big(|\Pi|\,\delta^{-1}\big)}{n_t}}.$$

Moreover, w.p. $\geq 1-\delta'$,

$$\frac{1}{\sqrt{n_t}} \;\leq\; \frac{1}{\sqrt{n\Big(V_{p_0}(\hat{\theta}_t) - \sqrt{\log(\delta'^{-1})/(2n)}\Big)}}.$$

Therefore, by a union bound over the two events, with probability at least $1-\delta-\delta'$,

$$V_{p_0}(\hat{\theta}_{t+1}) \;\geq\; 1 - \gamma - \frac{1}{c}\sqrt{\frac{2\log\big(|\Pi|\delta^{-1}\big)/n}{V_{p_0}(\hat{\theta}_t) \;-\; \sqrt{\log(\delta'^{-1})/(2n)}}}\,.$$

Define

$$F(x) \;:=\; 1-\gamma-\frac{c_\delta \nu}{c\sqrt{x - c_{\delta'}\nu}}, \qquad \nu := \sqrt{\frac{1}{n}}, \qquad c_\delta := \sqrt{2\log\big(|\Pi|\,\delta^{-1}\big)}, \qquad c_{\delta'} := \sqrt{\frac{\log(\delta'^{-1})}{2}},$$

on its natural domain $x > c_{\delta'}\nu$.Then the preceding bound can be rewritten as

$$V_{p_0}(\hat{\theta}_{t+1}) \;\geq\; F\big(V_{p_0}(\hat{\theta}_t)\big).$$

Finally, since F is monotone increasing on its domain, iterating the one-step inequality yields that, with probability at least $1-t(\delta+\delta')$, for every integer $t \geq 0$,

$$V_{p_0}(\hat{\theta}_t) \;\geq\; F^{\circ t}\big(V_{p_0}(\hat{\theta}_0)\big),$$

where $F^{\circ t}$ denotes the t-fold composition of F. This completes the proof.

□

B.3 Proof of Proposition 4.5

Proof. This proposition is an immediate specialization of Corollary B.4 by taking $a = 1$. □

Lemma B.3. *Let*

$$F(x) \;=\; 1 - \frac{\sigma}{\sqrt{x}}, \qquad x \in (0,1),$$

with parameter $\sigma > 0$. Assume $0 < \sigma < \sqrt{4/27}$. Let $x_- < x_+$ denote the two solutions in $(0,1)$ of $x(1-x)^2 = \sigma^2$. Then,

$$x_+ - x_- \geq 1 - \frac{3\sqrt{3}}{2}\,\sigma,$$

and for every $x \in (x_-, x_+)$ and every integer $t \geq 0$, all iterates $F^{\circ t}(x)$ stay in (x_-, x_+) and satisfy $F^{\circ(t+1)}(x) \;>\; F^{\circ t}(x)$.

Proof. Consider

$$h(x) \;:=\; x(1-x)^2, \qquad x \in [0,1].$$

Then $h'(x) = (1-x)^2 - 2x(1-x) = (1-x)(1-3x)$, so h is strictly increasing on $(0, 1/3)$ and strictly decreasing on $(1/3, 1)$, with a unique interior maximizer at $x = 1/3$. Moreover, $h(1/3) = 4/27$. So whenever $\sigma^2 < 4/27$, the equation $h(x) = \sigma^2$ has exactly two distinct solutions in $(0,1)$; we denote them by $x_- < x_+$.

Moreover, expanding $h(x) = \sigma^2$ gives the cubic $x^3 - 2x^2 + x - \sigma^2 = 0$. Set $x = z + 2/3$, we have

$$x^3 - 2x^2 + x - \sigma^2 = z^3 - \frac{1}{3}z + \Big(\frac{2}{27} - \sigma^2\Big) = 0.$$

Trigonometric solution of the roots of this depressed cubic is

$$z_\ell = \frac{2}{3}\cos\Big[\frac{1}{3}\arccos\Big(-1+\frac{27}{2}\sigma^2\Big)-\frac{2\pi\ell}{3}\Big], \qquad \ell = 0,1,2.$$

so

$$x_\ell = \frac{2}{3}+\frac{2}{3}\cos\Big[\frac{1}{3}\arccos\Big(-1+\frac{27}{2}\sigma^2\Big)-\frac{2\pi\ell}{3}\Big], \qquad \ell = 0,1,2.$$

Define

$$u \ := \ \frac{1}{3}\arccos\Big(-1+\frac{27}{2}\sigma^2\Big) \in \Big(0, \frac{\pi}{3}\Big),$$

then

$$x_+ = \frac{2}{3}+\frac{2}{3}\cos\Big(u-\frac{2\pi}{3}\Big), \qquad x_- = \frac{2}{3}+\frac{2}{3}\cos\Big(u-\frac{4\pi}{3}\Big).$$

Hence

$$x_+ - x_- = \frac{2}{\sqrt{3}}\sin u = \frac{2}{\sqrt{3}}\ \sin\Big(\frac{1}{3}\arccos\Big(-1+\frac{27}{2}\sigma^2\Big)\Big).$$

We now prove the desired bound $x_+ - x_- \ \geq \ 1 - \frac{3\sqrt{3}}{2}\,\sigma$. Let $\cos\theta := (3\sqrt{3}\sigma)/2 \in (0,1)$ with $\theta \in [0, \pi/2]$, then

$$-1+\frac{27}{2}\sigma^2 = 2\cos^2\theta - 1 = \cos(2\theta),$$

Hence $x_+ - x_- = \frac{2}{\sqrt{3}}\sin\left(\frac{2\theta}{3}\right)$. and $1 - \frac{3\sqrt{3}}{2}\,\sigma = 1 - \cos\theta$.

Thus it suffices to show that for all $\theta \in [0, \pi/2]$,

$$\frac{2}{\sqrt{3}}\sin\Big(\frac{2\theta}{3}\Big) \ \geq \ 1 - \cos\theta.$$

Define

$$g(\theta) := \frac{2}{\sqrt{3}}\sin\Big(\frac{2\theta}{3}\Big) - \big(1-\cos\theta\big), \qquad \theta \in [0, \frac{\pi}{2}].$$

We have $g(0) = g(\pi/2) = 0$. Moreover,

$$g''(\theta) = -\frac{8}{9\sqrt{3}}\sin\Big(\frac{2\theta}{3}\Big) - \cos\theta \leq 0$$

for $\theta \in [0, \pi/2]$ since we have $\sin(2\theta/3) \geq 0$ and $\cos\theta \geq 0$, thus g is concave on $[0, \pi/2]$ and vanishes at both endpoints. So $g(\theta) \geq 0$, which is equivalent to

$$x_+ - x_- \ \geq \ 1 - \frac{3\sqrt{3}}{2}\,\sigma.$$

Next, we characterize where $F(x) > x$. For $x \in (0,1)$, this is equivalent to $\sigma^2 < x(1-x)^2$. Recall that $x_\pm$ are the two solutions to $x(1-x)^2 = \sigma^2$ in $(0,1)$, and that $h(x) = x(1-x)^2$ is strictly increasing on $(0, 1/3)$ and strictly decreasing on $(1/3, 1)$. Hence $x(1-x)^2 > \sigma^2$ is equivalent to $x \in (x_-, x_+)$, so we have $F(x) > x$ is equivalent to $x \in (x_-, x_+)$. Let $x \in (x_-, x_+)$ and define $x_t := F^{\circ t}(x)$ for $t \geq 0$.

We prove by induction that

$$x_- < x_t < x_+, \qquad x_{t+1} > x_t, \qquad \forall\, t \geq 0.$$

For $t = 0$, we have $x_0 = x \in (x_-, x_+)$ by assumption, and $x_1 = F(x_0) > x_0$. Using monotonicity of F and $F(x_\pm) = x_\pm$,

$$x_- = F(x_-) < F(x_0) = x_1 < F(x_+) = x_+,$$

so $x_1 \in (x_-, x_+)$. Assume now that $x_t \in (x_-, x_+)$ and $x_t > x_{t-1}$ for some $t \geq 1$. Then, since $x_t \in (x_-, x_+)$, $x_{t+1} = F(x_t) > x_t$. Furthermore, monotonicity of F and the fixed-point property at $x_\pm$ give

$$x_- = F(x_-) < F(x_t) = x_{t+1} < F(x_+) = x_+,$$

so $x_{t+1} \in (x_-, x_+)$. This completes the induction and shows that for every integer $t \geq 0$, $F^{\circ(t+1)}(x) \ > \ F^{\circ t}(x)$. □

Corollary B.4. *Let*

$$F(x;a,\nu)\ :=\ 1-\gamma-\frac{c_\delta\nu}{c\sqrt{ax-c_{\delta'}\nu}},\qquad x>\frac{c_{\delta'}\nu}{a}.$$

Assume (a,ν) satisfies the validity conditions in Definition A.1, so that the interval

$$\mathcal{I}(a,\nu)\ :=\ \big(x_-(a,\nu),\,x_+(a,\nu)\big)\ \subset\ \Big(\frac{c_{\delta'}\nu}{a},\,1-\gamma\Big)$$

in Definition A.1 is well-defined. Then, for any $x\in\mathcal{I}(a,\nu)$ and any integer $t\geq 0$, $F^{\circ t}(x;a,\nu)\in\mathcal{I}(a,\nu)$ and $F^{\circ(t+1)}(x;a,\nu)>F^{\circ t}(x;a,\nu)$. Moreover, the interval length satisfies

$$|\mathcal{I}(a,\nu)|\ \geq\ \Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big)-\frac{3\sqrt{3}}{2}\cdot\frac{c_\delta\nu}{c\sqrt{a(1-\gamma)-c_{\delta'}\nu}}.$$

Furthermore, the family $\{\mathcal{I}(a,\nu)\}$ is monotone in the sense of inclusion: (i) for fixed ν, if $a_2>a_1>0$, then $\mathcal{I}(a_1,\nu)\subset\mathcal{I}(a_2,\nu)$; (ii) for fixed a, if $\nu_2>\nu_1$, then $\mathcal{I}(a,\nu_2)\subset\mathcal{I}(a,\nu_1)$.

Proof. We follow the notations in Definition A.1. Define the affine change of variables

$$x=\frac{c_{\delta'}\nu}{a}+\Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big)\,y,\qquad y\in(0,1),$$

A direct substitution shows that for every $y\in(0,1)$,

$$\begin{aligned}F\Big(\frac{c_{\delta'}\nu}{a}+\big(1-\gamma-\frac{c_{\delta'}\nu}{a}\big)\,y;a,\nu\Big)&=1-\gamma-\frac{c_\delta\nu}{c\sqrt{a\big(\frac{c_{\delta'}\nu}{a}+(1-\gamma-\frac{c_{\delta'}\nu}{a})y\big)-c_{\delta'}\nu}}\\&=1-\gamma-\frac{c_\delta\nu}{c\sqrt{\big(a(1-\gamma)-c_{\delta'}\nu\big)\,y}}\\&=\frac{c_{\delta'}\nu}{a}+\Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big)\Big(1-\frac{a\,c_\delta\nu}{c\,(a(1-\gamma)-c_{\delta'}\nu)^{3/2}}\cdot\frac{1}{\sqrt{y}}\Big).\end{aligned}$$

Therefore, if we denote

$$g_{a,\nu}(y):=1-\frac{a\,c_\delta\nu}{c\,(a(1-\gamma)-c_{\delta'}\nu)^{3/2}}\cdot\frac{1}{\sqrt{y}},\qquad y\in(0,1),$$

then we have

$$F\Big(\frac{c_{\delta'}\nu}{a}+\big(1-\gamma-\frac{c_{\delta'}\nu}{a}\big)\,y;a,\nu\Big)=\frac{c_{\delta'}\nu}{a}+\Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big)\,g_{a,\nu}(y).$$

The map $g_{a,\nu}$ is exactly of the form in Lemma B.3 with parameter

$$\frac{a\,c_\delta\nu}{c\,(a(1-\gamma)-c_{\delta'}\nu)^{3/2}}\in\Big(0,\sqrt{\frac{4}{27}}\Big).$$

Then, applying Lemma B.3 to $g_{a,\nu}$ yields: for every $y\in\big(y_-(a,\nu),y_+(a,\nu)\big)$ and every $t\geq 0$,

$$g_{a,\nu}^{\circ t}(y)\in\big(y_-(a,\nu),y_+(a,\nu)\big)\quad\text{and}\quad g_{a,\nu}^{\circ(t+1)}(y)>g_{a,\nu}^{\circ t}(y),$$

and moreover

$$y_+(a,\nu)-y_-(a,\nu)\ \geq\ 1-\frac{3\sqrt{3}}{2}\cdot\frac{a\,c_\delta\nu}{c\,(a(1-\gamma)-c_{\delta'}\nu)^{3/2}}.$$

Now take any $x\in\mathcal{I}(a,\nu)$ and write it as $x=\frac{c_{\delta'}\nu}{a}+\big(1-\gamma-\frac{c_{\delta'}\nu}{a}\big)\,y$ with $y\in\big(y_-(a,\nu),y_+(a,\nu)\big)$. Iterating the conjugacy identity gives

$$F^{\circ t}(x;a,\nu)=\frac{c_{\delta'}\nu}{a}+\Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big)\,g_{a,\nu}^{\circ t}(y),\qquad\forall\,t\geq 0,$$

so $g_{a,\nu}^{\circ t}(y)\in\big(y_-(a,\nu),y_+(a,\nu)\big)$ implies $F^{\circ t}(x;a,\nu)\in\mathcal{I}(a,\nu)$, and $g_{a,\nu}^{\circ(t+1)}(y)>g_{a,\nu}^{\circ t}(y)$ implies $F^{\circ(t+1)}(x;a,\nu)>F^{\circ t}(x;a,\nu)$.

For the interval length, using the lemma's bound,

$$
\begin{aligned}
|\mathcal{I}(a,\nu)| = x_+(a,\nu) - x_-(a,\nu) &= \Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big)\big(y_+(a,\nu)-y_-(a,\nu)\big) \\
&\ge \Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big)\left(1-\frac{3\sqrt{3}}{2}\cdot\frac{a\,c_\delta\nu}{c\,(a(1-\gamma)-c_{\delta'}\nu)^{3/2}}\right) \\
&= \Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big) - \frac{3\sqrt{3}}{2}\cdot\frac{c_\delta\nu}{c\sqrt{a(1-\gamma)-c_{\delta'}\nu}}.
\end{aligned}
$$

For any parameters (a,ν) in the validity range of Definition A.1, recall that the endpoints $x_-(a,\nu) < x_+(a,\nu)$ are the two fixed points of $F(\cdot;a,\nu)$, i.e.,

$$F\big(x_\pm(a,\nu);a,\nu\big) = x_\pm(a,\nu).$$

Define the fixed-point equation

$$\Phi(x,a,\nu) := x - F(x;a,\nu) = 0.$$

Whenever $\partial_x\Phi\big(x_\pm(a,\nu),a,\nu\big) \neq 0$, the implicit function theorem gives

$$\frac{\partial}{\partial a}x_\pm(a,\nu) = -\frac{\partial_a\Phi\big(x_\pm(a,\nu),a,\nu\big)}{\partial_x\Phi\big(x_\pm(a,\nu),a,\nu\big)}, \qquad \frac{\partial}{\partial \nu}x_\pm(a,\nu) = -\frac{\partial_\nu\Phi\big(x_\pm(a,\nu),a,\nu\big)}{\partial_x\Phi\big(x_\pm(a,\nu),a,\nu\big)}.$$

Fix ν and view Φ as a function of (a,x). First, for any $x > c_{\delta'}\nu/a$,

$$\partial_a\Phi(x,a,\nu) = -\partial_a F(x;a,\nu) = -\frac{c_\delta\nu}{c}\cdot\frac{x}{2\,(ax-c_{\delta'}\nu)^{3/2}} \;<\; 0.$$

Next, $\partial_x\Phi(x,a,\nu) = 1-\partial_x F(x;a,\nu)$. To determine its sign at the fixed points, consider the affine map

$$x = \frac{c_{\delta'}\nu}{a} + \Big(1-\gamma-\frac{c_{\delta'}\nu}{a}\Big)\,y,$$

which maps $F(x;a,\nu)$ to $g_{a,\nu}(y) = 1-\sigma(a,\nu)/\sqrt{y}$. At a fixed point $y = g_{a,\nu}(y)$ we have $\sigma(a,\nu) = (1-y)\sqrt{y}$, hence

$$g'_{a,\nu}(y) = \frac{\sigma(a,\nu)}{2y^{3/2}} = \frac{1-y}{2y}.$$

Since $y_-(a,\nu) \in (0,1/3)$ and $y_+(a,\nu) \in (1/3,1)$, $g'_{a,\nu}\big(y_-(a,\nu)\big) > 1$ and $g'_{a,\nu}\big(y_+(a,\nu)\big) < 1$. Therefore $\partial_x\Phi\big(x_-(a,\nu),a,\nu\big) < 0$ and $\partial_x\Phi\big(x_+(a,\nu),a,\nu\big) > 0$. Combining with $\partial_a\Phi(x_\pm(a,\nu),a,\nu) < 0$ yields

$$\frac{\partial}{\partial a}x_-(a,\nu) < 0, \qquad \frac{\partial}{\partial a}x_+(a,\nu) > 0.$$

Consequently, for any $a_2 > a_1$ (with the same fixed ν),

$$x_-(a_2,\nu) < x_-(a_1,\nu), \qquad x_+(a_2,\nu) > x_+(a_1,\nu),$$

i.e., $\mathcal{I}(a_1,\nu) \subset \mathcal{I}(a_2,\nu)$.

Finally, fix a and view Φ as a function of (ν,x). For any $x > c_{\delta'}\nu/a$, $\partial_\nu\Phi(x,a,\nu) = -\partial_\nu F(x;a,\nu)$. A direct differentiation shows $\partial_\nu F(x;a,\nu) < 0$, hence $\partial_\nu\Phi(x,a,\nu) > 0$. Moreover, as established above, $\partial_x\Phi\big(x_-(a,\nu),a,\nu\big) < 0$ and $\partial_x\Phi\big(x_+(a,\nu),a,\nu\big) > 0$. Therefore,

$$\frac{\partial}{\partial \nu}x_-(a,\nu) > 0, \qquad \frac{\partial}{\partial \nu}x_+(a,\nu) < 0,$$

Consequently, for any $\nu_2 > \nu_1$ (with the same fixed a).

$$x_-(a,\nu_2) > x_-(a,\nu_1), \qquad x_+(a,\nu_2) < x_+(a,\nu_1),$$

which implies $\mathcal{I}(a,\nu_2) \subset \mathcal{I}(a,\nu_1)$. $\square$

C. PROOFS FOR SECTION 5

C.1 PROOF OF THEOREM 5.2

Proof. We now prove the first part of the theorem. By Definition A.2, the conditions $\mathcal{M}_i\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0$ for all $i \in [4]$ are equivalent to requiring that both $V_{p_0}(\hat{\theta}_0)$ and $1 - \gamma - c_\delta \nu / (c\sqrt{a_0 V_{p_0}(\hat{\theta}_0) - c_{\delta'}\nu})$ lie in the open interval $(x_-(2^{-\beta}, \nu),\, x_+(2^{-\beta}, \nu))$. By Corollary B.4, since $2^{-\beta} < 1$, we have $V_{p_0}(\hat{\theta}_0) \in \mathcal{I}(2^{-\beta}, \nu) \subset \mathcal{I}(1, \nu)$. Therefore, by Corollary 4.4 and Proposition 4.5, it follows that, with high probability,

$$V_{p_0}(\hat{\theta}_L^{\mathrm{B}}) \;\geq\; F^{\circ L}\Big(V_{p_0}(\hat{\theta}_0)\Big),$$

and the sequence $\{F^{\circ t}(V_{p_0}(\hat{\theta}_0))\}_{t \geq 0}$ is monotonically increasing in t, where

$$F(x) \;=\; 1 - \gamma - \frac{c_\delta \nu}{c\sqrt{x - c_{\delta'}\nu}}.$$

Next, recall that $\hat{\theta}_0^{\mathrm{E2H}} = \hat{\theta}_0$ and for easy-to-hard, during iteration t we use distribution p_{t+1}. By Assumption 5.1, for every $i \in [L-1]$ and every θ,

$$\frac{V_{p_i}(\theta)}{V_{p_{i+1}}(\theta)} \;\geq\; \frac{i^{-\beta'}}{(i+1)^{-\beta'}}.$$

Equivalently, $V_{p_{i+1}}(\theta) \leq (i/(i+1))^{\beta'} V_{p_i}(\theta)$. Iterating from 1 to $i-1$ yields, $V_{p_i}(\theta) \leq i^{-\beta'}\, V_{p_1}(\theta)$. Applying this with $\theta = \hat{\theta}_0$ and averaging over i gives

$$V_{p_0}(\hat{\theta}_0) \;=\; \frac{1}{L}\sum_{i=1}^{L} V_{p_i}(\hat{\theta}_0) \;\leq\; \frac{1}{L}\Big(\sum_{i=1}^{L} i^{-\beta'}\Big)\, V_{p_1}(\hat{\theta}_0),$$

hence

$$V_{p_1}(\hat{\theta}_0) \;\geq\; a_0\, V_{p_0}(\hat{\theta}_0), \qquad a_0 \;:=\; \frac{L}{\sum_{i=1}^{L} i^{-\beta'}}.$$

Note that since $i^{-\beta'} < 1$ for all $i \geq 2$, we have $\sum_{i=1}^{L} i^{-\beta'} < L$ (for $L \geq 2$), and therefore $a_0 > 1$.

Then, at iteration t, easy-to-hard trains on p_{t+1}. By Corollary 4.4 (substitute p_0 with p_{t+1}), with high probability,

$$V_{p_{t+1}}(\hat{\theta}_{t+1}^{\mathrm{E2H}}) \;\geq\; F\Big(V_{p_{t+1}}(\hat{\theta}_t^{\mathrm{E2H}})\Big), \qquad t = 0, 1, \ldots, L-1.$$

Again by Assumption 5.1, for every $t \in [L-1]$ and every $\theta \in \Theta$,

$$\frac{V_{p_t}(\theta)}{V_{p_{t+1}}(\theta)} \;\leq\; \frac{t^{-\beta}}{(t+1)^{-\beta}},$$

equivalently,

$$V_{p_{t+1}}(\theta) \;\geq\; a_t\, V_{p_t}(\theta), \qquad a_t \;:=\; \frac{(t+1)^{-\beta}}{t^{-\beta}} \;<\; 1.$$

Define

$$H_t(x) \;:=\; F(a_t x) \;=\; 1 - \gamma - \frac{c_\delta \nu}{c\sqrt{a_t x - c_{\delta'}\nu}}, \qquad t = 0, 1, \ldots, L-1,$$

with $\{a_t\}$ defined above, We now chain the previous steps. First, since $V_{p_1}(\hat{\theta}_0) \geq a_0 V_{p_0}(\hat{\theta}_0)$, we have

$$V_{p_1}(\hat{\theta}_1^{\mathrm{E2H}}) \;\geq\; F\Big(V_{p_1}(\hat{\theta}_0)\Big) \;\geq\; F\Big(a_0 V_{p_0}(\hat{\theta}_0)\Big) \;=\; H_0\Big(V_{p_0}(\hat{\theta}_0)\Big).$$

Next, for $t = 1$,

$$V_{p_2}(\hat{\theta}_2^{\mathrm{E2H}}) \;\geq\; F\Big(V_{p_2}(\hat{\theta}_1^{\mathrm{E2H}})\Big) \;\geq\; F\Big(a_1 V_{p_1}(\hat{\theta}_1^{\mathrm{E2H}})\Big) \;=\; H_1\Big(V_{p_1}(\hat{\theta}_1^{\mathrm{E2H}})\Big).$$

Continuing this argument inductively for $t = 2, \ldots, L-1$ yields the recursion

$$V_{p_{t+1}}(\hat{\theta}_{t+1}^{\mathrm{E2H}}) \;\geq\; H_t\Big(V_{p_t}(\hat{\theta}_t^{\mathrm{E2H}})\Big), \qquad t = 0, 1, \ldots, L-1,$$

and therefore, after L steps,

$$V_{p_L}(\hat{\theta}_L^{\mathrm{E2H}}) \;\geq\; (H_{L-1} \circ H_{L-2} \circ \cdots \circ H_0)\Big(V_{p_0}(\hat{\theta}_0)\Big).$$

We finally lower bound $V_{p_0}(\hat{\theta}_L^{\mathrm{E2H}})$ in terms of $V_{p_L}(\hat{\theta}_L^{\mathrm{E2H}})$. Using Assumption 5.1 and iterating as in for a_0, for every $i \in \{1, \ldots, L\}$ and every $\theta \in \Theta$ we have

$$V_{p_i}(\theta) \;\geq\; \Big(\frac{i^{-\beta'}}{L^{-\beta'}}\Big) V_{p_L}(\theta) \;=\; \Big(\frac{L}{i}\Big)^{\beta'} V_{p_L}(\theta).$$

Averaging over i gives

$$V_{p_0}(\theta) \;=\; \frac{1}{L}\sum_{i=1}^{L} V_{p_i}(\theta) \;\geq\; \frac{1}{L}\sum_{i=1}^{L}\Big(\frac{L}{i}\Big)^{\beta'} V_{p_L}(\theta) \;=\; \frac{\sum_{i=1}^{L} i^{-\beta'}}{L^{1-\beta'}}\, V_{p_L}(\theta).$$

Applying this with $\theta = \hat{\theta}_L^{\mathrm{E2H}}$ yields

$$V_{p_0}(\hat{\theta}_L^{\mathrm{E2H}}) \;\geq\; a_L\, V_{p_L}(\hat{\theta}_L^{\mathrm{E2H}}), \qquad a_L \;:=\; \frac{\sum_{i=1}^{L} i^{-\beta'}}{L^{1-\beta'}},$$

and clearly $a_L > 1$ since $\sum_{i=1}^{L} i^{-\beta'} > L \cdot L^{-\beta'} = L^{1-\beta'}$. Defining $G(x) := a_L x$, we obtain

$$V_{p_0}(\hat{\theta}_L^{\mathrm{E2H}}) \;\geq\; (G \circ H_{L-1} \circ H_{L-2} \circ \cdots \circ H_0)\Big(V_{p_0}(\hat{\theta}_0)\Big),$$

as claimed. Moreover, note that for $t \in [L-1]$, a_t is strictly increasing in t. Therefore, by Corollary B.4, the associated intervals $\mathcal{I}(a_t, \nu)$ expand as t increases. Since $V_{p_0}(\hat{\theta}_0)$ and $H_t(V_{p_0}(\hat{\theta}_0)) = 1 - \gamma - c_\delta \nu / (c\sqrt{a_0 V_{p_0}(\hat{\theta}_0) - c_{\delta'}\nu})$ lies in the smallest admissible interval $(x_-(2^{-\beta}, \nu),\, x_+(2^{-\beta}, \nu))$, the chained lower bounds stay within the corresponding invariant intervals and satisfy$\{(H_t \circ H_{t-1} \circ \cdots \circ H_0)\big(V_{p_0}(\hat{\theta}_0)\big)\}_{t \geq 0}$ is monotonically increasing in t.

We now prove the second part of the theorem. Although the statement of Theorem 5.2 concerns $\nu > 0$, the maps involved are continuous in ν, and it is convenient to first analyze the case $\nu = 0$. To distinguish the dependence on ν, we write the baseline map as $F_\nu(\cdot)$ and the easy-to-hard maps as $H_{t,\nu}(\cdot)$.

When $\nu = 0$, the maps simplify to constants:

$$F_0(x) \;=\; 1 - \gamma, \qquad H_{t,0}(x) \;=\; 1 - \gamma, \qquad t = 0, 1, \ldots, L-1, \qquad G(x) = a_L x.$$

Define the comparison gap

$$\Delta(\nu, x) \;:=\; \big(G \circ H_{L-1,\nu} \circ H_{L-2,\nu} \circ \cdots \circ H_{0,\nu}\big)(x) \;-\; F_\nu^{\circ L}(x).$$

Then we have

$$\Delta(0, x) = G\big(H_{L-1,0} \circ \cdots \circ H_{0,0}(x)\big) - F_0^{\circ L}(x) = a_L(1-\gamma) - (1-\gamma) = (a_L - 1)(1-\gamma).$$

Since $a_L > 1$ and $1 - \gamma > 0$, it follows that $\Delta\big(0, V_{p_0}(\hat{\theta}_0)\big) = (a_L - 1)(1-\gamma) \;>\; 0$. Therefore, our next step is to lower bound $\Delta\big(\nu, V_{p_0}(\hat{\theta}_0)\big) - \Delta\big(0, V_{p_0}(\hat{\theta}_0)\big)$. For convenience, denote $x_0 := V_{p_0}(\hat{\theta}_0)$.[4] Observe that the above difference involves two scalar sequences induced by the iterated maps. For baseline, define $\{x_t^{\mathrm{B}}(\nu)\}_{t=0}^{L}$ by

$$x_0^{\mathrm{B}}(\nu) = x_0, \qquad x_{t+1}^{\mathrm{B}}(\nu) = F_\nu\big(x_t^{\mathrm{B}}(\nu)\big), \qquad t = 0, 1, \ldots, L-1.$$

By construction, $F_\nu^{\circ L}(x_0) \;=\; x_L^{\mathrm{B}}(\nu)$. Similarly, for easy-to-hard define $\{x_t^{\mathrm{E2H}}(\nu)\}_{t=0}^{L}$ by

$$x_0^{\mathrm{E2H}}(\nu) = x_0, \qquad x_{t+1}^{\mathrm{E2H}}(\nu) = H_{t,\nu}\big(x_t^{\mathrm{E2H}}(\nu)\big), \qquad t = 0, 1, \ldots, L-1.$$

[4] We will use this shorthand throughout the remainder of this proof (and in subsequent proofs) whenever the dependence on the initialization is through $V_{p_0}(\hat{\theta}_0)$.

By construction, $\big(G \circ H_{L-1,\nu} \circ H_{L-2,\nu} \circ \cdots \circ H_{0,\nu}\big)(x_0) \;=\; a_L\, x_L^{\mathrm{E2H}}(\nu)$.

Combining the above identities, we obtain

$$\begin{aligned}\Delta(\nu, x_0) - \Delta(0, x_0) &= \Big(a_L x_L^{\mathrm{E2H}}(\nu) - x_L^{\mathrm{B}}(\nu)\Big) - \Big(a_L x_L^{\mathrm{E2H}}(0) - x_L^{\mathrm{B}}(0)\Big) \\ &= \big(x_L^{\mathrm{B}}(0) - x_L^{\mathrm{B}}(\nu)\big) - a_L\big(x_L^{\mathrm{E2H}}(0) - x_L^{\mathrm{E2H}}(\nu)\big).\end{aligned}$$

Therefore, if we define the deviation sequences

$$e_t^{\mathrm{B}}(\nu) \;:=\; x_t^{\mathrm{B}}(0) - x_t^{\mathrm{B}}(\nu), \qquad e_t^{\mathrm{E2H}}(\nu) \;:=\; x_t^{\mathrm{E2H}}(0) - x_t^{\mathrm{E2H}}(\nu),$$

then the quantity of interest can be written succinctly as $\Delta(\nu, x_0) - \Delta(0, x_0) = e_L^{\mathrm{B}}(\nu) - a_L\, e_L^{\mathrm{E2H}}(\nu)$. Hence, it suffices to derive a lower bound on $e_L^{\mathrm{B}}(\nu)$ and an upper bound on $e_L^{\mathrm{E2H}}(\nu)$.

For the lower bound on $e_L^{\mathrm{B}}(\nu)$, ignoring the trivial case $t = 0$, we start from

$$e_1^{\mathrm{B}}(\nu) = F_0(x_0) - F_\nu(x_0) = \frac{c_\delta \nu}{c\sqrt{x_0 - c_{\delta'}\nu}}.$$

More generally, for every $t \geq 1$ we can write

$$\begin{aligned}e_t^{\mathrm{B}}(\nu) &= F_0\big(x_{t-1}^{\mathrm{B}}(0)\big) - F_\nu\big(x_{t-1}^{\mathrm{B}}(\nu)\big) \\ &= \Big(F_0\big(x_{t-1}^{\mathrm{B}}(0)\big) - F_\nu\big(x_{t-1}^{\mathrm{B}}(0)\big)\Big) + \Big(F_\nu\big(x_{t-1}^{\mathrm{B}}(0)\big) - F_\nu\big(x_{t-1}^{\mathrm{B}}(\nu)\big)\Big).\end{aligned}$$

The first term is nonnegative because $F_\nu(x) \leq F_0(x) = 1 - \gamma$ for all admissible x when $\nu > 0$. The second term is also nonnegative because F_ν is increasing and $x_{t-1}^{\mathrm{B}}(0) \geq x_{t-1}^{\mathrm{B}}(\nu)$. Hence,

$$e_t^{\mathrm{B}}(\nu) \geq 0, \qquad \forall t \geq 1.$$

Next we derive a quantitative lower bound. The first difference is explicit:

$$F_0\big(x_{t-1}^{\mathrm{B}}(0)\big) - F_\nu\big(x_{t-1}^{\mathrm{B}}(0)\big) = \frac{c_\delta \nu}{c\sqrt{1 - \gamma - c_{\delta'}\nu}}.$$

For the second difference, note that

$$F_\nu'(x) = \frac{c_\delta \nu}{2c\,(x - c_{\delta'}\nu)^{3/2}},$$

which is monotonically decreasing in x over its domain. By the mean value theorem, there exists $\xi_{t-1} \in \big[x_{t-1}^{\mathrm{B}}(\nu),\, 1-\gamma\big]$ such that $F_\nu\big(x_{t-1}^{\mathrm{B}}(0)\big) - F_\nu\big(x_{t-1}^{\mathrm{B}}(\nu)\big) = F_\nu'(\xi_{t-1})\, e_{t-1}^{\mathrm{B}}(\nu)$. Since F_ν' is decreasing and $\xi_{t-1} \leq 1 - \gamma$, we have $F_\nu'(\xi_{t-1}) \geq F_\nu'(1-\gamma)$, and thus

$$F_\nu\big(x_{t-1}^{\mathrm{B}}(0)\big) - F_\nu\big(x_{t-1}^{\mathrm{B}}(\nu)\big) \;\geq\; \frac{c_\delta \nu}{2c\,(1 - \gamma - c_{\delta'}\nu)^{3/2}}\; e_{t-1}^{\mathrm{B}}(\nu).$$

Combining the two pieces yields the recursion: for every $t = 2, 3, \ldots, L$,

$$e_t^{\mathrm{B}}(\nu) \;\geq\; \frac{c_\delta \nu}{c\sqrt{1 - \gamma - c_{\delta'}\nu}} \;+\; \frac{c_\delta \nu}{2c\,(1 - \gamma - c_{\delta'}\nu)^{3/2}}\; e_{t-1}^{\mathrm{B}}(\nu).$$

Iterating from $t = 2$ up to $t = L$ gives

$$e_L^{\mathrm{B}}(\nu) \;\geq\; \frac{c_\delta \nu}{c\sqrt{1 - \gamma - c_{\delta'}\nu}} \cdot \frac{1 - \left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1 - \frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}} + \left(\frac{c_\delta \nu}{2c\,(1 - \gamma - c_{\delta'}\nu)^{3/2}}\right)^{L-1} \frac{c_\delta \nu}{c\sqrt{x_0 - c_{\delta'}\nu}}.$$

For the upper bound on $e_L^{\mathrm{E2H}}(\nu)$, ignoring the trivial case $t = 0$, we start from

$$e_1^{\mathrm{E2H}}(\nu) = H_{0,0}(x_0) - H_{0,\nu}(x_0) = \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}.$$

More generally, for every $t \geq 1$ we can write

$$e_t^{\mathrm{E2H}}(\nu) = H_{t-1,0}\big(x_{t-1}^{\mathrm{E2H}}(0)\big) - H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(\nu)\big)$$

$$= \Big(H_{t-1,0}\big(x_{t-1}^{\mathrm{E2H}}(0)\big) - H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(0)\big)\Big) + \Big(H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(0)\big) - H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(\nu)\big)\Big).$$

Similarly, these two terms are nonnegative. Hence, $e_t^{\mathrm{E2H}}(\nu) \geq 0, \forall t \geq 1$.

Next we derive a quantitative upper bound. For the first difference, note that $x_{t-1}^{\mathrm{E2H}}(0) = 1-\gamma$ for all $t \geq 1$, and thus

$$H_{t-1,0}\big(x_{t-1}^{\mathrm{E2H}}(0)\big) - H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(0)\big) = \frac{c_\delta \nu}{c\sqrt{a_{t-1}(1-\gamma) - c_{\delta'}\nu}}.$$

For the second difference, we compute the derivative

$$H'_{t-1,\nu}(x) = \frac{c_\delta \nu}{2c} \cdot \frac{a_{t-1}}{(a_{t-1}x - c_{\delta'}\nu)^{3/2}},$$

which is monotonically decreasing in x over its domain. By the mean value theorem, there exists $\zeta_{t-1} \in \big[x_{t-1}^{\mathrm{E2H}}(\nu),\, 1-\gamma\big]$ such that $H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(0)\big) - H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(\nu)\big) = H'_{t-1,\nu}(\zeta_{t-1})\, e_{t-1}^{\mathrm{E2H}}(\nu)$. Since $H'_{t-1,\nu}$ is decreasing and $\zeta_{t-1} \geq x_{t-1}^{\mathrm{E2H}}(\nu)$, we have $H'_{t-1,\nu}(\zeta_{t-1}) \leq H'_{t-1,\nu}(x_{t-1}^{\mathrm{E2H}}(\nu))$. Moreover, since $\{x_t^{\mathrm{E2H}}(\nu)\}_{t\geq 0}$ is monotonically increasing and

$$x_1^{\mathrm{E2H}}(\nu) = H_{0,\nu}(x_0) = 1-\gamma - \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}},$$

we have $x_{t-1}^{\mathrm{E2H}}(\nu) \geq x_1^{\mathrm{E2H}}(\nu)$ for all $t \geq 2$. Using again that $H'_{t-1,\nu}$ is decreasing, it follows that for all $t \geq 2$, $H'_{t-1,\nu}(x_{t-1}^{\mathrm{E2H}}(\nu)) \leq H'_{t-1,\nu}\big(x_1^{\mathrm{E2H}}(\nu)\big)$. Combining the above displays yields, for every $t = 2, 3, \ldots, L$,

$$H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(0)\big) - H_{t-1,\nu}\big(x_{t-1}^{\mathrm{E2H}}(\nu)\big) \;\leq\; H'_{t-1,\nu}\big(x_1^{\mathrm{E2H}}(\nu)\big)\; e_{t-1}^{\mathrm{E2H}}(\nu),$$

where

$$H'_{t-1,\nu}\big(x_1^{\mathrm{E2H}}(\nu)\big) = \frac{c_\delta \nu}{2c} \cdot \frac{a_{t-1}}{\Big(a_{t-1}x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}.$$

Therefore, combining the two pieces, we obtain the recursion: for every $t = 2, 3, \ldots, L$,

$$e_t^{\mathrm{E2H}}(\nu) \;\leq\; \frac{c_\delta \nu}{c\sqrt{a_{t-1}(1-\gamma) - c_{\delta'}\nu}} \;+\; \frac{c_\delta \nu}{2c} \cdot \frac{a_{t-1}}{\Big(a_{t-1}x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\; e_{t-1}^{\mathrm{E2H}}(\nu).$$

Unrolling gives

$$\begin{aligned} e_L^{\mathrm{E2H}}(\nu) \leq \sum_{j=1}^{L-1} &\left(\prod_{s=j+1}^{L-1} \frac{c_\delta \nu}{2c} \cdot \frac{a_s}{\Big(a_s x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right) \cdot \frac{c_\delta \nu}{c\sqrt{a_j(1-\gamma) - c_{\delta'}\nu}} \\ &+ \left(\prod_{s=1}^{L-1} \frac{c_\delta \nu}{2c} \cdot \frac{a_s}{\Big(a_s x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right) \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}. \end{aligned}$$

Since $a_t = (t/(t+1))^\beta$ is increasing in t for $t \geq 1$, we have $a_t \geq a_1 = 2^{-\beta}$ for all $t \in [L-1]$. Hence, for every $j \in [L-1]$,

$$\frac{1}{\sqrt{a_j(1-\gamma) - c_{\delta'}\nu}} \;\leq\; \frac{1}{\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}},$$

and for every $s \in [L-1]$,

$$\frac{1}{\Big(a_s x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}} \;\leq\; \frac{1}{\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}.$$

Applying these bounds to the unrolled expression yields

$$e_L^{\mathrm{E2H}}(\nu) \leq \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \sum_{j=1}^{L-1} \left(\frac{c_\delta \nu}{2c\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1-j} \left(\prod_{s=j+1}^{L-1} a_s\right)$$

$$+\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} \left(\prod_{s=1}^{L-1} a_s\right) \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}.$$

Next, For $0 \le j \le L-1$, $\prod_{s=j+1}^{L-1} a_s = \prod_{s=j+1}^{L-1} \left(\frac{s}{s+1}\right)^{\beta} = \left(\frac{j+1}{L}\right)^{\beta}$. Substituting these identities gives

$$\begin{aligned}
e_L^{\mathrm{E2H}}(\nu) &\le \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \sum_{j=1}^{L-1} \left(\frac{c_\delta \nu}{2c\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1-j} \Big(\frac{j+1}{L}\Big)^{\beta} \\
&\quad + \left(\frac{c_\delta \nu}{2c\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}} \\
&= \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \sum_{m=0}^{L-2} \left(\frac{c_\delta \nu}{2c\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right)^{m} \Big(1-\frac{m}{L}\Big)^{\beta} \\
&\quad + \left(\frac{c_\delta \nu}{2c\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}.
\end{aligned}$$

Finally, since for $m \in \{0, 1, \ldots, L-2\}$, $(1-\frac{m}{L})^{\beta} \le e^{-\frac{\beta}{L} m}$,

$$\begin{aligned}
e_L^{\mathrm{E2H}}(\nu) &\le \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \sum_{m=0}^{L-2} \left[\frac{c_\delta \nu}{2c\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L}\right]^{m} \\
&\quad + \left(\frac{c_\delta \nu}{2c\Big(2^{-\beta} x_1^{\mathrm{E2H}}(\nu) - c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}} \\
&= \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \cdot \frac{1 - \left[\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}\right]^{L-1}}{1 - \frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}} \\
&\quad + \left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big) - c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}.
\end{aligned}$$

Therefore, given $\Delta(\nu, x_0) - \Delta(0, x_0) = e_L^{\mathrm{B}}(\nu) - a_L\, e_L^{\mathrm{E2H}}(\nu)$, combining the lower bound on $e_L^{\mathrm{B}}(\nu)$ and the upper bound on $e_L^{\mathrm{E2H}}(\nu)$ yields

$$\begin{aligned}
&\Delta(\nu, x_0) - \Delta(0, x_0) \\
&\ge \left[\frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} \cdot \frac{1-\left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1-\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}} + \left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1} \frac{c_\delta \nu}{c\sqrt{x_0 - c_{\delta'}\nu}}\right]
\end{aligned}$$

$$
- a_L \left[\frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \cdot \frac{1 - \left[\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L} \right]^{L-1}}{1 - \frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}} \right.
$$

$$
\left. + \left(\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}} \right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}} \right]
$$

$$
> \frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} \cdot \frac{1 - \left(\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1 - \frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}}
$$

$$
- a_L \left[\frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \cdot \frac{1}{1 - \frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}} \right.
$$

$$
\left. + \left(\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}} \right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}} \right]
$$

When the derived lower bound is greater than $-\frac{1}{2}\,\Delta(0,x_0) = -\frac{1}{2}\,(a_L-1)(1-\gamma)$, then it follows immediately that

$$
\Delta(\nu, x_0) = \Delta(0,x_0) + \big(\Delta(\nu,x_0) - \Delta(0,x_0)\big) \;>\; \frac{1}{2}\,(a_L-1)(1-\gamma) \;>\; 0.
$$

Therefore, we define the corresponding constraint by $\mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat\theta_0)\big) < 0$ as in Definition A.2, under this constraint, we conclude that

$$
\big(G \circ H_{L-1,\nu} \circ \cdots \circ H_{0,\nu}\big)(x_0) \;>\; F_\nu^{\circ L}(x_0),
$$

i.e., the easy-to-hard lower bound is strictly larger than the baseline lower bound. $\square$

C.2 Proof of Corollary 5.3

Proof. Let

$$
H_0(V) \;:=\; 1-\gamma - \frac{c_\delta \nu}{c\sqrt{a_0 V - c_{\delta'}\nu}},
$$

which is strictly increasing in V. Then, the inequality $x_-(2^{-\beta},\nu) < H_0\big(V_{p_0}(\hat\theta_0)\big) < x_+(2^{-\beta},\nu)$ is equivalent to $V_-(\nu) < V_{p_0}(\hat\theta_0) < V_+(\nu)$, where $V_\pm(\nu)$ are the unique solutions to $H_0(V) = x_\pm(2^{-\beta},\nu)$, respectively. Solving this gives

$$
\sqrt{a_0 V_\pm(\nu) - c_{\delta'}\nu} = \frac{c_\delta \nu}{c\,(1-\gamma - x_\pm(2^{-\beta},\nu))}.
$$

Moreover, using the fact that $x_\pm(2^{-\beta},\nu)$ are the fixed points, i.e.,

$$
x_\pm(2^{-\beta},\nu) = 1-\gamma - \frac{c_\delta\nu}{c\sqrt{2^{-\beta}x_\pm(2^{-\beta},\nu) - c_{\delta'}\nu}},
$$

we obtain

$$
\frac{c_\delta \nu}{c\,(1-\gamma - x_\pm(2^{-\beta},\nu))} = \sqrt{2^{-\beta}x_\pm(2^{-\beta},\nu) - c_{\delta'}\nu}\,.
$$

Therefore, $V_\pm(\nu) = \frac{2^{-\beta}}{a_0}\,x_\pm(2^{-\beta},\nu)$. Consequently, the second constraint is equivalent to

$$
\frac{2^{-\beta}}{a_0} x_-(2^{-\beta},\nu) < V_{p_0}(\hat\theta_0) < \frac{2^{-\beta}}{a_0} x_+(2^{-\beta},\nu).
$$

Intersecting with the first constraint $x_-(2^{-\beta},\nu) < V_{p_0}(\hat{\theta}_0) < x_+(2^{-\beta},\nu)$ yields that the feasible set of $V_{p_0}(\hat{\theta}_0)$ is

$$\mathcal{I}_{\mathcal{M}}(\beta',\beta,\nu) = \Big(x_-(2^{-\beta},\nu),\ \frac{2^{-\beta}}{a_0}x_+(2^{-\beta},\nu)\Big).$$

For the interval length $\big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,\nu)\big|$, we have

$$\begin{aligned}\big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,\nu)\big| &= \big(x_+(2^{-\beta},\nu) - x_-(2^{-\beta},\nu)\big) - \Big(1-\frac{2^{-\beta}}{a_0}\Big)x_+(2^{-\beta},\nu) \\ &\geq \big(x_+(2^{-\beta},\nu) - x_-(2^{-\beta},\nu)\big) - \Big(1-\frac{2^{-\beta}}{a_0}\Big)(1-\gamma).\end{aligned}$$

Applying Corollary B.4 with parameter $2^{-\beta}$ gives

$$\big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,\nu)\big| \ \geq\ \frac{2^{-\beta}}{a_0}(1-\gamma) - 2^{\beta}c_{\delta'}\nu - \frac{3\sqrt{3}}{2}\cdot\frac{c_\delta\nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}}.$$

By Corollary B.4, we have $x_-(2^{-\beta},0) = 0$ and $x_+(2^{-\beta},0) = 1-\gamma$, hence $\big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,0)\big| = \frac{2^{-\beta}}{a_0}(1-\gamma)$. Therefore,

$$\big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,0)\big| - \big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,\nu)\big| \ \leq\ 2^{\beta}c_{\delta'}\nu + \frac{3\sqrt{3}}{2}\cdot\frac{c_\delta\nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}}.$$

Moreover,

$$\begin{aligned}\big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,0)\big| - \big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,\nu)\big| &= \frac{2^{-\beta}}{a_0}(1-\gamma) - \Big(\frac{2^{-\beta}}{a_0}x_+(2^{-\beta},\nu) - x_-(2^{-\beta},\nu)\Big) \\ &= \frac{2^{-\beta}}{a_0}\big(1-\gamma-x_+(2^{-\beta},\nu)\big) + x_-(2^{-\beta},\nu).\end{aligned}$$

Since $x_+(2^{-\beta},\nu) < 1-\gamma$, the first term above is nonnegative, hence $\big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,0)\big| - \big|\mathcal{I}_{\mathcal{M}}(\beta',\beta,\nu)\big| \geq x_-(2^{-\beta},\nu)$. By the definition of $x_-(2^{-\beta},\nu)$,

$$x_-(2^{-\beta},\nu) = 2^{\beta}c_{\delta'}\nu + \Big(1-\gamma-2^{\beta}c_{\delta'}\nu\Big)\,y_-(\nu)\ \geq\ 2^{\beta}c_{\delta'}\nu,$$

which completes the proof. □

C.3 Proof of Proposition 5.5

Proof. Proposition 5.5 follows immediately by combining Lemma C.1, Lemma C.2, and Lemma C.3. □

Lemma C.1. *Under the notation of Proposition 5.5, for fixed (β',β), the length of the interval $\mathcal{I}_{\mathcal{N}}(\beta',\beta,\nu)$ for which $\mathcal{N}\big(\beta',\beta,\nu,V_{p_0}(\hat{\theta}_0)\big) < 0$ is monotonically decreasing in ν, and $\mathcal{I}_{\mathcal{N}}(\beta',\beta,\nu)$ can be written in the form*

$$\mathcal{I}_{\mathcal{N}}(\beta',\beta,\nu) = \big(x(\beta',\beta,\nu),\ 1-\gamma\big).$$

Proof. Let $x_0 := V_{p_0}(\hat{\theta}_0)$ for convenience. First, we show that the additional constraint $\mathcal{N}(\beta',\beta,\nu,x_0) < 0$ induces an admissible region that is monotonically decreasing in ν.

Define

$$\mathcal{E}(\beta',\beta,\nu,x_0) := \frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} \cdot \frac{1-\left(\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1-\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}}$$
$$- a_L\left[\frac{c_\delta\nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}} \cdot \frac{1}{1-\frac{c_\delta\nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\cdot e^{-\beta/L}}\right.$$
$$\left.+\left(\frac{c_\delta\nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\right)^{L-1} L^{-\beta}\cdot\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right].$$

Then, by Theorem 5.2, $\mathcal{N}(\beta',\beta,\nu,x_0)<0$ is equivalent to $\mathcal{E}(\beta',\beta,\nu,x_0) > -\frac{1}{2}\,(a_L-1)(1-\gamma)$. Moreover, the right-hand side above is a fixed constant since β' is fixed.

Next, note that $\mathcal{E}(\beta',\beta,\nu,x_0)$ is monotonically increasing in x_0. Moreover, as x_0 decreases to 0, for denominators, the term

$$1-\frac{c_\delta\nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\cdot e^{-\beta/L}$$

is the first to approach 0, which implies that $\mathcal{E}(\beta',\beta,\nu,x_0)\to-\infty$. On the other hand, when $\nu=0$ we have

$$\mathcal{E}(\beta',\beta,0,x_0)=0>-\frac{1}{2}\,(a_L-1)(1-\gamma)$$

for all x_0. Hence, by continuity, for sufficiently small $\nu>0$, the set of x_0 satisfying $\mathcal{E}(\beta',\beta,\nu,x_0) > -\frac{1}{2}\,(a_L-1)(1-\gamma)$ is non-empty. Combining the above, as ν increases gradually from 0, the set of x_0 such that $\mathcal{N}(\beta',\beta,\nu,x_0)<0$ must lie in a non-empty interval of the form

$$\mathcal{I}_{\mathcal{N}}(\beta',\beta,\nu)=\big(x(\beta',\beta,\nu),\,1-\gamma\big).$$

Therefore, it suffices to show that the minimal admissible threshold $x(\beta',\beta,\nu)$ is monotonically increasing in ν. Define

$$\Phi(\beta',\beta,\nu,x) \;:=\; \mathcal{E}(\beta',\beta,\nu,x)+\frac{1}{2}\,(a_L-1)(1-\gamma),$$

By definition of $x(\beta',\beta,\nu)$, we have $\Phi\big(\beta',\beta,\nu,x(\beta',\beta,\nu)\big)=0$. The implicit function theorem implies that

$$x'(\nu)=-\frac{\partial_\nu\Phi\big(\beta',\beta,\nu,x(\beta',\beta,\nu)\big)}{\partial_x\Phi\big(\beta',\beta,\nu,x(\beta',\beta,\nu)\big)}=-\frac{\partial_\nu\mathcal{E}\big(\beta',\beta,\nu,x(\beta',\beta,\nu)\big)}{\partial_x\mathcal{E}\big(\beta',\beta,\nu,x(\beta',\beta,\nu)\big)}.$$

Since $\partial_x\mathcal{E}(\beta',\beta,\nu,x)>0$, to show $x'(\nu)>0$, it suffices to show that $\mathcal{E}(\beta',\beta,\nu,x_0)$ is monotonically decreasing in ν.

Note first that the last term inside the brackets,

$$\left(\frac{c_\delta\nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\right)^{L-1} L^{-\beta}\cdot\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}},$$

is increasing as ν increases. Hence, after multiplying by the coefficient $-a_L$, this contribution is decreasing in ν. It therefore suffices to prove that the remaining part,

$$\frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} \cdot \frac{1-\left(\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1-\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}}$$

$$- \; a_L \left[\frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \cdot \frac{1}{1 - \frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}} \right].$$

has strictly negative derivative with respect to ν.

To this end, it is enough to verify the following two comparison statements: first, both the function value and the derivative (with respect to ν) of

$$U(\nu) = \frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}}$$

are strictly smaller than those of

$$\widetilde{U}(\nu) = \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}},$$

and second, both the function value and the derivative (with respect to ν) of

$$V(\nu) = \frac{1 - \left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1 - \frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}}$$

are strictly smaller than those of

$$\widetilde{V}(\nu) = \frac{1}{1 - \frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}}.$$

Indeed, by the product rule,

$$\frac{\mathrm{d}}{\mathrm{d}\nu}\Big(U(\nu)V(\nu) - a_L\,\widetilde{U}(\nu)\widetilde{V}(\nu)\Big) = \Big(U'(\nu)V(\nu)+U(\nu)V'(\nu)\Big) - a_L\Big(\widetilde{U}'(\nu)\widetilde{V}(\nu)+\widetilde{U}(\nu)\widetilde{V}'(\nu)\Big) < 0,$$

since $a_L > 1$, and

$$U(\nu), \widetilde{U}(\nu), V(\nu), \widetilde{V}(\nu), U'(\nu), \widetilde{U}'(\nu), V'(\nu), \widetilde{V}'(\nu) > 0.$$

We first prove that $U(\nu) < \widetilde{U}(\nu)$. Since $0 < 2^{-\beta} < 1$, we have

$$2^{-\beta}(1-\gamma) - c_{\delta'}\nu \;<\; 1-\gamma-c_{\delta'}\nu,$$

and hence $U(\nu) < \widetilde{U}(\nu)$.

We second prove that $U'(\nu) < \widetilde{U}'(\nu)$. By direct differentiation, we have

$$U'(\nu) = \frac{c_\delta}{c}\left((1-\gamma-c_{\delta'}\nu)^{-1/2} + \frac{c_{\delta'}\nu}{2}(1-\gamma-c_{\delta'}\nu)^{-3/2}\right).$$

Similarly,

$$\widetilde{U}'(\nu) = \frac{c_\delta}{c}\left((2^{-\beta}(1-\gamma)-c_{\delta'}\nu)^{-1/2} + \frac{c_{\delta'}\nu}{2}(2^{-\beta}(1-\gamma)-c_{\delta'}\nu)^{-3/2}\right).$$

Since $2^{-\beta}(1-\gamma) - c_{\delta'}\nu < (1-\gamma) - c_{\delta'}\nu$, $U'(\nu) < \widetilde{U}'(\nu)$.

We third prove that $V(\nu) < \widetilde{V}(\nu)$ We can rewrite $V(\nu)$ as the finite geometric sum

$$V(\nu) = \sum_{j=0}^{L-2}\left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^j.$$

Similarly,

$$\widetilde{V}(\nu) = \sum_{j=0}^{\infty}\left(\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}\right)^j.$$

Therefore, it is enough to show that the common ratio of the geometric sum defining $V(\nu)$ is strictly smaller than the common ratio of the geometric series defining $\widetilde{V}(\nu)$, namely,

$$\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}} \;<\; \frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L}.$$

This is equivalent to

$$2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big)-c_{\delta'}\nu \;<\; e^{-2\beta/(3L)}\,(1-\gamma-c_{\delta'}\nu).$$

Since $L \geq 2$ implies $\frac{1}{L} < \frac{3}{2}\log 2$, we have $e^{\beta(\frac{3}{2}\log 2 - \frac{1}{L})} > 1$. Rearranging gives $e^{-2\beta/(3L)} > 2^{-\beta}$. Moreover,

$$2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big)-c_{\delta'}\nu \;\leq\; 2^{-\beta}(1-\gamma)-c_{\delta'}\nu,$$

and

$$2^{-\beta}(1-\gamma)-c_{\delta'}\nu \;\leq\; e^{-2\beta/(3L)}(1-\gamma)-e^{-2\beta/(3L)}c_{\delta'}\nu = e^{-2\beta/(3L)}(1-\gamma-c_{\delta'}\nu).$$

Combining the last two displays gives

$$2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big)-c_{\delta'}\nu \;<\; e^{-2\beta/(3L)}(1-\gamma-c_{\delta'}\nu),$$

which proves the desired ratio inequality.

We finally prove that $V'(\nu) < \widetilde{V}'(\nu)$. Differentiating term-by-term, we obtain

$$V'(\nu) = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)\cdot \sum_{j=1}^{L-2} j\left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{j-1},$$

and

$$\widetilde{V}'(\nu) = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L}\right)$$
$$\cdot \sum_{j=1}^{\infty} j\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L}\right)^{j-1}.$$

Since we have already shown that

$$\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}} \;<\; \frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L},$$

it remains to prove that

$$\frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right) \;<\; \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L}\right).$$

For the left-hand side, direct differentiation gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right) &= \frac{c_\delta}{2c}\left((1-\gamma-c_{\delta'}\nu)^{-3/2}+\frac{3}{2}c_{\delta'}\nu\,(1-\gamma-c_{\delta'}\nu)^{-5/2}\right)\\ &= \frac{c_\delta}{2c}\cdot\frac{1-\gamma+\frac{1}{2}c_{\delta'}\nu}{(1-\gamma-c_{\delta'}\nu)^{5/2}}.\end{aligned}$$

For the right-hand side, direct differentiation yields

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}}\cdot e^{-\beta/L}\right)\\
&\quad=\frac{c_\delta}{2c}\,e^{-\beta/L}\Bigg[\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{-3/2}\\
&\qquad-\frac{3}{2}\nu\,\frac{\mathrm{d}}{\mathrm{d}\nu}\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)\cdot\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{-5/2}\Bigg]\\
&\quad=\frac{c_\delta}{2c}\,e^{-\beta/L}\cdot\frac{2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu-\frac{3}{2}\nu\,\frac{\mathrm{d}}{\mathrm{d}\nu}\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)}{\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{5/2}}.
\end{aligned}
$$

Next we compute the inner derivative explicitly:

$$
\frac{\mathrm{d}}{\mathrm{d}\nu}\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)=2^{-\beta}\left(-\frac{c_\delta}{c\sqrt{a_0x_0-c_{\delta'}\nu}}-\frac{c_\delta\nu}{c}\cdot\frac{c_{\delta'}}{2}\,(a_0x_0-c_{\delta'}\nu)^{-3/2}\right)-c_{\delta'}.
$$

Substituting this expression back and simplifying, we obtain

$$
\begin{aligned}
&2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu-\frac{3}{2}\nu\,\frac{\mathrm{d}}{\mathrm{d}\nu}\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)\\
&\quad=2^{-\beta}(1-\gamma)+2^{-\beta-1}\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}+\frac{1}{2}\,c_{\delta'}\nu+3\cdot 2^{-\beta-2}\frac{c_\delta c_{\delta'}\nu^2}{c\,(a_0x_0-c_{\delta'}\nu)^{3/2}}\\
&\quad\ge 2^{-\beta}(1-\gamma)+\frac{1}{2}\,c_{\delta'}\nu.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}}\cdot e^{-\beta/L}\right)\\
&\qquad\ge\frac{c_\delta}{2c}\,e^{-\beta/L}\cdot\frac{2^{-\beta}(1-\gamma)+\frac{1}{2}c_{\delta'}\nu}{\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{5/2}}.
\end{aligned}
$$

Moreover, we have

$$
2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\;\le\;2^{-\beta}(1-\gamma)-c_{\delta'}\nu\;\le\;2^{-\beta}(1-\gamma-c_{\delta'}\nu),
$$

Hence, plugging this into the previous display yields

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}}\cdot e^{-\beta/L}\right)\\
&\qquad\ge\frac{c_\delta}{2c}\cdot\frac{e^{-\beta/L}\Big(2^{3\beta/2}(1-\gamma)+2^{5\beta/2}\cdot\frac{1}{2}c_{\delta'}\nu\Big)}{(1-\gamma-c_{\delta'}\nu)^{5/2}}.
\end{aligned}
$$

Finally, since $L\ge 2$ implies $\frac{1}{L}<\frac{3}{2}\log 2$, we have $e^{-\beta/L}2^{3\beta/2}=e^{\beta(\frac{3}{2}\log 2-\frac{1}{L})}>1$, and similarly (because $\frac{1}{L}<\frac{5}{2}\log 2$) we have $e^{-\beta/L}2^{5\beta/2}=e^{\beta(\frac{5}{2}\log 2-\frac{1}{L})}>1$. Therefore,

$$
e^{-\beta/L}\Big(2^{3\beta/2}(1-\gamma)+2^{5\beta/2}\cdot\frac{1}{2}c_{\delta'}\nu\Big)>1-\gamma+\frac{1}{2}c_{\delta'}\nu,
$$

which implies

$$\frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}}\cdot e^{-\beta/L}\right)$$
$$> \frac{c_\delta}{2c}\cdot\frac{1-\gamma+\frac{1}{2}c_{\delta'}\nu}{(1-\gamma-c_{\delta'}\nu)^{5/2}} = \frac{\mathrm{d}}{\mathrm{d}\nu}\left(\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)$$

as desired. □

Lemma C.2. *Under the notation of Proposition 5.5 and Lemma C.1, for fixed (β', β), the threshold $x(\nu)$ satisfies $x(0) = 0$ and*

$$x'(\nu) = \frac{c_{\delta'}}{a_0} + \frac{2}{a_0}\Big(\frac{c_\delta}{c(1-\gamma)}\Big)^2\nu + O\big(\nu^{5/3}\big) \quad \textit{as } \nu \to 0,$$

where $a_0 = L/\sum_{i=1}^{L} i^{-\beta'}$.

Proof. By the notation and proof of Lemma C.1, for fixed (β', β), it suffices to prove that, as $\nu \to 0$,

$$x(\nu) = \frac{c_{\delta'}}{a_0}\nu + \frac{1}{a_0}\Big(\frac{c_\delta}{c(1-\gamma)}\Big)^2\nu^2 + O(\nu^{8/3}).$$

By definition of the threshold $x(\nu)$, it satisfies the boundary condition $\mathcal{E}(\beta', \beta, \nu, x(\nu)) = -\frac{1}{2}\,(a_L - 1)(1-\gamma)$. Note that the right-hand side is $\Theta(1)$. Consequently, as $\nu \to 0$, the left-hand side must also be $\Theta(1)$. Equivalently, if we expand $\mathcal{E}(\beta', \beta, \nu, x(\nu))$ in a Puiseux series, then the smallest exponent of ν appearing in the expansion of $\mathcal{E}(\beta', \beta, \nu, x(\nu))$ must be 0.

We analyze $\mathcal{E}(\beta', \beta, \nu, x(\nu))$ term-by-term. Consider first

$$T_1(\nu) := \frac{c_\delta\nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}}\cdot\frac{1-\Big(\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\Big)^{L-1}}{1-\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}}.$$

As $\nu \to 0$,

$$\frac{c_\delta\nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} = \frac{c_\delta}{c\sqrt{1-\gamma}}\,\nu + O(\nu^2).$$

Next, define

$$q(\nu) := \frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}.$$

Then we have

$$q(\nu) = \frac{c_\delta}{2c(1-\gamma)^{3/2}}\,\nu + O(\nu^2),$$

Moreover,

$$\frac{1-q(\nu)^{L-1}}{1-q(\nu)} = \sum_{j=0}^{L-2} q(\nu)^j = 1 + q(\nu) + O\big(q(\nu)^2\big) = 1 + O(\nu),$$

Therefore,

$$T_1(\nu) = \frac{c_\delta}{c\sqrt{1-\gamma}}\,\nu + O(\nu^2).$$

In particular, the smallest power of ν appearing in the Puiseux expansion of $T_1(\nu)$ is ν. Consequently, $T_1(\nu)$ cannot contribute a $\Theta(1)$ term to $\mathcal{E}(\beta', \beta, \nu, x(\nu))$ as $\nu \to 0$.

Next we analyze

$$T_2(\nu) := \left(\frac{c_\delta\nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0x(\nu)-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta}\cdot\frac{c_\delta\nu}{c\sqrt{a_0x(\nu)-c_{\delta'}\nu}}.$$

Our goal is to determine whether there exists a choice of $x(\nu)$ such that $T_2(\nu)$ can contribute a $\Theta(1)$ term to $\mathcal{E}(\beta', \beta, \nu, x(\nu))$ as $\nu \to 0$. To this end, consider the Puiseux expansion of

$$r(\nu) := \frac{c_\delta \nu}{c\sqrt{a_0 x(\nu) - c_{\delta'}\nu}}.$$

When $r(\nu)$ has a negative leading exponent, $r(\nu) \to +\infty$ as $\nu \to 0$. Then the inner radicand $2^{-\beta}\Big(1-\gamma-r(\nu)\Big) - c_{\delta'}\nu$ tends to $-\infty$, and hence becomes negative for all sufficiently small ν. This violates our standing well-definedness requirement that every denominator appearing in $\mathcal{E}(\beta', \beta, \nu, x(\nu))$ remain strictly positive.

When $r(\nu)$ has a positive leading exponent, $r(\nu) = o(1)$ as $\nu \to 0$. In order for $T_2(\nu)$ to be $\Theta(1)$, the factor

$$\left(\frac{c_\delta \nu}{2c\Big(2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1}$$

must contribute a negative power of ν, which forces

$$\frac{c_\delta \nu}{2c\Big(2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu\Big)^{3/2}} \to +\infty \qquad \text{as } \nu \to 0.$$

However, this implies that in the term

$$T_3(\nu) := \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \cdot \frac{1}{1 - \frac{c_\delta \nu}{2c\Big(2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L}},$$

the denominator becomes negative for all sufficiently small ν, again contradicting the requirement that all denominators remain strictly positive.

Consequently, the only remaining possibility consistent with well-definedness is that $r(\nu)$ has leading exponent 0. Therefore,

$$\frac{c_\delta \nu}{2c\Big(2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu\Big)^{3/2}}$$

must also have leading exponent 0. We now show that this is equivalent to the existence of a constant $d_2 > 0$ such that $r(\nu) = 1 - \gamma - d_2\,\nu^{2/3} + o(\nu^{2/3})$. Since the above quantity is $\Theta(1)$, its denominator must satisfy $2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu = \Theta(\nu^{2/3})$. Since $\nu = o(\nu^{2/3})$ as $\nu \to 0$, we have $1 - \gamma - r(\nu) = \Theta(\nu^{2/3})$. Thus there exists a constant $d > 0$ such that $1 - \gamma - r(\nu) = d_2\,\nu^{2/3} + o(\nu^{2/3})$, where the sign $d > 0$ is required to ensure the radicand $2^{-\beta}(1-\gamma-r(\nu)) - c_{\delta'}\nu$ (and hence the denominator) remains positive. Note that the value of d is not universal: it is determined by the matching condition coming from the $\Theta(1)$ order balance (in particular, it depends on $-\frac{1}{2}\,(a_L-1)(1-\gamma)$). Equivalently, $r(\nu) = 1 - \gamma - d_2\,\nu^{2/3} + o(\nu^{2/3})$, as claimed.

Next, from

$$\frac{c_\delta \nu}{c\sqrt{a_0 x(\nu) - c_{\delta'}\nu}} = (1-\gamma) - d_2\,\nu^{2/3} + o(\nu^{2/3}),$$

rearranging gives

$$\sqrt{a_0 x(\nu) - c_{\delta'}\nu} = \frac{c_\delta \nu}{c\big((1-\gamma) - d_2\,\nu^{2/3} + o(\nu^{2/3})\big)}.$$

Using the expansion

$$\frac{1}{(1-\gamma) - d_2\,\nu^{2/3} + o(\nu^{2/3})} = \frac{1}{1-\gamma}\left(1 + \frac{d_2}{1-\gamma}\nu^{2/3} + o(\nu^{2/3})\right),$$

we obtain

$$\sqrt{a_0 x(\nu) - c_{\delta'}\nu} = \frac{c_\delta}{c(1-\gamma)}\nu + \frac{c_\delta d_2}{c(1-\gamma)^2}\nu^{5/3} + o(\nu^{5/3}).$$

Squaring yields

$$a_0 x(\nu) - c_{\delta'}\nu = \Big(\frac{c_\delta}{c(1-\gamma)}\Big)^2 \nu^2 + O(\nu^{8/3}),$$

Therefore,

$$x(\nu) = \frac{c_{\delta'}}{a_0}\nu + \frac{1}{a_0}\Big(\frac{c_\delta}{c(1-\gamma)}\Big)^2 \nu^2 + O(\nu^{8/3}),$$

as claimed.

Finally, we analyze the choice of $x(\nu)$ for which

$$T_3(\nu) := \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \cdot \frac{1}{1 - \frac{c_\delta \nu}{2c\Big(2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L}}$$

contributes a $\Theta(1)$ term. First, as $\nu \to 0$ we have

$$\frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} = \frac{c_\delta}{c\sqrt{2^{-\beta}(1-\gamma)}}\,\nu + O(\nu^2).$$

Hence, in order for $T_3(\nu)$ to be $\Theta(1)$,

$$1 - \frac{c_\delta \nu}{2c\Big(2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L} = \Theta(\nu)$$

Equivalently, there exists a constant $d_3 > 0$ such that

$$\frac{c_\delta \nu}{2c\Big(2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu\Big)^{3/2}} = e^{\beta/L}\big(1 - d_3\nu + o(\nu)\big).$$

Invert the above display to obtain

$$\Big(2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu\Big)^{3/2} = \frac{c_\delta}{2c}\, e^{-\beta/L}\,\nu \cdot \frac{1}{1 - d_3'\nu + o(\nu)} = \frac{c_\delta}{2c}\, e^{-\beta/L}\,\nu\big(1 + d_3'\nu + o(\nu)\big),$$

i.e.,

$$2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu = \Big(\frac{c_\delta}{2c}\Big)^{2/3} e^{-2\beta/(3L)}\,\nu^{2/3}\Big(1 + d_3'\nu + o(\nu)\Big)^{2/3}.$$

Using the binomial expansion $(1+t)^{2/3} = 1 + \frac{2}{3}t + o(t)$ as $t \to 0$, we obtain

$$2^{-\beta}\big(1-\gamma-r(\nu)\big) - c_{\delta'}\nu = \Big(\frac{c_\delta}{2c}\Big)^{2/3} e^{-2\beta/(3L)}\,\nu^{2/3} + \frac{2}{3}\Big(\frac{c_\delta}{2c}\Big)^{2/3} e^{-2\beta/(3L)}\, d_3'\,\nu^{5/3} + o(\nu^{5/3}).$$

Rearranging yields

$$r(\nu) = 1 - \gamma - 2^{\beta}\Big(\frac{c_\delta}{2c}\Big)^{2/3} e^{-2\beta/(3L)}\,\nu^{2/3} + o(\nu^{2/3}).$$

Substituting this asymptotic form of $r(\nu)$ back into its definition $r(\nu) = \frac{c_\delta \nu}{c\sqrt{a_0 x(\nu) - c_{\delta'}\nu}}$ and repeating the same algebra as above yields

$$x(\nu) = \frac{c_{\delta'}}{a_0}\nu + \frac{1}{a_0}\Big(\frac{c_\delta}{c(1-\gamma)}\Big)^2 \nu^2 + O(\nu^{8/3}).$$

Moreover, it is straightforward to verify that if we substitute the $r(\nu)$ obtained from the $\Theta(1)$-balancing of $T_2(\nu)$ into $T_3(\nu)$, or conversely substitute the $r(\nu)$ obtained from the $\Theta(1)$-balancing of $T_3(\nu)$ into $T_2(\nu)$, then in both cases the resulting Puiseux expansion does not introduce any negative leading power of ν. In particular, neither substitution violates the well-definedness regime.

Consequently, regardless of whether the $\Theta(1)$ constant term in $\mathcal{E}(\beta', \beta, \nu, x(\nu))$ is predominantly contributed by $T_2(\nu)$ or by $T_3(\nu)$, and regardless of the specific constant value on the right-hand side $-\frac{1}{2}(a_L - 1)(1-\gamma)$, both scenarios yield the same expansion for the threshold $x(\nu)$ up to order ν^2:

$$x(\nu) = \frac{c_{\delta'}}{a_0}\nu + \frac{1}{a_0}\Big(\frac{c_\delta}{c(1-\gamma)}\Big)^2 \nu^2 + O(\nu^{8/3}).$$

The only difference lies in higher-order coefficients (beyond the ν^2 term), which does not affect our conclusion. Which of $T_2(\nu)$ or $T_3(\nu)$ provides the dominant $\Theta(1)$ contribution depends on the finer $\Theta(1)$ matching (i.e., the constant-level balance) in the boundary condition $\mathcal{E}(\beta', \beta, \nu, x(\nu)) = -\frac{1}{2}(a_L - 1)(1-\gamma)$, and hence on the specific interplay among (β', β, L) and the constants $(c_\delta, c_{\delta'}, c, \gamma)$ through the corresponding $\Theta(1)$ coefficients. $\square$

Lemma C.3. *Under the notation of Proposition 5.5 and Lemma C.1, for fixed (β', β), let*

$$\mathcal{N}_\infty(\nu) \;:=\; \lim_{V_{p_0}(\hat{\theta}_0)\to\infty} \mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big).$$

Then there exists a unique $\nu_c > 0$ such that $\mathcal{N}_\infty(\nu_c) = 0$, and there exists a constant $C(\nu_c) > 0$ such that

$$x'(\nu) \;=\; \frac{C(\nu_c)}{(\nu_c - \nu)^3}\big(1 + O(\nu_c - \nu)\big) \quad \textit{as } \nu \uparrow \nu_c.$$

Proof. By the notation and proof of Lemma C.1, for fixed (β', β), define $\mathcal{E}_\infty(\nu) := \lim_{x_0\to\infty} \mathcal{E}(\beta', \beta, \nu, x_0)$. Then,

$$\mathcal{E}_\infty(\nu) = \frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} \cdot \frac{1 - \left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1 - \frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}}$$
$$-\; a_L \cdot \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \cdot \frac{1}{1 - \frac{c_\delta \nu}{2c\left(2^{-\beta}(1-\gamma)-c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}}.$$

We first prove $\mathcal{N}'_\infty(\nu) > 0$ and the existence of a unique $\nu_c > 0$ such that $\mathcal{N}_\infty(\nu_c) = 0$. It suffices to prove that $\mathcal{E}'_\infty(\nu) < 0$ and that there exists a unique $\nu_c > 0$ such that $\mathcal{E}_\infty(\nu_c) = -\frac{1}{2}\,(a_L - 1)(1-\gamma)$.

In the proof of Lemma C.1, it is straightforward to verify that $\mathcal{E}(\beta', \beta, \nu, x_0)$ is monotonically decreasing in ν also holds when $x_0 \to \infty$, i.e., setting

$$r(\nu, x_0) \;:=\; \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}} \;=\; 0,$$

and the same monotonicity argument continues to hold. Hence $\mathcal{E}'_\infty(\nu) \;<\; 0$. Moreover, we clearly have $\mathcal{E}_\infty(0) = 0$. On the other hand, when ν is large enough so that

$$1 - \frac{c_\delta \nu}{2c\big(2^{-\beta}(1-\gamma) - c_{\delta'}\nu\big)^{3/2}} \cdot e^{-\beta/L}$$

approaches 0, the second term in $\mathcal{E}_\infty(\nu)$ diverges to $-\infty$, and thus $\mathcal{E}_\infty(\nu) \to -\infty$. Therefore, by continuity and strict monotonicity of $\mathcal{E}_\infty(\nu)$ in ν, there exists a unique $\nu_c > 0$ such that $\mathcal{E}_\infty(\nu_c) = -\frac{1}{2}\,(a_L - 1)(1-\gamma)$.

Next, define

$$T(\nu, x_0) \;:=\; 1 - \frac{c_\delta \nu}{2c\Big(2^{-\beta}(1-\gamma-r(\nu,x_0)) - c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L},$$

and

$$T(\nu, \infty) \;=\; 1 - \frac{c_\delta \nu}{2c\Big(2^{-\beta}(1-\gamma) - c_{\delta'}\nu\Big)^{3/2}} \cdot e^{-\beta/L}.$$

Moreover, since

$$2^{-\beta}(1-\gamma-r(\nu,x_0)) - c_{\delta'}\nu = \Big(2^{-\beta}(1-\gamma) - c_{\delta'}\nu\Big) - 2^{-\beta} r(\nu, x_0),$$

a Taylor expansion yields, as $x_0 \to \infty$,

$$T(\nu,\infty) - T(\nu, x_0) = \frac{3c_\delta \nu\, 2^{-\beta} e^{-\beta/L}}{4c\,\big(2^{-\beta}(1-\gamma) - c_{\delta'}\nu\big)^{5/2}}\, r(\nu, x_0) \;+\; O\big(r(\nu,x_0)^2\big),$$

Next, since $T(\nu,\infty) > 0$ in the well-definedness regime, we may expand the reciprocal around $T(\nu,\infty)$ as

$$\frac{1}{T(\nu,x_0)} - \frac{1}{T(\nu,\infty)} = \frac{T(\nu,\infty)-T(\nu,x_0)}{T(\nu,\infty)^2} \;+\; O\Big(\big(T(\nu,\infty)-T(\nu,x_0)\big)^2\Big), \qquad x_0 \to \infty.$$

Furthermore, since $T(\nu,\infty)-T(\nu,x_0) = O\big(r(\nu,x_0)\big)$,

$$\frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}}\left(\frac{1}{T(\nu,x_0)} - \frac{1}{T(\nu,\infty)}\right) = C_1(\nu)\, r(\nu,x_0) + O\big(r(\nu,x_0)^2\big), \qquad x_0 \to \infty,$$

where, after combining like terms, the coefficient $C_1(\nu)$ is given by

$$C_1(\nu) = \frac{3\cdot 2^{-\beta} e^{-\beta/L}(c_\delta\nu)^2}{4}\,\frac{1}{c^2\Big(2^{-\beta}(1-\gamma)-c_{\delta'}\nu\Big)^3\left(1-\frac{c_\delta\nu}{2c\Big(2^{-\beta}(1-\gamma)-c_{\delta'}\nu\Big)^{3/2}}\,e^{-\beta/L}\right)^2}.$$

Second, we can similarly obtain, by a Taylor expansion around $r(\nu,x_0)=0$, that

$$\begin{aligned}&\left(\frac{c_\delta\nu}{2c\Big(2^{-\beta}(1-\gamma-r(\nu,x_0))-c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta}\cdot r(\nu,x_0)\\ &\quad= \left(\frac{c_\delta\nu}{2c\Big(2^{-\beta}(1-\gamma)-c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta}\cdot r(\nu,x_0) \;+\; O\big(r(\nu,x_0)^2\big), \qquad x_0\to\infty.\end{aligned}$$

Therefore, if we define

$$C_2(\nu) := \left(\frac{c_\delta\nu}{2c\Big(2^{-\beta}(1-\gamma)-c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta},$$

then combining the previous expansion with the definition of $\mathcal{E}_\infty(\nu)$ yields, as $x_0\to\infty$,

$$\mathcal{E}_\infty(\nu) - \mathcal{E}(\beta',\beta,\nu,x_0) = a_L\Big(C_1(\nu)+C_2(\nu)\Big)\, r(\nu,x_0) \;+\; O\big(r(\nu,x_0)^2\big).$$

Moreover, by the boundary equation defining $x(\nu)$, $\mathcal{E}(\beta',\beta,\nu,x(\nu)) = -\frac{1}{2}(a_L-1)(1-\gamma)$. Therefore

$$\mathcal{E}_\infty(\nu) + \frac{1}{2}(a_L-1)(1-\gamma) = a_L\Big(C_1(\nu)+C_2(\nu)\Big)\, r\big(\nu,x(\nu)\big) \;+\; O\big(r\big(\nu,x(\nu)\big)^2\big).$$

We next justify that $x(\nu)$ goes to infinity as ν approaches ν_c. Suppose, for contradiction, that $x(\nu)$ does not diverge as $\nu\uparrow\nu_c$. Then there exist a sequence $\nu_n\uparrow\nu_c$ and a constant $M<\infty$ such that $x(\nu_n)\le M$ for all n. By passing to a subsequence, we may assume $x(\nu_n)\to x_\star$ for some $x_\star\in(0,M]$. By continuity of $\mathcal{E}(\beta',\beta,\nu,x_0)$ in (ν,x_0) within the well-definedness regime, and using the boundary equation $\mathcal{E}(\beta',\beta,\nu_n,x(\nu_n)) = -\frac{1}{2}(a_L-1)(1-\gamma)$, we obtain after taking $n\to\infty$ that

$$\mathcal{E}(\beta',\beta,\nu_c,x_\star) = -\frac{1}{2}(a_L-1)(1-\gamma).$$

However, since $x_\star<\infty$, the strict inequality above gives

$$\mathcal{E}(\beta',\beta,\nu_c,x_\star) \;<\; \mathcal{E}_\infty(\nu_c) \;=\; -\frac{1}{2}(a_L-1)(1-\gamma),$$

a contradiction. Therefore $x(\nu)$ must be unbounded as $\nu\uparrow\nu_c$. Finally, since Lemma C.1 shows that $x(\nu)$ is monotonically increasing in ν, the only possibility is $x(\nu)$ goes to infinity as ν approaches ν_c as claimed.

We next start from the expansion obtained above:

$$\mathcal{E}_\infty(\nu)+\frac{1}{2}(a_L-1)(1-\gamma) = a_L\big(C_1(\nu)+C_2(\nu)\big)\, r\big(\nu, x(\nu)\big)+R(\nu), \qquad R(\nu) = O\Big(r\big(\nu, x(\nu)\big)^2\Big),$$

as $\nu \uparrow \nu_c$. Obviously $C_1(\nu) > 0$ and $C_2(\nu) > 0$, hence $a_L\big(C_1(\nu) + C_2(\nu)\big) > 0$. Then, by continuity, there exist $\varepsilon > 0$ and a constant $M', m > 0$ such that $a_L\big(C_1(\nu) + C_2(\nu)\big) \geq m > 0$ and $|R(\nu)| \leq M'\, r\big(\nu, x(\nu)\big)^2$ for all $\nu \in (\nu_c - \varepsilon, \nu_c)$. Since $x(\nu) \to \infty$ as $\nu \uparrow \nu_c$, we have $r\big(\nu, x(\nu)\big) \to 0$. Hence we may further assume that

$$M\, r\big(\nu, x(\nu)\big) \leq \frac{1}{2}\, m.$$

when $\nu \in (\nu_c - \varepsilon, \nu_c)$. Then,

$$\begin{aligned}
\mathcal{E}_\infty(\nu) + \frac{1}{2}(a_L - 1)(1 - \gamma) &\geq a_L\big(C_1(\nu) + C_2(\nu)\big)\, r\big(\nu, x(\nu)\big) - |R(\nu)| \\
&\geq m\, r\big(\nu, x(\nu)\big) - M\, r\big(\nu, x(\nu)\big)^2 \\
&\geq \frac{1}{2}\, m\, r\big(\nu, x(\nu)\big),
\end{aligned}$$

Consequently,

$$r\big(\nu, x(\nu)\big)^2 = O\left(\left(\mathcal{E}_\infty(\nu) + \frac{1}{2}(a_L - 1)(1 - \gamma)\right)^2\right).$$

Now,

$$r\big(\nu, x(\nu)\big) = \frac{\mathcal{E}_\infty(\nu) + \frac{1}{2}(a_L - 1)(1 - \gamma)}{a_L\big(C_1(\nu) + C_2(\nu)\big)} - \frac{R(\nu)}{a_L\big(C_1(\nu) + C_2(\nu)\big)}.$$

Therefore

$$\frac{R(\nu)}{a_L\big(C_1(\nu) + C_2(\nu)\big)} = O\Big(r\big(\nu, x(\nu)\big)^2\Big) = O\left(\left(\mathcal{E}_\infty(\nu) + \frac{1}{2}(a_L - 1)(1 - \gamma)\right)^2\right),$$

and

$$r\big(\nu, x(\nu)\big) = \frac{\mathcal{E}_\infty(\nu) + \frac{1}{2}(a_L - 1)(1 - \gamma)}{a_L\big(C_1(\nu) + C_2(\nu)\big)} + O\left(\left(\mathcal{E}_\infty(\nu) + \frac{1}{2}(a_L - 1)(1 - \gamma)\right)^2\right), \qquad \nu \uparrow \nu_c.$$

By differentiability of $\mathcal{E}_\infty$ at ν_c and the identity $\mathcal{E}_\infty(\nu_c) = -\frac{1}{2}(a_L - 1)(1 - \gamma)$, we have the first-order expansion

$$\mathcal{E}_\infty(\nu)+\frac{1}{2}(a_L-1)(1-\gamma) = \mathcal{E}'_\infty(\nu_c)\,(\nu-\nu_c)+o(\nu-\nu_c) = \big(-\mathcal{E}'_\infty(\nu_c)\big)\,(\nu_c-\nu)+o(\nu_c-\nu), \qquad \nu \uparrow \nu_c,$$

where $-\mathcal{E}'_\infty(\nu_c) > 0$. Hence, we obtain

$$r\big(\nu, x(\nu)\big) = \frac{-\mathcal{E}'_\infty(\nu_c)}{a_L\big(C_1(\nu) + C_2(\nu)\big)}\,(\nu_c-\nu) + O\big((\nu_c-\nu)^2\big) = \frac{-\mathcal{E}'_\infty(\nu_c)(\nu_c - \nu)}{a_L\big(C_1(\nu) + C_2(\nu)\big)}\Big(1+O(\nu_c-\nu)\Big).$$

Now invert the definition

$$r\big(\nu, x(\nu)\big) = \frac{c_\delta \nu}{c\sqrt{a_0 x(\nu) - c_{\delta'}\nu}}$$

to get the identity

$$a_0 x(\nu) - c_{\delta'}\nu = \left(\frac{c_\delta \nu}{c\, r\big(\nu, x(\nu)\big)}\right)^2.$$

Since

$$\left(\frac{c_\delta \nu}{c\, r\big(\nu, x(\nu)\big)}\right)^2 = \left(\frac{c_\delta \nu\, a_L\big(C_1(\nu) + C_2(\nu)\big)}{c\,\big(-\mathcal{E}'_\infty(\nu_c)\big)}\right)^2 \frac{1}{(\nu_c - \nu)^2}\Big(1 + O(\nu_c - \nu)\Big), \qquad \nu \uparrow \nu_c,$$

we have,

$$x(\nu) = \frac{c_{\delta'}\nu}{a_0} + \frac{1}{a_0}\left(\frac{c_\delta \nu\, a_L\big(C_1(\nu)+C_2(\nu)\big)}{c\left(-\mathcal{E}_\infty'(\nu_c)\right)}\right)^2 \frac{1}{(\nu_c-\nu)^2}\Big(1+O(\nu_c-\nu)\Big).$$

Finally, since $\nu = \nu_c + O(\nu_c-\nu)$ and $C_1(\nu)+C_2(\nu) = C_1(\nu_c)+C_2(\nu_c)+O(\nu_c-\nu)$ by continuity, we may replace ν by ν_c and $C_1(\nu)+C_2(\nu)$ by $C_1(\nu_c)+C_2(\nu_c)$ inside the prefactor at the cost of a multiplicative $(1+O(\nu_c-\nu))$ factor, giving

$$x(\nu) = \frac{1}{a_0}\left(\frac{c_\delta \nu_c\, a_L\big(C_1(\nu_c)+C_2(\nu_c)\big)}{c\left(-\mathcal{E}_\infty'(\nu_c)\right)}\right)^2 \frac{1}{(\nu_c-\nu)^2}\Big(1+O(\nu_c-\nu)\Big).$$

Differentiating the asymptotic expansion gives

$$x'(\nu) = \frac{C(\nu_c)}{(\nu_c-\nu)^3}\Big(1+O(\nu_c-\nu)\Big),$$

where

$$C(\nu_c) = \frac{2}{a_0}\left(\frac{c_\delta \nu_c\, a_L\big(C_1(\nu_c)+C_2(\nu_c)\big)}{c\left(-\mathcal{E}_\infty'(\nu_c)\right)}\right)^2.$$

This completes the proof. □

C.4 Proof of Corollary 5.6

Proof. By Proposition 5.5, for fixed (β',β) we have $\mathcal{I}_\mathcal{N}(\beta',\beta,\nu) = \big(x(\beta',\beta,\nu),\,1-\gamma\big)$ and $x(\beta',\beta,\nu)$ is monotonically increasing in ν. Therefore, for any $0 < \nu < \nu^\star(\beta',\beta)$ we have $x(\beta',\beta,\nu) < V_{p_0}(\hat\theta_0)$, so $V_{p_0}(\hat\theta_0) \in \mathcal{I}_\mathcal{N}(\beta',\beta,\nu)$, i.e., $\mathcal{N}\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big) < 0$. Conversely, for any $\nu \ge \nu^\star(\beta',\beta)$ we have $x(\beta',\beta,\nu) \ge V_{p_0}(\hat\theta_0)$, so $V_{p_0}(\hat\theta_0) \notin \mathcal{I}_\mathcal{N}(\beta',\beta,\nu)$, i.e., $\mathcal{N}\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big) \ge 0$. Hence,

$$\nu^\star(\beta',\beta) \;=\; \sup\Big\{\,\nu>0:\ \mathcal{N}\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big)<0\,\Big\}.$$

The remaining claims follow by directly combining Lemma C.4, Lemma C.5, Lemma C.6, and Lemma C.7. □

Lemma C.4. *In Proposition 5.5, fix any initialization $V_{p_0}(\hat\theta_0) \in (0,1-\gamma)$, and define*

$$\nu^\star(\beta',\beta) \;=\; \sup\Big\{\,\nu>0:\ \mathcal{N}\big(\beta',\beta,\nu,V_{p_0}(\hat\theta_0)\big)<0\,\Big\}.$$

Then, fixing β', for $\beta>\beta'$, $\nu^\star(\beta',\beta)$ is decreasing in β.

Proof. Let $x_0 := V_{p_0}(\hat\theta_0)$ for convenience. By Theorem 5.2, the inequality $\mathcal{N}(\beta',\beta,\nu,x_0)<0$ is equivalent to

$$\mathcal{E}(\beta',\beta,\nu,x_0) > -\frac{1}{2}\,(a_L-1)(1-\gamma).$$

Note that when $\beta'>0$ is fixed, the right-hand side above is a fixed constant. Moreover, Lemma C.1 has already shown that $\mathcal{E}(\beta',\beta,\nu,x_0)$ is monotonically (strictly) decreasing in ν. Therefore, to prove that $\nu^\star(\beta',\beta)$ is monotonically decreasing in β, it suffices to show that for each fixed (β',ν,x_0) in the well-definedness regime, $\mathcal{E}(\beta',\beta,\nu,x_0)$ is strictly decreasing as a function of β (with $\beta>\beta'$). Indeed, define

$$\Phi(\beta',\beta,\nu) \;:=\; \mathcal{E}(\beta',\beta,\nu,x_0) + \frac{1}{2}\,(a_L-1)(1-\gamma).$$

Then $\Phi\big(\beta',\beta,\nu^\star(\beta',\beta)\big) = 0$. If $\mathcal{E}(\beta',\beta,\nu,x_0)$ is strictly decreasing in β, equivalently $\partial_\beta\Phi(\beta',\beta,\nu)<0$, since $\partial_\nu\Phi(\beta',\beta,\nu)<0$ and $\partial_\nu\Phi\big(\beta',\beta,\nu^\star(\beta',\beta)\big)\neq 0$, the implicit function theorem applies to the equation $\Phi(\beta',\beta,\nu)=0$ around $\nu=\nu^\star(\beta',\beta)$ and yields

$$\frac{\partial}{\partial\beta}\nu^\star(\beta',\beta) \;=\; -\frac{\partial_\beta\Phi\big(\beta',\beta,\nu^\star(\beta',\beta)\big)}{\partial_\nu\Phi\big(\beta',\beta,\nu^\star(\beta',\beta)\big)} \;<\; 0.$$

Therefore, it suffices to prove that

$$\frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}} \cdot \frac{1}{1-\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}}$$

and

$$\left(\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}$$

are both strictly increasing as a function of β. For

$$\widetilde{U}(\beta) = \frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}}.$$

Clearly, as β increases, $2^{-\beta}$ decreases, and hence $\widetilde{U}(\beta)$ is strictly increasing in β. Next consider

$$\widetilde{V}(\beta) = \frac{1}{1-\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}}.$$

Differentiating gives

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\beta}\log\left(\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}\right) \\
&= -\frac{3}{2}\cdot\frac{-\log 2\cdot 2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)}{2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu} - \frac{1}{L} \\
&= \frac{3}{2}\log 2\cdot\frac{2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)}{2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu} - \frac{1}{L} \\
&> \frac{3}{2}\log 2 - \frac{1}{L}
\end{aligned}$$

Since $L \geq 2$, we have $\frac{3}{2}\log 2 - \frac{1}{L} > 0$, and hence the above derivative is strictly positive. Thus

$$\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}$$

is strictly increasing in β, which implies that $\widetilde{V}(\beta)$ is strictly increasing in β. Combining with the monotonicity of $\widetilde{U}(\beta)$, we conclude that $\widetilde{U}(\beta)\widetilde{V}(\beta)$ is strictly increasing in β.

Next, consider

$$T_2(\beta) = \left(\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}.$$

We compute the derivative of $\log T_2(\beta)$. Since the factor $\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}$ does not depend on β, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\beta}\log T_2(\beta) &= (L-1)\cdot\frac{\mathrm{d}}{\mathrm{d}\beta}\log\left(\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu\right)^{3/2}}\right) - \log L \\
&= (L-1)\cdot\frac{3}{2}\log 2\cdot\frac{2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)}{2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0x_0-c_{\delta'}\nu}}\right)-c_{\delta'}\nu} - \log L
\end{aligned}$$

$$> (L-1)\cdot\frac{3}{2}\log 2 - \log L.$$

In particular, since $L \geq 2$, we have $(L-1)\cdot\frac{3}{2}\log 2 - \log L > 0$, and therefore $T_2(\beta)$ is strictly increasing in β. Consequently, the β-dependent terms inside the bracket in $\mathcal{E}(\beta', \beta, \nu, x_0)$ are strictly increasing in β, and since the bracket is multiplied by the negative coefficient $-a_L$, it follows that $\mathcal{E}(\beta', \beta, \nu, x_0)$ is strictly decreasing in β. This completes the proof. □

Lemma C.5. *In Proposition 5.5, fix any initialization $V_{p_0}(\hat{\theta}_0) \in (0, 1-\gamma)$, and define*

$$\nu^\star(\beta', \beta) \;=\; \sup\Big\{\,\nu > 0 : \,\mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0\,\Big\}.$$

Then, fixing β, for $\beta' \in (0, \beta)$, $\nu^\star(\beta', \beta)$ is increasing in β'.

Proof. Let $x_0 := V_{p_0}(\hat{\theta}_0)$ for convenience. Define

$$T_1(\nu) := \frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}}\cdot\frac{1-\left(\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1-\frac{c_\delta\nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}},$$

$$T_2(\nu, \beta') := \left(\frac{c_\delta\nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0(\beta')x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}}\right)^{L-1} L^{-\beta}\cdot\frac{c_\delta\nu}{c\sqrt{a_0(\beta')x_0-c_{\delta'}\nu}},$$

and

$$T_3(\nu, \beta') := \frac{c_\delta\nu}{c\sqrt{2^{-\beta}(1-\gamma)-c_{\delta'}\nu}}\cdot\frac{1}{1-\frac{c_\delta\nu}{2c\Big(2^{-\beta}\Big(1-\gamma-\frac{c_\delta\nu}{c\sqrt{a_0(\beta')x_0-c_{\delta'}\nu}}\Big)-c_{\delta'}\nu\Big)^{3/2}}\cdot e^{-\beta/L}}.$$

Then $\mathcal{N}(\beta', \beta, \nu, x_0) < 0$ is equivalent to

$$T_1(\nu) - a_L(\beta')\big(T_2(\nu,\beta') + T_3(\nu,\beta')\big) > -\frac{1}{2}\,(a_L(\beta')-1)(1-\gamma).$$

Equivalently,

$$T_1(\nu) - \frac{1-\gamma}{2} + a_L(\beta')\left(\frac{1-\gamma}{2} - T_2(\nu,\beta') - T_3(\nu,\beta')\right) > 0.$$

We first verify that both $a_0(\beta')$ and $a_L(\beta')$ are strictly increasing in β'. Recall the definitions

$$a_0(\beta') \;=\; \frac{L}{\sum_{i=1}^L i^{-\beta'}}, \qquad a_L(\beta') \;=\; \frac{\sum_{i=1}^L i^{-\beta'}}{L^{1-\beta'}}.$$

$a_0(\beta')$ is strictly increasing in β' since

$$\frac{\mathrm{d}}{\mathrm{d}\beta'}\left(\sum_{i=1}^L i^{-\beta'}\right) < 0,$$

Next,

$$a_L(\beta') = \frac{1}{L}\sum_{i=1}^L\left(\frac{L}{i}\right)^{\beta'}.$$

Therefore $a_L(\beta')$ is strictly increasing in β'. Then, we conclude that both $T_2(\nu, \beta')$ and $T_3(\nu, \beta')$ are strictly decreasing in β'.

Suppose that for every $\beta' \in (0, \beta)$, we have

$$\frac{1-\gamma}{2} - T_2\big(\nu^\star(\beta',\beta),\beta'\big) - T_3\big(\nu^\star(\beta',\beta),\beta'\big) \geq 0.$$

Then for any $0 < \beta_1' < \beta_2' < \beta$, using that $a_L(\beta')$ is strictly increasing in β' and that $T_2(\nu, \beta')$ and $T_3(\nu, \beta')$ are strictly decreasing in β', we obtain

$$T_1\big(\nu^\star(\beta_1', \beta)\big) - \frac{1-\gamma}{2} + a_L(\beta_2') \left(\frac{1-\gamma}{2} - T_2\big((\nu^\star(\beta_1', \beta), \beta_2'\big) - T_3\big((\nu^\star(\beta_1', \beta), \beta_2'\big) \right)$$
$$> T_1\big(\nu^\star(\beta_1', \beta)\big) - \frac{1-\gamma}{2} + a_L(\beta_1') \left(\frac{1-\gamma}{2} - T_2\big((\nu^\star(\beta_1', \beta), \beta_1'\big) - T_3\big((\nu^\star(\beta_1', \beta), \beta_1'\big) \right) = 0.$$

Since Lemma C.1 shows that $\mathcal{E}(\beta', \beta, \nu, x_0)$ is strictly decreasing in ν, it follows that $\nu^\star(\beta_2', \beta) > \nu^\star(\beta_1', \beta)$. Therefore, it remains to prove that for every $\beta' \in (0, \beta)$,

$$\frac{1-\gamma}{2} - T_2\big(\nu^\star(\beta', \beta), \beta'\big) - T_3\big(\nu^\star(\beta', \beta), \beta'\big) \geq 0.$$

Note that

$$T_1\big(\nu^\star(\beta', \beta)\big) - \frac{1-\gamma}{2} + a_L(\beta') \left(\frac{1-\gamma}{2} - T_2\big(\nu^\star(\beta', \beta), \beta'\big) - T_3\big(\nu^\star(\beta', \beta), \beta'\big) \right) = 0.$$

Since $a_L(\beta') > 0$, it suffices to prove that $T_1\big(\nu^\star(\beta', \beta)\big) - \frac{1-\gamma}{2} < 0$.

We next prove that $T_1(\nu)$ is strictly increasing in ν within the well-definedness regime. This is because $\frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}}$ is positive and clearly strictly increasing in ν. Moreover,

$$\frac{1 - \left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}} \right)^{L-1}}{1 - \frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}} = \sum_{j=0}^{L-2} \left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}} \right)^j,$$

where $\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}$ is positive and clearly strictly increasing in ν. Therefore, let ν_T denote a (necessarily unique, by strict monotonicity) value satisfying $T_1(\nu_T) - \frac{1-\gamma}{2} = 0$. Then, since $T_1(\nu)$ is increasing in ν, it suffices to prove that $0 < \nu^\star(\beta', \beta) < \nu_T$.

By the proof of Lemma C.1, we have $T_3(\nu, \beta') > T_1(\nu)$. Hence $T_3(\nu, \beta') + T_2(\nu, \beta') > T_1(\nu)$, and therefore,

$$\begin{aligned} &T_1\big(\nu_T\big) - \frac{1-\gamma}{2} + a_L(\beta') \left(\frac{1-\gamma}{2} - T_2\big(\nu_T, \beta'\big) - T_3\big(\nu_T, \beta'\big) \right) \\ &= a_L(\beta') \left(\frac{1-\gamma}{2} - T_2\big(\nu_T, \beta'\big) - T_3\big(\nu_T, \beta'\big) \right) \\ &< a_L(\beta') \left(\frac{1-\gamma}{2} - T_1(\nu_T) \right) = 0. \end{aligned}$$

On the other hand, since $T_1(0) = T_2(0, \beta') = T_3(0, \beta') = 0$, we have

$$T_1(0) - \frac{1-\gamma}{2} + a_L(\beta') \left(\frac{1-\gamma}{2} - T_2(0, \beta') - T_3(0, \beta') \right) = \big(a_L(\beta') - 1\big) \frac{1-\gamma}{2} > 0.$$

Finally, since

$$T_1(\nu) - \frac{1-\gamma}{2} + a_L(\beta') \left(\frac{1-\gamma}{2} - T_2(\nu, \beta') - T_3(\nu, \beta') \right)$$

is strictly decreasing in ν as well (for each fixed β'), we must have $0 < \nu^\star(\beta', \beta) < \nu_T$, as desired. □

Lemma C.6. *In Proposition 5.5, fix any initialization $V_{p_0}(\hat{\theta}_0) \in (0, 1-\gamma)$, and define*

$$\nu^\star(\beta', \beta) \;=\; \sup \Big\{ \nu > 0 : \mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0 \Big\}.$$

Fix $\Delta = \beta - \beta' > 0$ and write $\nu^\star(\beta', \beta) = \nu^\star(\beta', \beta' + \Delta)$. When β' is small, $\nu^\star(\beta', \beta' + \Delta)$ is increasing in β' and

$$\nu^\star(\beta', \beta' + \Delta) = \frac{c(1-\gamma)^{3/2} \log\Big(\frac{L}{(L!)^{1/L}} \Big)}{2c_\delta \big(2^{\Delta/2} - 1\big)} \,\beta' \;+\; o(\beta') \qquad \textit{as } \beta' \to 0.$$

Proof. Let $x_0 := V_{p_0}(\hat{\theta}_0)$ and $\nu^\star_\Delta(\beta') := \nu^\star(\beta', \beta' + \Delta)$ convenience. Following the notation of Theorem 5.2, we write

$$\mathcal{N}(\beta', \beta' + \Delta, \nu, x_0) = -\mathcal{E}(\beta', \beta' + \Delta, \nu, x_0) - \frac{1}{2}\big(a_L(\beta') - 1\big)(1-\gamma).$$

Moreover, $\mathcal{N}(\beta', \beta' + \Delta, \nu, x_0)$ is increasing in ν, and for every $\beta' > 0$ we have $\mathcal{N}(\beta', \beta' + \Delta, 0, x_0) < 0$. Furthermore, $\nu^\star_\Delta(\beta')$ satisfies

$$\mathcal{N}\big(\beta', \beta' + \Delta, \nu^\star_\Delta(\beta'), x_0\big) = 0,$$

and note that when $\beta' = 0$, the corresponding threshold satisfies $\nu^\star_\Delta(0) = 0$.

We first compute the partial derivatives of $\mathcal{N}(\beta', \beta' + \Delta, \nu, x_0)$ with respect to ν and β' at $(\beta', \nu) = (0, 0)$, namely $\partial_\nu \mathcal{N}(0, \Delta, 0, x_0)$ and $\partial_{\beta'} \mathcal{N}(0, \Delta, 0, x_0)$. For $\partial_\nu \mathcal{N}(0, \Delta, 0, x_0) = -\partial_\nu \mathcal{E}(0, \Delta, 0, x_0)$, note that the derivative of $\mathcal{E}(\beta', \beta' + \Delta, \nu, x_0)$ at $\nu = 0$ only depends on its linear term in ν. Since

$$\frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} \cdot \frac{1 - \left(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\right)^{L-1}}{1 - \frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}} = \frac{c_\delta}{c\sqrt{1-\gamma}}\,\nu + o(\nu),$$

$$\frac{c_\delta \nu}{c\sqrt{2^{-\beta}(1-\gamma) - c_{\delta'}\nu}} \cdot \frac{1}{1 - \frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}} \cdot e^{-\beta/L}} = 2^{\Delta/2}\frac{c_\delta}{c\sqrt{1-\gamma}}\,\nu + o(\nu),$$

and

$$\left(\frac{c_\delta \nu}{2c\left(2^{-\beta}\left(1-\gamma-\frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}}\right) - c_{\delta'}\nu\right)^{3/2}}\right)^{L-1} L^{-\beta} \cdot \frac{c_\delta \nu}{c\sqrt{a_0 x_0 - c_{\delta'}\nu}} = O(\nu^L),$$

we obtain

$$\partial_\nu \mathcal{N}(0, \Delta, 0, x_0) = -\partial_\nu \mathcal{E}(0, \Delta, 0, x_0) = \left(2^{\Delta/2} - 1\right)\frac{c_\delta}{c\sqrt{1-\gamma}} > 0.$$

Next we compute $\partial_{\beta'} \mathcal{N}(0, \Delta, 0, x_0)$. Since $\mathcal{E}(\beta', \beta' + \Delta, \nu, x_0) \equiv 0$ at $\nu = 0$, we have $\partial_{\beta'} \mathcal{N}(0, \Delta, 0, x_0) = -\frac{1}{2}\, a_L'(0)(1-\gamma)$. Recall that $a_L(\beta') = \frac{\sum_{i=1}^L i^{-\beta'}}{L^{1-\beta'}}$. Hence

$$\log a_L(\beta') = \log\left(\sum_{i=1}^L i^{-\beta'}\right) - \log L + \beta' \log L,$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\beta'} \log a_L(\beta')\bigg|_{\beta'=0} = \frac{-\sum_{i=1}^L \log i}{L} + \log L = \log\Big(\frac{L}{(L!)^{1/L}}\Big).$$

Since $a_L(0) = \frac{\sum_{i=1}^L 1}{L} = 1$, it follows that

$$a_L'(0) = a_L(0) \cdot \frac{\mathrm{d}}{\mathrm{d}\beta'} \log a_L(\beta')\bigg|_{\beta'=0} = \log\Big(\frac{L}{(L!)^{1/L}}\Big) > 0.$$

Therefore,

$$\partial_{\beta'} \mathcal{N}(0, \Delta, 0, x_0) = -\frac{1}{2}\, a_L'(0)(1-\gamma) = -\frac{1}{2}\,\log\Big(\frac{L}{(L!)^{1/L}}\Big)\,(1-\gamma) < 0.$$

Finally, since $\nu^\star_\Delta(\beta')$ is defined implicitly by $\mathcal{N}\big(\beta', \beta' + \Delta, \nu^\star_\Delta(\beta'), x_0\big) = 0$, the implicit function theorem yields

$$\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu^\star_\Delta(0) = -\frac{\partial_{\beta'} \mathcal{N}(0, \Delta, 0, x_0)}{\partial_\nu \mathcal{N}(0, \Delta, 0, x_0)} = \frac{c(1-\gamma)^{3/2}\log\Big(\frac{L}{(L!)^{1/L}}\Big)}{2c_\delta\big(2^{\Delta/2} - 1\big)} \; > \; 0.$$

Therefore, as $\beta' \to 0$,

$$\nu^\star_\Delta(\beta') = \nu^\star_\Delta(0) + \frac{\mathrm{d}}{\mathrm{d}\beta'}\nu^\star_\Delta(0)\,\beta' + o(\beta') = \frac{c(1-\gamma)^{3/2}\log\Big(\frac{L}{(L!)^{1/L}}\Big)}{2c_\delta\big(2^{\Delta/2} - 1\big)}\,\beta' \; + \; o(\beta').$$

This completes the proof. $\square$

Lemma C.7. *In Proposition 5.5, fix any initialization $V_{p_0}(\hat{\theta}_0) \in (0, 1-\gamma)$, and define*

$$\nu^\star(\beta', \beta) \;=\; \sup\Big\{\, \nu > 0 : \,\mathcal{N}\big(\beta', \beta, \nu, V_{p_0}(\hat{\theta}_0)\big) < 0 \,\Big\}.$$

Fix $\Delta = \beta - \beta' > 0$ and write $\nu^\star(\beta', \beta) = \nu^\star(\beta', \beta' + \Delta)$. Let $\nu_T > 0$ be the solution to

$$\frac{c_\delta \nu}{c\sqrt{1-\gamma-c_{\delta'}\nu}} \cdot \frac{1 - \Big(\frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}\Big)^{L-1}}{1 - \frac{c_\delta \nu}{2c\,(1-\gamma-c_{\delta'}\nu)^{3/2}}} = \frac{1-\gamma}{2}.$$

Then there exists a constant $\nu_0 > 0$ such that whenever $\nu_T < \nu_0$, we have $\nu^\star(\beta', \beta' + \Delta) < \nu_T$, and $\nu^\star(\beta', \beta' + \Delta)$ is first increasing and then decreasing in β', with a unique maximizer. Moreover, for sufficiently large β', the tail scaling satisfies

$$\nu^\star(\beta', \beta' + \Delta) = \Theta\big(2^{-\beta'/2}\big).$$

Proof. Let $x_0 := V_{p_0}(\hat{\theta}_0)$ and $\nu^\star_\Delta(\beta') := \nu^\star(\beta', \beta' + \Delta)$ convenience. First, by the proof of Lemma C.5, we directly obtain $0 < \nu^\star_\Delta(\beta') < \nu_T < \nu_0$.

Next, in $\mathcal{N}(\beta', \beta' + \Delta, \nu, x_0)$ we focus on the roles of β' and ν, and denote it by $\mathcal{N}(\beta', \nu)$. Then, by the proof of Lemma C.6, when ν_T is sufficiently small, we have the expansion

$$\mathcal{N}(\beta', \nu) = -\frac{1}{2}\,\big(a_L(\beta') - 1\big)(1-\gamma) + \frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2} - 1\Big)\,\nu + R(\beta', \nu),$$

where there exist constants $C_0, C_1, C_2 > 0$ such that for any $\nu \le \nu_T$,

$$|R(\beta', \nu)| \le C_0 \nu^2, \qquad \big|\partial_\nu R(\beta', \nu)\big| \le C_1 \nu, \qquad \big|\partial_{\beta'} R(\beta', \nu)\big| \le C_2 \nu^2.$$

Moreover, note that for fixed $\Delta > 0$, we always have

$$\frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2} - 1\Big) > 0.$$

Define the linear approximation of $\mathcal{N}(\beta', \nu)$ by dropping the remainder term:

$$\mathcal{N}_\ell(\beta', \nu) = -\frac{1}{2}\,\big(a_L(\beta') - 1\big)(1-\gamma) + \frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2} - 1\Big)\,\nu.$$

Then the linearized root is given by

$$\nu^\ell_\Delta(\beta') = \frac{c(1-\gamma)^{3/2}\big(a_L(\beta') - 1\big)}{2c_\delta\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2} - 1\Big)}.$$

Note that the behavior of $\nu^\ell_\Delta(\beta')$ as $\beta' \to 0$ is consistent with Lemma C.6. We now verify this. Recall that $a_L(\beta') = \frac{\sum_{i=1}^{L} i^{-\beta'}}{L^{1-\beta'}}$. With explicit expansions we have

$$a_L(\beta') = 1 + \log\Big(\frac{L}{(L!)^{1/L}}\Big)\,\beta' + o(\beta').$$

Next, since

$$2^{(\beta'+\Delta)/2} = 2^{\Delta/2}\Big(1 + \frac{\log 2}{2}\beta' + o(\beta')\Big),$$

we obtain

$$a_L(\beta')\,2^{(\beta'+\Delta)/2} - 1 = \big(2^{\Delta/2} - 1\big) + 2^{\Delta/2}\Big(\log\Big(\frac{L}{(L!)^{1/L}}\Big) + \frac{\log 2}{2}\Big)\beta' + o(\beta').$$

Therefore, as $\beta' \to 0$,

$$\nu^\ell_\Delta(\beta') = \frac{c(1-\gamma)^{3/2}\log\Big(\frac{L}{(L!)^{1/L}}\Big)}{2c_\delta\big(2^{\Delta/2} - 1\big)}\,\beta' \;+\; o(\beta'),$$

which is consistent with Lemma C.6. Therefore, when β' is sufficiently small, the above expansion implies that $\nu_\Delta^\ell(\beta')$ is strictly increasing in β'. On the other hand, when β' is sufficiently large, we have $a_L(\beta') \to \infty$, and hence

$$\nu_\Delta^\ell(\beta') = \frac{c(1-\gamma)^{3/2}\big(a_L(\beta')-1\big)}{2c_\delta\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)} \;\sim\; \frac{c(1-\gamma)^{3/2}\,a_L(\beta')}{2c_\delta\,a_L(\beta')\,2^{(\beta'+\Delta)/2}} = \frac{c(1-\gamma)^{3/2}}{2c_\delta}\,2^{-(\beta'+\Delta)/2} \;\to\; 0.$$

Consequently, $\nu_\Delta^\ell(\beta')$ must start decreasing at some point in between, and therefore it admits at least one maximizer (i.e., at least one local maximum) over $\beta' > 0$.

Next we prove that there exists a unique critical point. This is equivalent to showing that the equation

$$\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu_\Delta^\ell(\beta') = 0$$

admits a unique solution on $(0,\infty)$. By directly differentiating $\nu_\Delta^\ell(\beta')$, one obtains that $\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu_\Delta^\ell(\beta') = 0$ is equivalent to

$$\frac{a_L(\beta')\big(a_L(\beta')-1\big)}{a_L'(\beta')} = \frac{2^{(\beta'+\Delta)/2}-1}{\frac{\mathrm{d}}{\mathrm{d}\beta'}2^{(\beta'+\Delta)/2}}.$$

Let the right-hand side be denoted by $g(\beta')$. Then

$$g(\beta') = \frac{2^{(\beta'+\Delta)/2}-1}{\frac{\mathrm{d}}{\mathrm{d}\beta'}2^{(\beta'+\Delta)/2}} = \frac{2}{\log 2}\left(1-2^{-(\beta'+\Delta)/2}\right),$$

which is strictly increasing in β', and satisfies

$$g(0) = \frac{2}{\log 2}\left(1-2^{-\Delta/2}\right) > 0, \qquad \lim_{\beta'\to\infty} g(\beta') = \frac{2}{\log 2}.$$

Let the left-hand side be denoted by $h(\beta')$, namely

$$h(\beta') := \frac{a_L(\beta')\big(a_L(\beta')-1\big)}{a_L'(\beta')}.$$

Lemma C.8 shows that $h(\beta')$ is also strictly increasing in β'. Moreover,

$$\lim_{\beta'\to 0} h(\beta') = 0, \qquad \lim_{\beta'\to\infty} h(\beta') = +\infty.$$

Therefore, there exists a unique $\bar\beta_c' \in (0,\infty)$ such that $h(\bar\beta_c') = g(\bar\beta_c')$, which is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu_\Delta^\ell(\bar\beta_c') = 0.$$

Therefore, $\nu_\Delta^\ell(\beta')$ is strictly increasing on $(0,\bar\beta_c')$ and is strictly decreasing on $(\bar\beta_c',\infty)$. It remains to show that $\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu_\Delta^\ell(\beta')$ and $\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu_\Delta^\star(\beta')$ do not deviate too much, so that the true root $\nu_\Delta^\star(\beta')$ inherits the unimodality and the uniqueness of the maximizer.

On the one hand, since

$$\partial_\nu \mathcal{N}(\beta',\nu) = \frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big) + \partial_\nu R(\beta',\nu),$$

$\big|\partial_\nu R(\beta',\nu)\big| \le C_1\nu \le C_1\nu_T$, and $\frac{c_\delta}{c\sqrt{1-\gamma}}\big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\big)$ admits a strictly positive lower bound, it follows that

$$\partial_\nu \mathcal{N}(\beta',\nu) = \frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)\big(1+O(\nu_T)\big).$$

On the other hand, $\partial_{\beta'}\mathcal{N}(\beta',\nu) = \partial_{\beta'}\mathcal{N}_\ell(\beta',\nu) + \partial_{\beta'}R(\beta',\nu)$, so plugging in $\nu = \nu_\Delta^\star(\beta')$ gives

$$\partial_{\beta'}\mathcal{N}\big(\beta',\nu_\Delta^\star(\beta')\big) = \partial_{\beta'}\mathcal{N}_\ell\big(\beta',\nu_\Delta^\star(\beta')\big) + \partial_{\beta'}R\big(\beta',\nu_\Delta^\star(\beta')\big).$$

Moreover, since

$$\partial_{\beta'}\mathcal{N}_\ell(\beta',\nu) = -\frac{1}{2}\,a_L'(\beta')(1-\gamma) + \frac{\mathrm{d}}{\mathrm{d}\beta'}\left[\frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)\right]\nu,$$

we have

$$\partial_{\beta'}\mathcal{N}_\ell\big(\beta',\nu^\star_\Delta(\beta')\big) = \partial_{\beta'}\mathcal{N}_\ell\big(\beta',\nu^\ell_\Delta(\beta')\big)+\frac{\mathrm{d}}{\mathrm{d}\beta'}\left[\frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)\right]\Big(\nu^\star_\Delta(\beta')-\nu^\ell_\Delta(\beta')\Big).$$

Hence

$$\partial_{\beta'}\mathcal{N}\big(\beta',\nu^\star_\Delta(\beta')\big) = \partial_{\beta'}\mathcal{N}_\ell\big(\beta',\nu^\ell_\Delta(\beta')\big)+\frac{\mathrm{d}}{\mathrm{d}\beta'}\left[\frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)\right]\Big(\nu^\star_\Delta(\beta')-\nu^\ell_\Delta(\beta')\Big)+\partial_{\beta'}R\big(\beta',\nu^\star_\Delta(\beta')\big).$$

For the remainder derivative, we use the bound $\big|\partial_{\beta'}R(\beta',\nu)\big| \le C_2\nu^2 \le C_2\nu_T^2$. Next, to control $\nu^\star_\Delta(\beta')-\nu^\ell_\Delta(\beta')$, note that $\nu^\star_\Delta(\beta')$ satisfies

$$-\frac{1}{2}\,\big(a_L(\beta')-1\big)(1-\gamma) + \frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)\,\nu^\star_\Delta(\beta') + R\big(\beta',\nu^\star_\Delta(\beta')\big) = 0,$$

while $\nu^\ell_\Delta(\beta')$ satisfies

$$-\frac{1}{2}\,\big(a_L(\beta')-1\big)(1-\gamma) + \frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)\,\nu^\ell_\Delta(\beta') = 0.$$

Subtracting the two equations yields

$$\nu^\star_\Delta(\beta')-\nu^\ell_\Delta(\beta') = -\frac{R\big(\beta',\nu^\star_\Delta(\beta')\big)}{\frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)}.$$

Since $|R(\beta',\nu)| \le C_0\nu^2 \le C_0\nu_T^2$, and the denominator admits a strictly positive lower bound, we conclude that $\nu^\star_\Delta(\beta')-\nu^\ell_\Delta(\beta') = O(\nu_T^2)$. Therefore,

$$\partial_{\beta'}\mathcal{N}\big(\beta',\nu^\star_\Delta(\beta')\big) = \partial_{\beta'}\mathcal{N}_\ell\big(\beta',\nu^\ell_\Delta(\beta')\big) + O(\nu_T^2).$$

Then, by the implicit function theorem,

$$\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu^\star_\Delta(\beta') = -\frac{\partial_{\beta'}\mathcal{N}\big(\beta',\nu^\star_\Delta(\beta')\big)}{\partial_\nu\mathcal{N}\big(\beta',\nu^\star_\Delta(\beta')\big)} = -\frac{\partial_{\beta'}\mathcal{N}_\ell\big(\beta',\nu^\ell_\Delta(\beta')\big)+O(\nu_T^2)}{\frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)\big(1+O(\nu_T)\big)}.$$

Equivalently,

$$\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu^\star_\Delta(\beta') = -\frac{\partial_{\beta'}\mathcal{N}_\ell\big(\beta',\nu^\ell_\Delta(\beta')\big)}{\frac{c_\delta}{c\sqrt{1-\gamma}}\Big(a_L(\beta')\,2^{(\beta'+\Delta)/2}-1\Big)} + O(\nu_T) = \frac{\mathrm{d}}{\mathrm{d}\beta'}\nu^\ell_\Delta(\beta') + O(\nu_T).$$

Hence, as long as $\nu_0 > 0$ (and thus $\nu_T < \nu_0$) is chosen sufficiently small, the term $O(\nu_T)$ is small enough so that $\frac{\mathrm{d}}{\mathrm{d}\beta'}\nu^\star_\Delta(\beta')$ still has a unique zero β'_c, and moreover it is positive on $(0,\beta'_c)$ and negative on (β'_c,∞).

Finally, as shown earlier, when β' is sufficiently large,

$$\nu^\ell_\Delta(\beta') \;\sim\; \frac{c(1-\gamma)^{3/2}}{2c_\delta}\,2^{-(\beta'+\Delta)/2},$$

and hence the tail scaling satisfies $\nu^\ell_\Delta(\beta') = \Theta\big(2^{-\beta'/2}\big)$. Moreover, we have established that $\nu^\star_\Delta(\beta')-\nu^\ell_\Delta(\beta') = O\big(\nu^\star_\Delta(\beta')^2\big)$. For $\nu^\star_\Delta(\beta')$, note that by the standing well-definedness requirement of $\mathcal{N}(\beta',\nu)$, every radicand in the denominators must remain strictly positive. In particular, since $\beta = \beta'+\Delta$, we must have $2^{-(\beta'+\Delta)}(1-\gamma) - c_{\delta'}\,\nu^\star_\Delta(\beta') > 0$, which implies the upper bound $\nu^\star_\Delta(\beta') = O\big(2^{-\beta'}\big)$ as $\beta'\to\infty$. Combining this with $\nu^\star_\Delta(\beta')-\nu^\ell_\Delta(\beta') = O\big(\nu^\star_\Delta(\beta')^2\big)$, we obtain

$$\nu^\star_\Delta(\beta') = \Theta\big(2^{-\beta'/2}\big), \qquad \beta'\to\infty.$$

□

Lemma C.8. *Let $L \geq 2$ be an integer and define $a_L(\beta') \coloneqq \frac{1}{L}\sum_{i=1}^{L}\left(\frac{L}{i}\right)^{\beta'}$. Then the mapping*

$$\beta' \;\longmapsto\; \frac{a_L(\beta')\big(a_L(\beta')-1\big)}{a_L'(\beta')}, \qquad \beta' > 0,$$

is strictly increasing on $(0,\infty)$. Moreover, it satisfies

$$\lim_{\beta'\to 0}\frac{a_L(\beta')\big(a_L(\beta')-1\big)}{a_L'(\beta')} = 0, \qquad \lim_{\beta'\to \infty}\frac{a_L(\beta')\big(a_L(\beta')-1\big)}{a_L'(\beta')} = +\infty.$$

Proof. Define

$$g(\beta') \coloneqq \frac{a_L(\beta')\big(a_L(\beta')-1\big)}{a_L'(\beta')}.$$

We first prove that $g(\beta')$ is increasing on $(0,\infty)$. For any $i \in [L]$, define $x_i \coloneqq \log\!\left(\frac{L}{i}\right)$. Then $x_i \in [0, \log L]$, and we can rewrite

$$a_L(\beta') = \frac{1}{L}\sum_{i=1}^{L} e^{\beta' x_i}.$$

We define a random variable X supported on $\{x_1, x_2, \ldots, x_L\}$ with probability mass function

$$p_i \coloneqq \frac{e^{\beta' x_i}}{\sum_{j=1}^{L} e^{\beta' x_j}} = \frac{e^{\beta' x_i}}{L\, a_L(\beta')}, \qquad i = 1, \ldots, L.$$

Note that this notation p_i is unrelated to the "question distribution" used elsewhere in the paper; we reuse the symbol p here only to follow standard convention. With this definition, we have

$$a_L'(\beta') = \frac{1}{L}\sum_{i=1}^{L} x_i e^{\beta' x_i} = a_L(\beta')\,\mathbb{E}[X], \qquad a_L''(\beta') = \frac{1}{L}\sum_{i=1}^{L} x_i^2 e^{\beta' x_i} = a_L(\beta')\,\mathbb{E}[X^2].$$

Moreover, note that

$$\frac{\mathrm{d}}{\mathrm{d}\beta'}\mathbb{E}[X] = \frac{\mathrm{d}}{\mathrm{d}\beta'}\Big(\frac{a_L'(\beta')}{a_L(\beta')}\Big) = \frac{a_L''(\beta')a_L(\beta') - \big(a_L'(\beta')\big)^2}{\big(a_L(\beta')\big)^2} = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Therefore,

$$g(\beta') = \frac{a_L(\beta')\big(a_L(\beta')-1\big)}{a_L(\beta')\,\mathbb{E}[X]} = \frac{a_L(\beta')-1}{\mathbb{E}[X]}.$$

Differentiating yields

$$g'(\beta') = \frac{a_L'(\beta')\,\mathbb{E}[X] - \big(a_L(\beta')-1\big)\frac{\mathrm{d}}{\mathrm{d}\beta'}\mathbb{E}[X]}{\mathbb{E}[X]^2} = \frac{a_L(\beta')\,\mathbb{E}[X]^2 - \big(a_L(\beta')-1\big)\,\mathrm{Var}(X)}{\mathbb{E}[X]^2}.$$

Hence, to prove that $g(\beta')$ is increasing, it suffices to show that $g'(\beta') > 0$, i.e.,

$$\mathrm{Var}(X) < \frac{a_L(\beta')}{a_L(\beta')-1}\,\mathbb{E}[X]^2.$$

By Lemma C.9, for every $t \in [0, \log L)$,

$$\mathbb{E}\big[X - t \,\big|\, X > t\big] \leq \mathbb{E}\big[X \,\big|\, X > 0\big].$$

Therefore, with $(u)_+ \coloneqq \max\{u, 0\}$, for all $t \in [0, \log L)$, we have

$$\mathbb{E}\big[(X-t)_+\big] = \Pr(X > t)\,\mathbb{E}\big[X - t \,\big|\, X > t\big] \leq \mathbb{E}\big[X \,\big|\, X > 0\big]\Pr(X > t).$$

Meanwhile, let $p(x) = \sum_{i=1}^{L} p_i \delta(x - x_i)$ with the Dirac delta function $\delta(\cdot)$. We notice that

$$\int_0^{\log L} \mathbb{E}\big[(X-t)_+\big]\,\mathrm{d}t = \int_0^{\log L}\int_t^{\log L} (x-t)p(x)\,\mathrm{d}x\,\mathrm{d}t$$

$$= \int_0^{\log L}\int_t^{\log L} xp(x)\,\mathrm{d}x\,\mathrm{d}t - \int_0^{\log L} t\Pr(X>t)\,\mathrm{d}t$$
$$= \int_0^{\log L} t^2 p(t)\,\mathrm{d}t - \frac{1}{2}\mathbb{E}[X^2] = \mathbb{E}[X^2] - \frac{1}{2}\mathbb{E}[X^2] = \frac{1}{2}\mathbb{E}[X^2].$$

Then, we obtain

$$\mathbb{E}[X^2] = 2\int_0^{\log L}\mathbb{E}\big[(X-t)_+\big]\,\mathrm{d}t \le 2\,\mathbb{E}\big[X \mid X>0\big]\int_0^{\log L}\Pr(X>t)\,\mathrm{d}t = 2\,\mathbb{E}\big[X \mid X>0\big]\cdot\mathbb{E}[X].$$

Since $x_L = 0$, we have $p_L = \frac{1}{L\,a_L(\beta')}$. Therefore,

$$\mathbb{E}\big[X \mid X>0\big] = \frac{\mathbb{E}[X]}{1-p_L} = \frac{\mathbb{E}[X]}{1-\frac{1}{L\,a_L(\beta')}},$$

$$\mathbb{E}[X^2] \le \frac{2}{1-\frac{1}{L\,a_L(\beta')}}\,\mathbb{E}[X]^2,$$

and

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \le \big(\frac{2}{1-\frac{1}{L\,a_L(\beta')}} - 1\big)\,\mathbb{E}[X]^2 = \frac{L\,a_L(\beta')+1}{L\,a_L(\beta')-1}\,\mathbb{E}[X]^2.$$

Hence, it suffices to show that for all $L \ge 2$,

$$\frac{L\,a_L(\beta')+1}{L\,a_L(\beta')-1} < \frac{a_L(\beta')}{a_L(\beta')-1}.$$

Since $a_L(\beta') > 1$ and $L \ge 2$, it's equivalent to $a_L(\beta')\,(2-L) < 1$, which holds since $2-L \le 0$. Therefore,

$$\mathrm{Var}(X) \le \frac{L\,a_L(\beta')+1}{L\,a_L(\beta')-1}\,\mathbb{E}[X]^2 < \frac{a_L(\beta')}{a_L(\beta')-1}\,\mathbb{E}[X]^2.$$

It remains to verify the two limits. By a first-order Taylor expansion,

$$a_L(\beta') = \frac{1}{L}\sum_{i=1}^{L} e^{\beta' x_i} = 1+\beta'\cdot\frac{1}{L}\sum_{i=1}^{L} x_i + o(\beta'), \qquad a_L'(\beta') = \frac{1}{L}\sum_{i=1}^{L} x_i e^{\beta' x_i} = \frac{1}{L}\sum_{i=1}^{L} x_i + o(1),$$

as $\beta' \to 0$. Since $\frac{1}{L}\sum_{i=1}^{L} x_i > 0$, it follows that

$$\lim_{\beta'\to 0} g(\beta') = \lim_{\beta'\to 0}\frac{a_L(\beta')\big(a_L(\beta')-1\big)}{a_L'(\beta')} = \lim_{\beta'\to 0}\frac{\big(1+o(1)\big)\big(\beta'\cdot\frac{1}{L}\sum_{i=1}^{L} x_i + o(\beta')\big)}{\frac{1}{L}\sum_{i=1}^{L} x_i + o(1)} = 0.$$

As $\beta' \to \infty$, the sum defining $a_L(\beta')$ is dominated by its largest term, i.e.,

$$a_L(\beta') = \frac{1}{L}\sum_{i=1}^{L} e^{\beta' x_i} \sim \frac{1}{L}e^{\beta' x_1}, \qquad a_L'(\beta') = \frac{1}{L}\sum_{i=1}^{L} x_i e^{\beta' x_i} \sim \frac{1}{L}x_1 e^{\beta' x_1},$$

and hence

$$\mathbb{E}[X] = \frac{a_L'(\beta')}{a_L(\beta')} \to x_1 = \log L, \qquad a_L(\beta') \to \infty.$$

Therefore,

$$\lim_{\beta'\to\infty} g(\beta') = \lim_{\beta'\to\infty}\frac{a_L(\beta')-1}{\mathbb{E}[X]} = +\infty.$$

This completes the proof. $\square$

Lemma C.9. *Let $L \geq 2$ be an integer, and let X be the discrete random variable supported on*

$$\{x_1, \ldots, x_L\}, \qquad x_i \coloneqq \log\Big(\frac{L}{i}\Big) \in [0, \log L],$$

with probability mass function

$$\Pr(X = x_i) = p_i \coloneqq \frac{e^{\beta' x_i}}{\sum_{j=1}^{L} e^{\beta' x_j}}, \qquad i = 1, \ldots, L,$$

where $\beta' > 0$. Then for any $t \in [0, \log L)$, it holds that

$$\mathbb{E}\big[X - t \,\big|\, X > t\big] \leq \mathbb{E}\big[X \,\big|\, X > 0\big].$$

Proof. For any $t \in [0, \log L)$, there exists an index $i \in \{1, \ldots, L-1\}$ such that $x_{i+1} \leq t < x_i$. Hence

$$\mathbb{E}\big[X - t \,\big|\, X > t\big] \;=\; \mathbb{E}\big[X - t \,\big|\, X > x_{i+1}\big] \;\leq\; \mathbb{E}\big[X - x_{i+1} \,\big|\, X > x_{i+1}\big]$$

Therefore, it suffices to prove it on the support points, i.e., for any $t = x_i$ ($i \geq 2$),

$$\mathbb{E}\big[X - t \,\big|\, X > t\big] \leq \mathbb{E}\big[X \,\big|\, X > 0\big].$$

Let $w_i \coloneqq i^{-\beta'}$. Then $p_i = w_i / H_L$, where $H_L \coloneqq \sum_{i=1}^{L} w_i$. Moreover, for any $k \in \{1, 2, \ldots, L\}$, define $H_k \coloneqq \sum_{i=1}^{k} w_i$. For $k \geq 2$, define

$$m(k) \coloneqq \mathbb{E}\big[X - x_k \,\big|\, X > x_k\big].$$

Since $X > x_k$ is equivalent to $X \in \{x_1, \ldots, x_{k-1}\}$, we have

$$m(k) = \sum_{i=1}^{k-1} \frac{w_i}{H_{k-1}} \Big(\log\Big(\frac{L}{i}\Big) - \log\Big(\frac{L}{k}\Big)\Big) = \frac{1}{H_{k-1}} \sum_{i=1}^{k-1} w_i \,\log\Big(\frac{k}{i}\Big).$$

Next, for any $k \in \{2, 3, \ldots, L-1\}$, we can write

$$\begin{aligned} m(k+1) - m(k) &= \frac{1}{H_k} \sum_{i=1}^{k} w_i \,\log\Big(\frac{k+1}{i}\Big) - \frac{1}{H_{k-1}} \sum_{i=1}^{k-1} w_i \,\log\Big(\frac{k}{i}\Big) \\ &= \frac{1}{H_k}\, w_k \,\log\Big(1 + \frac{1}{k}\Big) + \sum_{i=1}^{k-1} w_i \left(\frac{\log\big(\frac{k+1}{i}\big)}{H_k} - \frac{\log\big(\frac{k}{i}\big)}{H_{k-1}} \right). \end{aligned}$$

For each $i \in \{1, \ldots, k-1\}$, we have

$$\begin{aligned} \frac{\log\big(\frac{k+1}{i}\big)}{H_k} - \frac{\log\big(\frac{k}{i}\big)}{H_{k-1}} &= \frac{H_{k-1} \log(k+1) - H_k \log k + (H_k - H_{k-1}) \log i}{H_{k-1} H_k} \\ &= \frac{H_{k-1} \log(k+1) - H_k \log k + w_k \log i}{H_{k-1} H_k} \\ &= \frac{H_{k-1} \log\big(1 + \frac{1}{k}\big) - w_k \log\big(\frac{k}{i}\big)}{H_{k-1} H_k}, \end{aligned}$$

Therefore,

$$\begin{aligned} m(k+1) - m(k) &= \frac{w_k}{H_k} \,\log\Big(1 + \frac{1}{k}\Big) + \sum_{i=1}^{k-1} w_i \cdot \frac{H_{k-1} \log\big(1 + \frac{1}{k}\big) - w_k \log\big(\frac{k}{i}\big)}{H_{k-1} H_k} \\ &= \frac{w_k}{H_k} \,\log\Big(1 + \frac{1}{k}\Big) + \frac{H_{k-1}}{H_k} \log\Big(1 + \frac{1}{k}\Big) - \frac{w_k}{H_k}\, m(k) \\ &= \log\Big(1 + \frac{1}{k}\Big) \;-\; \frac{w_k}{H_k}\, m(k). \end{aligned}$$

Note that $\mathbb{E}\big[X \,\big|\, X > 0\big] = m(L)$. Therefore, it suffices to show that $m(k)$ is nondecreasing in k, i.e.,

$$\log\Big(1 + \frac{1}{k}\Big) - \frac{w_k}{H_k}\, m(k) > 0, \qquad \forall\, k \in \{2, 3, \dots, L-1\},$$

equivalently,

$$m(k) < \log\Big(1 + \frac{1}{k}\Big)\, \frac{H_k}{w_k}, \qquad \forall\, k \in \{2, 3, \dots, L-1\}.$$

First,

$$\begin{aligned} H_{k-1} m(k) &= \sum_{i=1}^{k-1} w_i \log\Big(\frac{k}{i}\Big) = \sum_{i=1}^{k-1} w_i \sum_{j=i}^{k-1} \log\Big(\frac{j+1}{j}\Big) \\ &= \sum_{j=1}^{k-1} \log\Big(\frac{j+1}{j}\Big) \sum_{i=1}^{j} w_i = \sum_{j=1}^{k-1} \log\Big(\frac{j+1}{j}\Big)\, H_j. \end{aligned}$$

Therefore,

$$m(k) = \frac{1}{H_{k-1}} \sum_{j=1}^{k-1} \log\Big(\frac{j+1}{j}\Big)\, H_j \le \frac{k\, \log\big(1 + \frac{1}{k}\big)}{H_{k-1}} \sum_{j=1}^{k-1} \frac{H_j}{j},$$

since

$$\log\Big(1 + \frac{1}{j}\Big) \le \frac{k}{j} \log\Big(1 + \frac{1}{k}\Big).$$

Second, we show that the sequence $\big\{\frac{H_k}{k w_k}\big\}_{k \ge 1}$ is non-decreasing in k. To this end, note that

$$\frac{H_k}{k w_k} = \frac{1}{k} \sum_{i=1}^{k} \Big(\frac{i}{k}\Big)^{-\beta'}.$$

Let $f(x) \coloneqq x^{-\beta'}$, which is convex on $(0, \infty)$ for $\beta' > 0$. For each $i \in \{1, \dots, k\}$, observe that

$$\frac{i}{k} = \Big(1 - \frac{i}{k}\Big) \frac{i}{k+1} + \frac{i}{k} \cdot \frac{i+1}{k+1}.$$

By convexity of f, we obtain

$$\Big(\frac{i}{k}\Big)^{-\beta'} \le \Big(1 - \frac{i}{k}\Big) \Big(\frac{i}{k+1}\Big)^{-\beta'} + \frac{i}{k} \Big(\frac{i+1}{k+1}\Big)^{-\beta'}.$$

Summing over $i = 1, \dots, k$ yields

$$\sum_{i=1}^{k} \Big(\frac{i}{k}\Big)^{-\beta'} \le \frac{1}{k} \sum_{i=1}^{k} (k-i) \Big(\frac{i}{k+1}\Big)^{-\beta'} + \frac{1}{k} \sum_{i=1}^{k} i \Big(\frac{i+1}{k+1}\Big)^{-\beta'}.$$

Re-indexing the second sum, we get

$$\sum_{i=1}^{k} \Big(\frac{i}{k}\Big)^{-\beta'} \le \frac{1}{k} \left[\sum_{j=1}^{k} (k-j) \Big(\frac{j}{k+1}\Big)^{-\beta'} + \sum_{j=2}^{k+1} (j-1) \Big(\frac{j}{k+1}\Big)^{-\beta'} \right] = \frac{k-1}{k} \sum_{j=1}^{k} \Big(\frac{j}{k+1}\Big)^{-\beta'} + 1.$$

Dividing both sides by k and using

$$\frac{H_{k+1}}{(k+1) w_{k+1}} = \frac{1}{k+1} \sum_{j=1}^{k+1} \Big(\frac{j}{k+1}\Big)^{-\beta'}, \qquad \sum_{j=1}^{k} \Big(\frac{j}{k+1}\Big)^{-\beta'} = (k+1) \frac{H_{k+1}}{(k+1) w_{k+1}} - 1,$$

we obtain

$$\frac{H_k}{k w_k} \le \Big(1 - \frac{1}{k^2}\Big) \frac{H_{k+1}}{(k+1) w_{k+1}} + \frac{1}{k^2}.$$

Since $\frac{H_{k+1}}{(k+1) w_{k+1}} \ge 1$, it follows that

$$\frac{H_k}{k w_k} \le \frac{H_{k+1}}{(k+1) w_{k+1}},$$

i.e., $\frac{H_k}{k w_k}$ is non-decreasing in k.

Consequently, for any $j \in \{1, 2, \ldots, k\}$, this implies $\frac{H_j}{j w_j} \le \frac{H_k}{k w_k}$, equivalently,

$$\frac{H_j}{j^{1-\beta'}} \le \frac{H_k}{k^{1-\beta'}}.$$

Therefore,

$$\sum_{j=1}^{k-1} \frac{H_j}{j} = \sum_{j=1}^{k-1} \frac{H_j}{j^{1-\beta'}} \cdot j^{-\beta'} \le \frac{H_k}{k^{1-\beta'}} \sum_{j=1}^{k-1} j^{-\beta'} = \frac{H_k H_{k-1}}{k^{1-\beta'}}.$$

Plugging this bound into the previous estimate

$$m(k) \le \frac{k \log\left(1+\frac{1}{k}\right)}{H_{k-1}} \sum_{j=1}^{k-1} \frac{H_j}{j},$$

we obtain

$$m(k) \le \frac{k \log\left(1+\frac{1}{k}\right)}{H_{k-1}} \cdot \frac{H_k H_{k-1}}{k^{1-\beta'}} = \log\Big(1+\frac{1}{k}\Big) H_k\, k^{\beta'} = \log\Big(1+\frac{1}{k}\Big) \frac{H_k}{w_k}.$$

This completes the proof. □

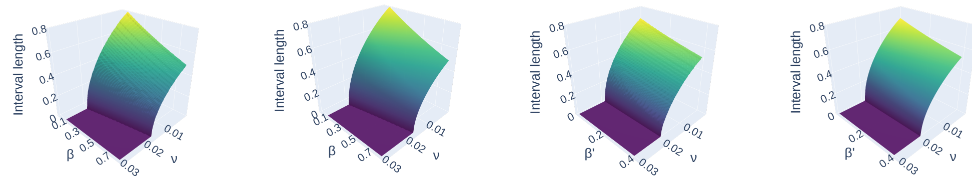
APPENDIX: ADDITIONAL RELATED WORK AND EXPERIMENTAL DETAILS

D. ADDITIONAL RELATED WORK

Self-improvement for LLM mathematical reasoning. Empirically, many pipelines for LLM reasoning tasks, especially mathematical reasoning, instantiate the idea of self-improvement via an iterative generate-and-filter loop: the model generates one or multiple candidate solutions for each problem, retains a subset of correct solutions, and then fine-tunes on the retained solutions to enhance performance (Zelikman et al., 2022; Xin et al., 2024a;b; Guo et al., 2025; Lin et al., 2025a;b; Ren et al., 2025; Zhang et al., 2025; Guan et al., 2025). While existing approaches vary in implementation details (e.g., incorporating long chain-of-thought (CoT) reasoning (Lin et al., 2025b), tactic annotations (Xin et al., 2024b), code-augmented CoT data (Guan et al., 2025), or reflection steps (Zhang et al., 2025)), their overall framework shares the same spirit of bootstrapping from model-generated attempts filtered by an explicit correctness signal. Additionally, a growing line of work (Ren et al., 2025; Koh et al., 2025; Lee et al., 2025) shows that combining self-improvement with an explicit easy-to-hard curriculum across rounds can further strengthen model performance.

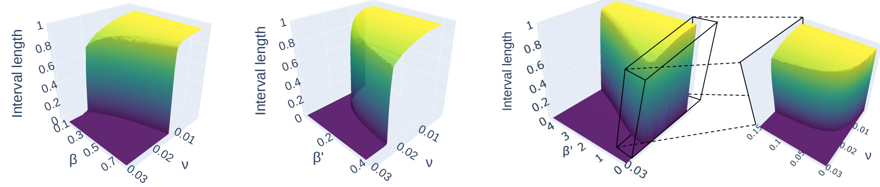
Self-distillation, self-consuming loops, and model collapse. First, several theoretical works on self-distillation analyze training a model with supervision signals generated by the model itself (Mobahi et al., 2020; Das & Sanghavi, 2023; Pareek et al., 2024). In contrast, we study LLM self-improvement for generative reasoning, which is typically outside the scope of standard self-distillation analyses (e.g., linear predictors). Second, another related line of work concerns self-consuming loops in generative models, a failure mode where repeatedly training on generated data can degrade performance and even lead to model collapse (Fu et al., 2024; 2025b). Recent theory shows that such degradation can be mitigated under suitable mechanisms (Gillman et al., 2024; Gerstgrasser et al., 2024; Ferbach et al., 2024; Feng et al., 2024; Fu et al., 2025a;b). While the objective of preventing model collapse in this literature differs from LLM self-improvement, Ferbach et al. (2024) is particularly relevant to our work as it explicitly characterizes how curated data can optimize a reward signal. However, it is not tailored to mathematical reasoning, and the resulting trends do not align as closely with empirical practice. A primary reason is that it does not account for the finite-sample regime, which, as we argue in this paper, is important to understanding self-improvement in mathematical reasoning.

E. DEFERRED EXPERIMENTAL FIGURES



(a) (b) (c) (d)

Figure 3: **Feasible initialization region.** Panels (a,c) report Monte-Carlo estimates of the length of the initialization interval $V_{p_0}(\hat{\theta}_0)$ for which $\{F^{\circ t}(V_{p_0}(\hat{\theta}_0))\}_{t\geq 0}$ and $\{(H_t \circ \cdots \circ H_0)(V_{p_0}(\hat{\theta}_0))\}_{t\geq 0}$ are both monotonically increasing in t, under different (β', β, ν) settings. Panels (b,d) show the length of the feasibility interval $\mathcal{I}_{\mathcal{M}}(\beta', \beta, \nu)$ in Corollary 5.3. Panels (a,b): fix $\beta' = 0.1$ and vary (β, ν). Panels (c,d): fix $\beta = 0.4$ and vary (β', ν).



(a) (b) (c)

Figure 4: **Improvement initialization region.** Panels (a)-(c) report Monte-Carlo estimates of the length of the initialization interval $V_{p_0}(\hat{\theta}_0)$ for which $(G \circ H_{L-1} \circ H_{L-2} \circ \cdots \circ H_0)\big(V_{p_0}(\hat{\theta}_0)\big) > F^{\circ L}\big(V_{p_0}(\hat{\theta}_0)\big)$ holds under different (β', β, ν) settings. Panel (a): fix $\beta' = 0.1$ and vary (β, ν). Panel (b): fix $\beta = 0.4$ and vary (β', ν). Panel (c): fix $\Delta = 0.1$ and vary (β', ν); the same panel also includes a zoomed-in view for small β'.

F. IMPLEMENTATION DETAILS AND TRAINING SETUP

F.1 GRAPH GENERATION AND SAMPLE POOL

We construct a balanced sample pool across configurations specified by the number of nodes N, target expected out-degree $\bar{d}$, and target distance l via rejection sampling. Concretely, we first generate a directed unweighted graph $\mathcal{G}$ on N labeled vertices by independently including each possible directed edge (v_i, v_j) with probability $\bar{d}/(N-1)$, so that the expected out-degree is $\bar{d}$. Given $\mathcal{G}$, we then sample two distinct query vertices $v_s \neq v_t$ and compute the shortest path length from v_s to v_t using a standard Breadth-First Search routine on directed edges. If the resulting distance equals the target l (with the convention $l = -1$ when v_t is unreachable from v_s), we add the instance $(\mathcal{G}, v_s, v_t, l)$ to the pool associated with $(N, \bar{d}, l)$. We repeat this procedure until we either collect 2000 instances for each $(N, \bar{d}, l)$ combination or reach a preset sampling limit; if fewer than 2000 instances are found for a combination, we keep all collected instances.

In our experiments, we take $N \in \{6, 8, 10, 12, 14, 16, 18\}$, $\bar{d} \in \{2, \ldots, \lfloor N/2 \rfloor\}$, and $l \in \{-1, 1, 2, 3\}$. Each retained instance is rendered into a natural language prompt using a unified template. The prompt template is:

```
You are given a directed unweighted graph with nodes labeled
1..N.
N = ⟨N⟩
Edges are listed as ordered pairs (u,v), where each (u,v)
represents a directed edge from u to v:
⟨edge list⟩
Start s = ⟨v_s⟩, Target t = ⟨v_t⟩
Question:  Output the length (number of edges) of the
shortest path from s to t.  If no path exists, output -1.
Answer with a single integer only.
```

The union of all instances across $(N, \bar{d}, l)$ forms our overall sample pool, from which we subsequently construct dataset splits for different experimental settings.

F.2 Warm-up and Self-Improvement Finetuning Details

Our synthetic shortest path task allows control of the initialization reward $V_{p_0}(\hat{\theta}_0)$ and $V_{p_i}(\hat{\theta}_0)$ for $i \in [L]$. Note that, in our binary reward setting, $V_p(\theta)$ corresponds to the (population) Pass@1 accuracy of the model θ evaluated on questions drawn from p. Across experiments, we obtain different initial Pass@1 accuracies (corresponding to different values of $V_p(\theta)$) by varying the initialization model via warm-up finetuning and by varying the task difficulty through selecting different subsets of the overall sample pool.

Concretely, our warm-up datasets and the datasets used for self-improvement are sampled as disjoint subsets from the overall sample pool. Each warm-up dataset has a balanced composition across different $(N, \bar{d}, l)$ combinations, i.e., it contains equal numbers of samples for each $(N, \bar{d}, l)$. We warm up the pretrained base LLM using different warm-up datasets under different training configurations (learning rate and random seed) to obtain different initialization models.

Next, for self-improvement, given an initialization model $\hat{\theta}_0$, we select an appropriate training set of size nL together with a held-out test set such that the empirical Pass@1 accuracy of $\hat{\theta}_0$ matches the target value $V_{p_0}(\hat{\theta}_0)$, enabling control of the initialization performance. We fix the test set across all L iterations, with a test size of 1000. To mitigate the effect of class imbalance, we enforce that, for both the per-iteration training set and the test set, the number of questions in each distance class $l \in \{-1, 1, 2, 3\}$ is approximately balanced (up to a small tolerance); moreover, within each class, we also match (up to a small tolerance) the number of questions that are initially answered correctly by $\hat{\theta}_0$. For experiments comparing easy-to-hard against the baseline, we require a different form of control. Specifically, the training set of size nL should admit two different partitions: (i) a baseline partition whose L subsets (each containing n questions for one iteration) have the approximately same (up to a small tolerance) initial Pass@1, matching $V_{p_0}(\hat{\theta}_0)$; and (ii) an easy-to-hard partition whose initial Pass@1 values across iterations satisfy Assumption 5.1. we impose these constraints jointly and solve the resulting data selection problem using a CP-SAT solver.

Hyperparameter	Value
Learning rate	2×10^{-4}
Batch size	8
LoRA rank	16
LoRA scaling	32
LoRA dropout	0.05

Table 2: Self-improvement finetuning hyperparameters used at each iteration.

Figure	Initial test Pass@1	n	m	β'	Δ
Figure 1(a)	–	5,000	1	–	–
Figure 1(b)	0.32	–	1	–	–
Figure 1(c)	0.32	4,000	–	–	–
Figure 2(a)	–	3,000	1	–	0.04
Figure 2(b)	–	–	1	0.25	0.04

Table 3: Experimental settings for Figures 1 and 2.

Given an initialization model $\hat{\theta}_0$ and a constructed dataset, we perform iterative self-improvement exactly following the setup in Sections 3 and 5.1. At each iteration, we use the same finetuning hyperparameters summarized in Table 2, and train for one epoch by default, with a minimum of 50 optimization steps for cases with too few accepted samples. Additionally, each data point in Figures 1 and 2 is obtained by averaging over five runs with different random seeds. Table 3 summarizes the experimental settings that are held fixed within each panel of Figures 1 and 2.