1 NOTATION.

Let $\mathbf{K}^\star$ be its optimal knowledge. $\mathbf{K}_t$: knowledge produced by the actor π_θ after t refinement rounds. $\mathcal{L}(\mathbf{K}_1, \mathbf{K}_2) \in [0,1]$: task-level distance. $e_t := \mathcal{L}(\mathbf{K}_t, \mathbf{K}^\star)$: error at round t. $\mathbf{r}_t := \pi_\psi(\mathbf{K}_t)$: feedback from the reflection model π_ψ. $\mathbf{K}_{t+1} = T(\mathbf{K}_t) := \pi_\theta(\mathbf{K}_t, \mathbf{r}_t)$: refinement operator: π_θ refines original knowledge with reflection. The loop terminates when π_ψ declares the candidate *reasonable*.

2 ASSUMPTIONS.

This is some assumptions.

A1: Reflection effectiveness

$\Pr\big[\pi_\psi(\mathbf{K}) \text{ detects errors}\big] \;\geq\; \alpha,\; \alpha \in (0,1]$.

With probability $\alpha \in (0,1]$ the reflection model π_ψ can detect errors in $\mathbf{K}$.

A2: Refinement contractivity

$\mathbb{E}\big[\mathcal{L}(\pi_\theta(\mathbf{K}, \pi_\psi(\mathbf{K})), \mathbf{K}^\star)\big] \leq \gamma\, \mathcal{L}(\mathbf{K}, \mathbf{K}^\star)$ for $\gamma \in (0,1)$ whenever the feedback is correct.

Conditional on receiving correct feedback, π_θ shrinks the loss by a factor γ, which means once the actor receives correct feedback, π_θ reduces the current error by a multiplicative factor

A3: Bounded refinement noise There exists a constant $\eta \geq 0$ such that for any knowledge state $\mathbf{K}$,

$$\mathbb{E}\Big[\big|\mathcal{L}\big(\pi_\theta(\mathbf{K},\, \pi_\psi(\mathbf{K})),\, \mathbf{K}\big)\big|\Big] \;\leq\; \eta,$$

i.e., a single refinement step can increase the expected error by at most η.

3 PROOF

At round t two cases arise:

Event	Probability	$\mathbb{E}[e_{t+1} \mid e_t]$ upper-bound
Reflection *hit*	α	$\gamma\, e_t + \eta$
Reflection *miss*	$1-\alpha$	$e_t + \eta$

Hence

$$\mathbb{E}[e_{t+1} \mid e_t] \leq \big[1 - \alpha(1-\gamma)\big]\, e_t + \eta. \tag{1}$$

Define $\rho := 1 - \alpha(1-\gamma) \in (0,1)$; then

$$E[e_{t+1}] \;\leq\; \rho\, e_t + \eta. \tag{2}$$

Solving (2) yields

$$E[e_t] \;\leq\; \rho^{\,t}\, e_0 \;+\; \frac{1-\rho^{\,t}}{1-\rho}\,\eta. \tag{3}$$

Consequently

$$\lim_{t\to\infty} \mathbb{E}[e_t] = \frac{\eta}{1-\rho} = \frac{\eta}{\alpha(1-\gamma)}. \tag{4}$$

This establishes that the reflection–refinement loop converges, and that the asymptotic error stabilises at $\frac{\eta}{\alpha(1-\gamma)}$.

Iterations to reach a target error ε. For any $\varepsilon > \eta/(1-\rho)$ the minimal integer $T(\varepsilon)$ satisfying $\mathbb{E}[e_t] \leq \varepsilon$ is

We want to guarantee $\mathbb{E}[e_t] \leq \varepsilon$ (with $\varepsilon > \eta/(1-\rho)$, otherwise the requirement is infeasible). From Eq.(3) it follows that

$$\rho^{\,t} e_0 \;\leq\; \varepsilon - \frac{1-\rho^{\,t}}{1-\rho}\,\eta \;<\; \varepsilon - \frac{\eta}{1-\rho}, \tag{5}$$

hence

$$\rho^{\,t} \;\leq\; \frac{\varepsilon - \eta/(1-\rho)}{e_0}. \tag{6}$$

Taking logarithms on both sides gives

$$t \;\geq\; \frac{\log\!\big((\varepsilon - \eta/(1-\rho))/e_0\big)}{\log \rho}. \tag{7}$$

4 CONCLUSION

We hope the above proof may address the reviewer question.

Q1: Why does the reflection–refinement loop converge? Eq. (3) and Eq. (4) Shows that the eometric convergence of the reflection–refinement loop under mild assumptions with rate ρ and stabilises at $\frac{\eta}{\alpha(1-\gamma)}$.

Q2: How many iterations are needed to reach a target error tolerance?

The number of iterations required to reach the target error tolerance is $\frac{\log\left((\varepsilon-\eta/(1-\rho))/e_0\right)}{\log \rho}$.