MORE OPTIMIZATION, WORSE ALIGNMENT: CYCLES AND SAFEGUARDS IN ITERATIVE DPO

Anonymous authors
Paper under double-blind review

ABSTRACT

Iterative preference learning fits a policy, adopts it as the next reference, and uses it to generate new comparisons. Better fitting appears desirable, but adoption changes the next optimization problem. We show that this feedback can turn deeper direct preference optimization (DPO) into a persistent reward loss. Under a fixed Bradley–Terry reward, we construct a regular nine-response policy family where one-step DPO converges globally, whereas 32-step DPO admits an attracting suboptimal two-cycle. Every frozen inner problem has globally convergent gradient descent, and the unique outer fixed point remains locally stable and reward-optimal within the family. Matched IPO converges globally at the same depth, scale, and learning rate, despite agreeing with DPO through second order at the optimum. We explain the separation by connecting the fourth loss derivative, comparison moments, and inner depth to the leading possible cubic difference. Even an arbitrarily small logistic component in an IPO mixture can produce cycles, with model-dependent depth and coexistence windows. A certified fifth-order analysis quantifies cycle onset and the GD step sizes needed to preserve coexistence. Finally, an IPO-envelope safeguard restores convergence in the product model by controlling the adopted displacement, while allowing arbitrary inner depths and exact minimization. Together, these results show why reliable iterative alignment requires control of policy adoption beyond accurate preference fitting.

1 INTRODUCTION

More accurate preference fitting is a natural goal inside an alignment round. Whether the fitted policy should then be adopted in full is a separate question. We show that this distinction can determine the long-run reward: the full update can sustain an attracting low-reward cycle, while a safeguarded fraction of the same update converges to the reward optimum.

Direct preference optimization (DPO) replaces an explicit reward-learning and policy-optimization pipeline by a logistic loss on policy log-ratios (Rafailov et al., 2023). In an iterative implementation, the current policy can also generate the next comparison batch and become the next reference policy. The resulting algorithm has two distinct dynamics: optimization of a frozen preference loss inside a round, and feedback through the policy between rounds. Convergence of the first does not establish convergence of the second. This distinction matters when the reward represented by a policy log-ratio cannot express the reward that generates preferences.

The issue is more specific than whether a misspecified preference objective can reduce reward, or whether feedback can create oscillations. Both are known concerns (Gopalan et al., 2025; Chen et al., 2026). We ask whether *inner optimization depth itself* can separate two iterative preference procedures when their inner objectives share the same quadratic geometry, their equal-depth outer maps agree through second order at the optimum, and each frozen optimization problem is solved by a globally stable gradient method. An affirmative answer would rule out a common form of local justification: combining a well-behaved inner solver with correct local curvature and a stable reward-optimal fixed point is still insufficient.

We give such an answer in a finite, fully specified model. Preferences are generated by a single fixed scalar reward through the Bradley–Terry (BT) rule. Thus preference intransitivity, an evolving reward model, singular Fisher information, and instability of the frozen inner gradient iteration are

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

absent. The policy family is restricted, however, and its log-ratios do not realize the true reward. The performance target is the maximum expected true reward *within that same family*; the target is attained at a unique finite parameter.

Our principal construction uses ordinary Euclidean gradient descent (GD). With the same learning rate and log-ratio scale, one-step DPO converges to the optimal policy from every finite initialization, whereas 32-step DPO has a full-dimensional basin of attraction to a two-cycle. The reward gap on the cycle is exactly $\gamma/8 > 0$. Every finite depth of the matched IPO procedure converges globally. The DPO cycle coexists with a locally stable, uniquely reward-optimal fixed point. These conclusions use exact inequalities; floating-point trajectories are not used as proofs.

Three results organize the paper. First, the explicit separation shows that stable inner optimization, an attainable reward optimum, and correct second-order local behavior do not certify reward consistency. Second, we connect loss nonlinearity and inner depth to attracting coexistence. A transferable formula uses the loss's fourth derivative, comparison moments, and the local sensitivity of the preference gradient. Even an arbitrarily small positive logistic component can admit cycles on an explicit model region. A certified fifth-order expansion then quantifies the onset in DPO: a quadratic coexistence window, linear reward gaps, and a square-root axial basin radius. Joint limits identify how small finite GD steps must be to retain that coexistence. Near IPO, explicit depth and window rates quantify the price of the counterexample. These guarantees concern refreshed outer dynamics beyond a correspondence between frozen losses. Third, an IPO-envelope safeguard restores convergence without truncating the deep fit, including unbounded history-dependent depths. Exact IPO already converges here, and sufficiently small constant damping can work; the safeguard proves a mechanism for recovery, without claiming an efficiency or performance advantage. Extensions for the explicit DPO construction cover solver limits, model perturbations, and finite-horizon sampling in the appendix.

The scope is a limitation theorem for logistic fitting on a restricted policy manifold with reference/sampling refresh. Numerical depths are normalization-specific; neither a universal algorithm ranking nor a neural policy guarantee follows. The appendix collects supporting lemmas and additional statements before the proofs and exact certificates.

2 Related Literature

Preference losses and their local correspondence. DPO expresses KL-regularized reward optimization through policy log-ratios (Rafailov et al., 2023); IPO fits a squared target margin (Gheshlaghi Azar et al., 2024). Generalized preference optimization unifies convex preference losses and establishes their second-order Taylor correspondence after rescaling (Tang et al., 2024, Section 2 and Appendix B.2). Within this convex framework, we connect the fourth loss derivative to inner depth, cubic outer response, and attracting cycles, quantifying the model-dependent depths and shrinking windows near IPO.

Representation constraints and reward gaps. Nika et al. (2024) compare RLHF and DPO with loglinear policies, separating statistical, approximation, and optimization errors. Early stopping and restricted policy classes can invalidate the standard DPO–RLHF equivalence (Kong et al., 2025). Gopalan et al. (2025) link sampling-weighted reward projection under misspecification to preference reversals and reward deterioration. Shi et al. (2025a) compare RLHF, DPO, and online DPO under reward/policy representation gaps, with exact-optimization and finite-sample separations targeting fixed-reference KL-regularized value. We track unregularized reward through refreshed rounds: the unique, locally stable fixed point attains the class optimum, yet a lower-reward attractor coexists.

Overoptimization and statistical safeguards. Rafailov et al. (2024) study direct-alignment overoptimization across policy sizes, regularization, and training progress. Worsening alignment with fitting is therefore known; we certify a dynamical mechanism under fixed BT preferences. χPO changes DPO's log-ratio link to implement pessimism and obtains offline sample-complexity guarantees under policy realizability and coverage conditions (Huang et al., 2025). This addresses statistical extrapolation from offline data. Our cycles already exist at the population level and survive the stated finite-horizon sampling limit. Our displacement envelope is not an alternative statistical guarantee to χPO.

Online sampling and preference games. Xiong et al. (2024) develop exploration and sample-efficiency guarantees for KL-constrained iterative preference learning. Shi et al. (2025b) separate linear and quadratic exact-gradient DPO convergence through sampler design. We match within-round sampling across losses and vary inner depth under joint reference refresh. Nash learning replaces scalar-reward maximization by a preference game (Munos et al., 2024), and online IPO connects to this framework (Calandriello et al., 2024). COMAL gives convergence guarantees for general preferences with approximate inner solutions and also exhibits iterative DPO/IPO cycling under intransitivity (Liu et al., 2024). Magnetic preference optimization uses successive regularized games to obtain last-iterate convergence to an unregularized Nash equilibrium (Wang et al., 2025). Chen et al. (2026, Proposition 33) analyze DPO oscillations in a Condorcet-cycle regime with reference/sampling feedback. Our comparison uses one scalar BT reward, specified finite GD, and an attained class optimum; the target is reward consistency, rather than game exploitability. Positive optimization guarantees under exponential-tilting realizability (He et al., 2024) remain compatible because our family lacks that closure.

Reference updates and controlled adoption. TR-DPO mitigates offline overoptimization by averaging the trained model into the reference or periodically copying it (Gorbatovski et al., 2025). We freeze the reference and sampling law inside each round, damp the candidate-policy displacement, and use the accepted policy for both next-round roles. This changes the update target and data regime. The envelope proves recovery while preserving deep inner fitting; it claims no computational or reward advantage over direct IPO or sufficiently small constant damping. Reference refresh and damping themselves are established tools.

Feedback and optimization dynamics. Performative prediction formalizes model-dependent data distributions (Perdomo et al., 2020); iterative RLHF can also create feedback through reward-model retraining (Gauthier et al., 2026). Our reward is fixed. For fixed logistic objectives, Meng et al. (2024) construct GD cycles, including coexistence with a stable minimizer, at steps beyond a global convergence bound. Here the inner loss is globally $1/3$-smooth and unit-step GD converges: the cycle belongs to the outer reference/sampling iteration. The supplementary frozen-curvature update uses classical quadratic majorization (Böhning & Lindsay, 1988).

3 Main Results

3.1 The iterative procedure and the reward target

Let $\mathcal{Y}$ be finite and

$$p_\theta(y) = h(y)\exp\{\theta^\top \phi(y) - \psi(\theta)\}, \qquad \theta \in \mathbb{R}^m, \tag{1}$$

where $h(y) > 0$ and the differences of the sufficient statistics span $\mathbb{R}^m$. At each outer round, set the reference to p_θ and draw I, J independently from p_θ. Conditional on (I, J), the label $A \in \{0, 1\}$ has probability $q_{IJ} = \sigma(r(I) - r(J))$ of equaling one, where $\sigma(u) = (1 + e^{-u})^{-1}$ and r is a fixed finite reward. Write

$$d = \phi(I) - \phi(J), \quad M_\theta = \mathbb{E}_\theta[dd^\top] = 2\mathcal{I}(\theta), \quad b_\theta = \mathbb{E}_\theta[d(q_{IJ} - 1/2)]. \tag{2}$$

The inner coordinate is $w = \beta(\vartheta - \theta)$, where ϑ is the candidate policy parameter and $\beta > 0$. The DPO pair logit is exactly $d^\top w$, since the log-partition terms cancel. Define

$$L_{\mathrm{D}}(w; \theta) = \mathbb{E}_\theta[\log(1 + e^{d^\top w}) - q_{IJ} d^\top w], \tag{3}$$

$$L_{\mathrm{Q}}(w; \theta) = \frac{1}{8}\mathbb{E}_\theta[(d^\top w - 2(2A - 1))^2] = \frac{1}{2} - b_\theta^\top w + \frac{1}{8} w^\top M_\theta w. \tag{4}$$

The second loss is IPO with its scale chosen to match the quadratic part of L_{D}. Under the raw IPO convention $\mathbb{E}[(\log(p_\vartheta(I)/p_\theta(I)) - \log(p_\vartheta(J)/p_\theta(J)) - (2A - 1)/(2\tau))^2]$, we set $\tau = \beta/4$ and multiply the loss by $\beta^2/8$. This scaling is material for GD and is part of the comparison.

For $X \in \{\mathrm{D}, \mathrm{Q}\}$, a depth-$K$ round is

$$w_{X,0} = 0, \quad w_{X,j+1} = w_{X,j} - \eta \nabla_w L_X(w_{X,j}; \theta), \quad F_{X,K}(\theta) = \theta + \beta^{-1} w_{X,K}(\theta). \tag{5}$$

Both the reference and sampling law are frozen throughout these K steps. GD in w is ordinary GD in ϑ with step η/β^2; no matrix preconditioning is present. The next round repeats the construction at $F_{X,K}(\theta)$. Our criterion is $J(\theta) = \mathbb{E}_{p_\theta} r(Y)$ and $J_* = \sup_{\theta \in \mathbb{R}^m} J(\theta)$.

3.2 A REWARD CYCLE UNDER GLOBALLY STABLE INNER GD

Consider nine responses $(x, z) \in \{-1, 0, 1\}^2$, features $\phi(x, z) = (x, z)$, and $\theta = (s, t)$, with

$$p_{s,t}(x, z) = \frac{h(x)e^{sx}}{2 + 2\cosh s} \frac{e^{tz}}{1 + 2\cosh t}, \qquad h(0) = 2,\ h(\pm 1) = 1, \qquad r(x, z) = -\gamma x^2 - \delta z^2, \tag{6}$$

where $\delta = 1/20$ and $\gamma > 0$ is the unique solution of

$$\frac{5}{9}\tanh\frac{\gamma}{2} + \frac{2}{9}\tanh\frac{\gamma - \delta}{2} + \frac{2}{9}\tanh\frac{\gamma + \delta}{2} = \frac{3}{4}. \tag{7}$$

Both coordinates affect the true reward and the preference law. The Fisher information is positive definite at every finite parameter. The true reward is not affine in (x, z), so it is not realizable by policy log-ratios in this family. Nevertheless, the within-family target is attained uniquely at zero:

$$J(s, t) = -\gamma\frac{\cosh s}{1 + \cosh s} - \delta\frac{2\cosh t}{1 + 2\cosh t}, \qquad J_* = -\gamma/2 - 2\delta/3. \tag{8}$$

Define a finite real sequence and the scale parameter by

$$v_0 = 0, \quad v_{j+1} = v_j + \frac{9}{64} - \frac{15}{64}\tanh(v_j/2) - \frac{9}{128}\tanh v_j, \qquad \beta_{32} = \frac{v_{32}}{2\log 3}. \tag{9}$$

These are exact definitions. In particular, $151/400 < \beta_{32} < 236/625$; its decimal value is approximately 0.3775410212.

Theorem 3.1 (Stable inner optimization, different outer rewards)**.** *For (6)–(7), $\eta = 1$ and $\beta = \beta_{32}$, all of the following hold.*

1. *For either inner loss and every finite outer parameter, GD converges to its unique minimizer. Each inner step satisfies*
$$L_X(w - \nabla L_X(w); \theta) \le L_X(w; \theta) - \frac{5}{6}\left\|\nabla L_X(w; \theta)\right\|^2. \tag{10}$$
2. *$F_{\mathrm{D},1}$ and $F_{\mathrm{Q},K}$ for every fixed finite $K \ge 1$ converge to zero from every finite initialization; their true reward increases strictly at every nonstationary outer step.*
3. *$F_{\mathrm{D},32}$ has the attracting two-cycle $a_+ = (\log 3, 0) \leftrightarrow a_- = (-\log 3, 0)$, with an open basin in $\mathbb{R}^2$. Its only fixed point is zero, which is locally asymptotically stable.*
4. *At both cycle points, $J_* - J(a_\pm) = \gamma/8$. The equal-depth maps $F_{\mathrm{D},32}$ and $F_{\mathrm{Q},32}$ have identical Taylor expansions through order two at zero.*

The cycle is not produced by a large-step failure on a fixed logistic objective. In this family $0 \prec \nabla_w^2 L_{\mathrm{D}} \preceq M_\theta/4 \preceq \mathrm{Id}/3$ and $\nabla_w^2 L_{\mathrm{Q}} = M_\theta/4$. Its attraction is also two-dimensional: the Jacobian at either cycle point is diagonal, with

$$-\frac{179}{200} < \lambda_{\mathrm{cyc},x} < -\frac{893}{1000}, \qquad \frac{8}{9} < \lambda_{\mathrm{cyc},z} < 1. \tag{11}$$

We prove the first interval by a finite rational certificate and the second analytically. Corollary B.3 gives an explicit rational interval of β containing β_{32} and shared by four depths. Furthermore, at a_+ the last inner step, from 31 to 32, strictly decreases both the DPO loss and the true reward of the resulting candidate policy. This statement does not assert convergence of the depth-31 outer procedure.

3.3 LOSS NONLINEARITY AND INNER DEPTH

The leading possible cubic difference in Theorem B.4 is the first place where equal-depth DPO and IPO can disagree near a stationary point. It vanishes at depth one. To identify the part due to the loss, consider even real-analytic potentials with

$$L_\Phi(w; \theta) = \mathbb{E}_\theta \Phi(d^\top w) - b_\theta^\top w, \qquad \Phi''(0) = \frac{1}{4}, \qquad \omega_\Phi = -8\Phi^{(4)}(0). \tag{12}$$

IPO has $\Phi_{\mathrm{Q}}(u) = u^2/8$ and DPO has $\Phi_{\mathrm{D}}(u) = \log(2\cosh(u/2))$, up to constants independent of w. Thus $\omega_{\mathrm{Q}} = 0$ and $\omega_{\mathrm{D}} = 1$. Use the GD rule (5) with L_Φ in place of L_X. This is an even-potential subclass with a linear preference term; we do not identify it with all convex losses in GPO.

Let $b_{\theta_*} = 0$ and let e be a unit vector such that, at $\theta_* + se$, the inner gradients at we stay parallel to e for both compared losses. Use s, w for this scalar restriction and assume reflection symmetry $L_\Phi(-w; -s) = L_\Phi(w; s)$ up to a term independent of w. For $d_e = e^\top d$, write

$$b_e(s) = -As + Bs^3 + O(s^5), \qquad \frac{1}{4}\mathbb{E}_s d_e^2 = H + H_2 s^2 + O(s^4), \qquad Q_4 = \mathbb{E}_0 d_e^4, \tag{13}$$

where $A, H > 0$. For $0 < \eta H < 1$, define

$$r = 1 - \eta H, \quad d_K = 1 - r^K, \quad u_K = \frac{K(1-r)r^{K-1}}{d_K}, \quad \Lambda = \frac{Q_4 A^2}{24 H^3},$$
$$W_K = \frac{1-r}{d_K}\sum_{j=0}^{K-1} r^{K-1-j}(1-r^j)^3, \qquad \beta_{c,K} = \frac{A d_K}{2H}. \tag{14}$$

For a concrete model region, we use three responses $x \in \{-1, 0, 1\}$:

$$p_s(x) = \frac{h_\alpha(x)e^{sx}}{\alpha + 2\cosh s}, \quad h_\alpha(0) = \alpha, \quad h_\alpha(\pm 1) = 1, \quad r_g(x) = -gx^2, \quad c = \tanh(g/2). \tag{15}$$

Theorem 3.2 (From loss nonlinearity to attracting reward cycles)**.** *At the common critical scale $\beta_{c,K}$, the restricted outer map is $-s + \chi_{\Phi,K} s^3 + O(s^5)$, where*

$$\chi_{\Phi,K} = \underbrace{\frac{2B}{A} + \frac{2H_2}{H}(1-u_K)}_{\chi_{\mathrm{Q},K}} - \omega_\Phi \Lambda W_K. \tag{16}$$

If $H_2 \le 0$ and $0 < \chi_{\mathrm{Q},\infty} := 2B/A + 2H_2/H < \omega_\Phi \Lambda$, the IPO coefficients stay positive while the Φ coefficients strictly decrease to a negative limit.

In (15), assume $0 < \Phi''(u) \le 1/4$ for all u, $\lim_{u\to\infty} \Phi'(u) > c/2$, and

$$1 < \alpha \le 4, \qquad \omega_\Phi c^2 > T(\alpha) := \frac{(\alpha-1)(\alpha+2)^2}{\alpha^2(\alpha+8)}. \tag{17}$$

At unit inner step size, every sufficiently large finite depth K then admits a nonempty open scale interval above $\beta_{c,K}$ where the unique, stable, reward-optimal fixed point coexists with an attracting lower-reward two-cycle. IPO at that same depth and scale converges globally to the optimum. Both frozen inner solvers converge globally. For every positive logistic mixture weight $\lambda \le 1$, these conclusions hold on an open region of (α, c) for $L_\lambda = (1-\lambda)L_{\mathrm{Q}} + \lambda L_{\mathrm{D}}$.

The mixture has $\omega_\Phi = \lambda$; the explicit rectangle $1 + \lambda/16 < \alpha < 1 + \lambda/8$, $1/2 < c < 3/4$ suffices. Since $T(\alpha) \to 0$ as $\alpha \downarrow 1$, no positive logistic weight is uniformly safe over the admissible models and inner depths. The potentials converge to IPO in every fixed C^m norm on each compact logit interval as $\lambda \downarrow 0$. The model region, required depth, and scale interval may all depend on λ; the theorem does not preserve a cycle in one fixed model as $\lambda \to 0$.

The tail-slope condition is a separate global requirement; the fourth derivative alone does not certify attraction.

How close to IPO, and how deep? The dependence on λ can be quantified. Fix $0 < c < 1$ and $0 < \vartheta < c^2$, and set $\alpha_\lambda = 1 + \vartheta\lambda$. At unit step, Proposition B.6 gives the first negative-cubic depth

$$K_{\mathrm{sgn}}(\lambda) = \frac{\log(1/\lambda) + \log\log(1/\lambda)}{\log(3/2)} + O(1). \tag{18}$$

For deeper finite choices satisfying $u_{K_\lambda} = o(\lambda)$, the local coexistence width is

$$\beta_{\mathrm{f}} - \beta_{c,K_\lambda} = \frac{5c(c^2-\vartheta)^2}{27}\lambda^2 + o(\lambda^2), \tag{19}$$

and cycle reward gaps are of order λ at a fixed interior relative scale. Thus approaching IPO delays cubic sign reversal and narrows this local failure window. The result supplies neither a fixed reward gap as $\lambda \to 0$ nor the first depth with any cycle. An exact finite-depth sign test and certified examples accompany the proof.

The rates have a common origin. At the endpoint $\lambda = 0$, $\alpha = 1$, the exact IPO map at its critical scale has the expansion

$$F_{0,\infty}(s) = s - \frac{6 \sinh s}{\cosh s + 2} = -s + \frac{s^5}{90} + O(s^7). \tag{20}$$

The cubic restoring margin vanishes while the quintic margin and inner curvature $H = 1/3$ stay positive. Along the model family, a negative cubic term of order λ competes with a finite-depth residual $K(2/3)^K$, giving the logarithm and log-log correction. Balancing the cubic and quintic terms gives $a^2 \asymp \lambda$ and the quadratic window width. This mechanism approaches a vanishing nonlinear restoring margin; it gives no instability guarantee for an arbitrary fixed model.

Interpreting the criterion. A measures the initial restoring preference gradient, B/A its relative cubic weakening, and H_2 the change of inner curvature. For a centered symmetric scalar feature X, with $\mu_j = \mathbb{E}_0 X^j$ and kurtosis $\varkappa = \mu_4/\mu_2^2$, Lemma A.12 gives $H = \mu_2/2$, $H_2 = (\mu_4 - 3\mu_2^2)/4$, and $Q_4 = 2\mu_4 + 6\mu_2^2$. The sufficient sign condition is therefore

$$\varkappa \le 3, \qquad 0 < \frac{2B}{A\mu_2} + \varkappa - 3 < \omega_\Phi \frac{\varkappa + 3}{6} \left(\frac{A}{H}\right)^2. \tag{21}$$

It compares IPO's remaining cubic restoring margin with the accumulated loss correction. Heavy tails are unnecessary; changing kurtosis alone changes both sides and does not predict the outcome.

The three-response example makes the representation constraint visible: a linear log-ratio cannot favor the central response over both endpoints, although the policy at $s = 0$ uniquely maximizes its probability and reward. The inner update points toward zero but can be adopted beyond it. Bounded inner displacement turns the local sign reversal into an attracting cycle. The nine-response construction adds transverse attraction and a common numerical scale interval. Further interpretation and the original three-response calculation appear in Appendix B.1. In that normalized nine-response model, the cubic sign changes between depths 15 and 16; this remains a local criticality threshold, not the first depth admitting any cycle or a universal threshold at fixed β.

3.4 From local criticality to coexistence

The negative cubic coefficient identifies an unstable local cycle on the stable side of the optimal fixed point. We combine this information with global control of the frozen DPO update to obtain a separate attracting cycle. For fixed $0 < \eta \le 3$ and finite K, set $\nu_K(s) = -w_{\mathrm{D},K,x}(s, 0)$ and define

$$g_{K,\beta}(s) = s - \nu_K(s)/\beta, \qquad \mathcal{A}_K(s) = \nu_K(s)/(2s)\ (s > 0), \qquad \mathcal{A}_K(0) = \beta_{c,K}. \tag{22}$$

The amplitude curve records when one outer update reaches the reflected parameter:

$$g_{K,\beta}(s) = -s \quad \Longleftrightarrow \quad \mathcal{A}_K(s) = \beta, \qquad s > 0. \tag{23}$$

The common step from a cubic sign to attracting coexistence is Lemma A.14. An analytic amplitude curve that initially rises above β_c and tends to zero has a positive interior maximum. Levels just below that maximum yield a smaller repelling cycle and a larger attracting one, while zero stays stable. At an intersection, the identity $g'(a) = -1 - 2a\mathcal{A}'(a)/\beta$ distinguishes the two branches. This argument is shared by the loss-class and DPO constructions; transverse attraction, reward gaps, and IPO convergence are separate model-specific checks. The full DPO conclusion is Theorem B.2: every negative-cubic depth admits such coexistence, without a nondegenerate or unique maximum assumption.

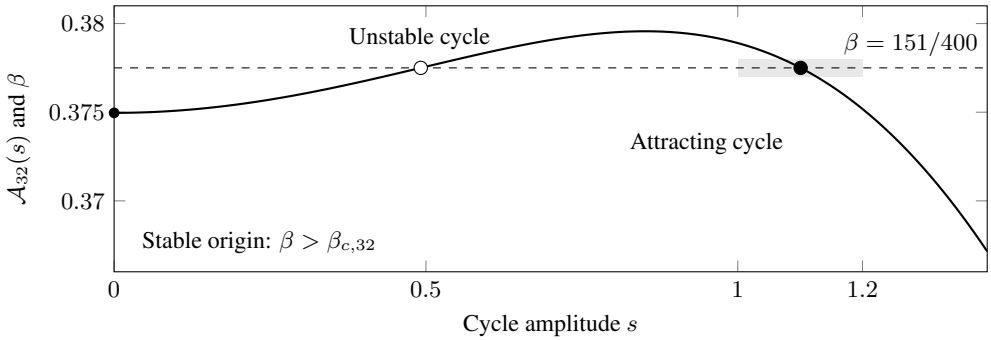


Figure 1: The rising intersection is unstable, the descending intersection attracts, and zero is stable. Shading marks the certified amplitude and scale region shared by depths 31–34. The curve evaluates the scalar recurrence numerically; intersection and stability claims are certified in Appendix D.32.

A common certified interval. Corollary B.3 proves simultaneous coexistence for all depths $K \in \{31, 32, 33, 34\}$ and scales $0.377 < \beta < 0.378$. Every attracting cycle has amplitude $1 < a < 6/5$ and reward gap exceeding $\gamma/10$.

The interval has width $1/1000$. It is a certified common subset, not the full coexistence region or evidence that the phenomenon is widespread. The certificate bounds derivatives over complete subintervals of $[1, 6/5]$; checking values at a numerical grid would not establish this claim.

3.5 How a coexistence window opens

The preceding existence result need not quantify a global maximum. Near the cubic transition, a fifth-order expansion instead identifies a specific local fold, where attracting and saddle cycles meet. Keep the outer update discrete and run $\dot{w} = -\nabla L_{\mathrm{D}}(w;\theta)$, $w(0) = 0$, for time τ; write $\mathsf{F}_{\mathrm{D},\tau}(\theta) = \theta + w(\tau)/\beta$. On the invariant axis, at the critical scale,

$$\mathsf{F}_{\mathrm{D},\tau,x}(s,0) = -s + \chi(\tau)s^3 + \psi(\tau)s^5 + O(s^7), \qquad \beta_c(\tau) = \frac{3}{8}(1 - e^{-\tau/4}). \tag{24}$$

Here χ is (95). The rational certificate proves

$$16.83612 < \tau_* < 16.83613, \qquad 0.01268 < \kappa := -\chi'(\tau_*) < 0.01269,$$
$$0.08681 < \psi_* := \psi(\tau_*) < 0.08683. \tag{25}$$

The five-decimal bounds are exact rational inequalities.

Theorem 3.3 (Width, reward, and basin at the local cycle fold)**.** *In the nine-response model, set $\varepsilon = \tau - \tau_* > 0$. For all sufficiently small ε, a local fold of two-cycles has amplitude a_{f} and scale β_{f} satisfying*

$$a_{\mathrm{f}}^2 = \frac{\kappa}{2\psi_*}\varepsilon + O(\varepsilon^2), \qquad \beta_{\mathrm{f}}(\tau) - \beta_c(\tau) = \frac{\beta_c(\tau_*)\kappa^2}{8\psi_*}\varepsilon^2 + O(\varepsilon^3). \tag{26}$$

For fixed $\rho \in (0,1)$, take $\beta = \beta_c + \rho(\beta_{\mathrm{f}} - \beta_c)$. Zero is stable and two cycle amplitudes obey

$$a_\pm^2 = \frac{\kappa}{2\psi_*}(1 \pm \sqrt{1-\rho})\varepsilon + O(\varepsilon^2). \tag{27}$$

The larger cycle attracts an open set in $\mathbb{R}^2$; the smaller is a saddle. Their reward gaps are

$$\Delta_\pm := J_* - J(a_\pm, 0) = \frac{\gamma\kappa}{16\psi_*}(1 \pm \sqrt{1-\rho})\varepsilon + O(\varepsilon^2). \tag{28}$$

The connected component containing zero of its basin on the invariant axis is exactly $(-a_-, a_-)$. Matched IPO flow converges globally at these same times and scales. All branch remainders are uniform for ρ in compact subsets of $(0,1)$.

The certified width coefficient in (26) lies between 0.0000855 and 0.0000857. The exponents follow from the cubic–quintic normal form; the DPO contribution is its verified realization and reward and basin consequences. This local fold need not be the global amplitude maximum. No explicit upper bound on admissible ε is asserted.

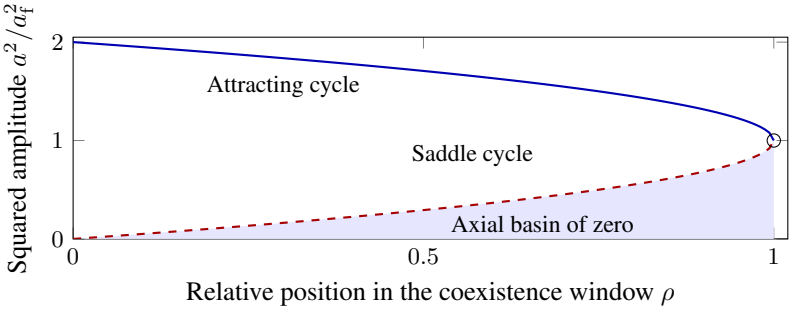


Figure 2: The limiting rescaled local branches as $\varepsilon \downarrow 0$. Their squared amplitudes tend to $1 \pm \sqrt{1-\rho}$; they meet at the fold, marked by an open circle. Below the saddle branch lies the connected basin of zero on the invariant axis, expressed in squared coordinates. The horizontal unit interval represents a physical scale width of order ε^2. These are analytic limiting curves, not simulated trajectories.

Finite GD must resolve the stability margin. The attracting cycle's longitudinal stability margin is of order ε^2, so fixed-horizon consistency alone does not settle the joint limit. Proposition B.12 takes $\varepsilon = K\eta - \tau_* \downarrow 0$ and shows that the exact GD critical scale satisfies

$$\beta_c^{\eta,K} - \beta_c(K\eta) = B_*\eta + O(\eta\varepsilon + \eta^2), \qquad B_* = 3\tau_* e^{-\tau_*/4}/256 > 0. \tag{29}$$

If the scale is kept at the flow's relative position ρ, $\eta = o(\varepsilon^2)$ preserves the leading cycle and reward constants. More sharply, if $\eta/\varepsilon^2 \to h$, coexistence persists for $h < \rho C_*/B_*$ with effective position $\rho - B_* h/C_*$, whereas zero is unstable for $h > \rho C_*/B_*$. Here $C_* = \beta_c(\tau_*)\kappa^2/(8\psi_*)$ and the certified ratio is $0.0291 < C_*/B_* < 0.0293$. The critical-scale shift is shared by matched DPO and IPO.

If instead the scale is centred at the exact discrete value $\beta_c^{\eta,K} = 3\{1-(1-\eta/4)^K\}/8$, the weaker condition $\eta = o(\varepsilon)$ preserves the fold scalings and attraction. Matching the linear critical scale removes the leading discretization error; uniform higher derivatives then control the remaining nonlinear error. These guarantees resolve the joint limit, while an explicit admissible ε range remains open.

3.6 Fit deeply, refresh conservatively

Can the same failure model support a positive guarantee without evaluating the true reward? Keep the full DPO inner computation, but limit how much of its output becomes the next policy. The comparison vector

$$z_\theta = 4M_\theta^{-1} b_\theta = w_{\mathrm{Q}}^*(\theta)$$

uses only comparison-feature moments and preference probabilities, since $b_\theta = \mathbb{E}[d(A - 1/2)]$. It does not require reward values or a reward gradient. For the DPO output $w = w_{\mathrm{D},K}(\theta)$, define

$$a(\theta, w) = \min\left\{1, \min_{i: w_i \neq 0} \frac{|z_{\theta,i}|}{|w_i|}\right\}, \qquad \theta^+ = \theta + \frac{a(\theta,w)}{\beta} w, \tag{30}$$

with $a(\theta, 0) = 1$. The same definition applies to the exact DPO minimizer. This is a single scalar damping of the DPO displacement, not a coordinatewise replacement by IPO. Both the next reference and the next sampling law use p_{θ^+}.

Theorem 3.4 (Depth-independent safeguarded convergence)**.** *In model (6), take $\eta = 1$ and $\beta > 3/8 + 1/7200$, including $\beta = \beta_{32}$. At every round choose any $K_n \in \{1, 2, \ldots, \infty\}$, where infinity means exact inner DPO minimization, and apply (30). For every finite initialization, reward increases strictly at each nonstationary round and $\theta_n \to 0$. More precisely, for each compact coordinate rectangle C centered at zero, there is an explicit $\rho_C \in (0,1)$, independent of all depth choices, such that*

$$\|\theta_n\|_1 \leq (1-\rho_C)^n \|\theta_0\|_1 \qquad (\theta_0 \in C). \tag{31}$$

The depths may depend on the entire past and need not be bounded. The adopted displacement also strictly lowers the frozen DPO loss outside zero.

Exact IPO already converges globally here. The safeguard is an intervention on the failure mechanism, without a computational or reward superiority claim. Proposition C.1 compares it with sufficient constant damping: the envelope chooses the largest fraction satisfying its current coordinate caps, requires no initial rectangle as input, and accepts one-step DPO in full. The proposition also gives the sharp axial condition for strict absolute-coordinate decrease under constant damping.

The proof supplies sign preservation and a depth-independent lower bound on accepted motion, ruling out stagnation. It uses population moments and product structure; arbitrary features and empirical estimates remain outside its scope.

Which parts require product structure? The finite-family cubic identity needs no product decomposition. The local part of Theorem 3.2 requires an invariant scalar direction and reflection symmetry. The scalar coexistence argument then uses analyticity, an initial rise of $\mathcal{A}_K$, and its decay at infinity, rather than independence itself. In the constructed two-dimensional model, coordinate reflections provide the invariant line and a diagonal cycle Jacobian; the independent second factor supplies the explicit transverse bound. For the safeguard, independence of the two comparison-feature differences makes the centered logistic loss even in each inner coordinate, yielding sign preservation, while additive coordinate rewards yield the reward Lyapunov function. These are distinct uses; a scalar sign reversal alone does not prove transverse attraction or a reward guarantee.

Extensions and limits of the certificate. The complete statements are collected in Appendix B. The coexistence construction survives inner gradient flow for $\tau > \tau_* \in (16, 17)$, exact minimization, and arbitrarily fine GD discretization at a fixed coexistence scale (Theorem B.10 and Corollary B.11). Small feature, weight, and reward perturbations preserve the cycle and a positive reward difference from IPO (Theorem B.13); exact global optimality requires additional structure (Corollary C.2). Unrestricted perturbation and full-dimensional basin radii remain qualitative. For each horizon T, using batch size N_T at every round of that run with $N_T/\log T \to \infty$ preserves the depth-32 reward gap in probability and expectation (Theorem B.14). This is a triangular-array limit, not a statement about increasing batches along one infinite run.

4 Discussion

Stable inner optimization, matched second-order geometry, and a stable reward-optimal fixed point do not jointly certify reward consistency. Higher loss derivatives matter through repeated inner fitting and outer adoption. Near IPO, logarithmic depth growth and a shrinking local window quantify this limitation. Near the DPO fold, finite GD must resolve a vanishing stability margin; calibrating its linear critical scale relaxes that requirement. Removing inner optimization error can retain the failure.

For the nine-response onset, the step-size conditions have a computational interpretation at a prescribed inner duration. Since $K\eta \to \tau_* > 0$, sufficient resolution rates are $K\varepsilon^2 \to \infty$ at the flow scale and $K\varepsilon \to \infty$ after recentering. These rates concern approximation of the certified branches, not the time required to reach them from an arbitrary initialization. The attracting cycle's longitudinal second-iterate multiplier remains $1 - \Theta(\varepsilon^2)$ at a fixed interior relative scale. Finer inner discretization therefore does not remove the linearized outer relaxation time of order ε^{-2}. Accurately approximating a cycle and approaching it rapidly are separate requirements.

The guarantees remain structural. The sign criterion needs a symmetric invariant direction; attracting coexistence also requires global amplitude control and, in higher dimensions, transverse stability. The explicit model regions do not classify arbitrary feature distributions. Unrestricted model-perturbation radii, full-dimensional basin radii, and admissible onset neighborhoods remain qualitative. General perturbations preserve local reward separation but need not preserve exact global optimality. Our target is an attainable optimum in a restricted policy class under one correct BT law, with jointly refreshed reference and sampling policies. Fixed-reference DPO, neural features, Adam, and fresh inner minibatches require further analysis.

The numerical depths describe this normalization. Extending the safeguard to empirical estimates and general features remains open. Preference fitting and policy adoption need separate certificates.

AI USE STATEMENT

In this work, we used generative AI tools for spell and grammar check of English writing. We have not used generative AI tools for help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, and the rest of the required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools for identify relevant literature. We have reviewed all AI-assisted work. We checked all LLM-assisted survey results. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All distributions, parameters, update rules, theorem assumptions, and proofs are specified in the paper and appendices. Appendix F gives a self-contained Python program using exact rational arithmetic to certify the finite inequalities in the constructions. The complete program and the figure coordinates are embedded in the single LaTeX source; references are supplied in the accompanying BibTeX file. Decimal values are descriptive; the proofs use the stated exact definitions and rational bounds.

REFERENCES

Dankmar Böhning and Bruce G. Lindsay. Monotonicity of quadratic-approximation algorithms. *Annals of the Institute of Statistical Mathematics*, 40(4):641–663, 1988. URL https://www.ism.ac.jp/editsec/aism/pdf/040_4_0641.pdf.

Daniele Calandriello, Zhaohan Daniel Guo, Rémi Munos, Mark Rowland, Yunhao Tang, Bernardo Avila Pires, Pierre Harvey Richemond, Charline Le Lan, Michal Valko, Tianqi Liu, Rishabh Joshi, Zeyu Zheng, and Bilal Piot. Human Alignment of Large Language Models through Online Preference Optimisation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 5409–5435, 2024. URL https://proceedings.mlr.press/v235/calandriello24a.html.

Yurong Chen, Yu He, Michael I. Jordan, and Fan Yao. How Sampling Shapes LLM Alignment: From One-Shot Optima to Iterative Dynamics. arXiv:2602.12180, 2026. URL https://arxiv.org/abs/2602.12180v1. Version 1.

Etienne Gauthier, Francis Bach, and Michael I. Jordan. Explaining and Preventing Alignment Collapse in Iterative RLHF. arXiv:2605.04266, 2026. URL https://arxiv.org/abs/2605.04266v1. Version 1.

Mohammad Gheshlaghi Azar, Zhaohan Daniel Guo, Bilal Piot, Rémi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. A general theoretical paradigm to understand learning from human preferences. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 4447–4455, 2024. URL https://proceedings.mlr.press/v238/gheshlaghi-azar24a.html.

Aditya Gopalan, Sayak Ray Chowdhury, and Debangshu Banerjee. Why DPO is a Misspecified Estimator and How to Fix It. arXiv:2510.20413, 2025. URL https://arxiv.org/abs/2510.20413v1. Version 1.

Alexey Gorbatovski, Boris Shaposhnikov, Alexey Malakhov, Nikita Surnachev, Yaroslav Aksenov, Ian Maksimov, Nikita Balagansky, and Daniil Gavrilov. Learn your reference model for real good alignment. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/94d13c2401fe119e57ba325b6fe526e0-Paper-Conference.pdf.

Jiafan He, Huizhuo Yuan, and Quanquan Gu. Accelerated Preference Optimization for Large Language Model Alignment. arXiv:2410.06293, 2024. URL https://arxiv.org/abs/2410.06293v1. Version 1.

Audrey Huang, Wenhao Zhan, Tengyang Xie, Jason D. Lee, Wen Sun, Akshay Krishnamurthy, and Dylan J. Foster. Correcting the Mythos of KL-Regularization: Direct Alignment without Overoptimization via Chi-Squared Preference Optimization. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e785d64baa2cde06c69d9b7a14636ef2-Abstract-Conference.html.

Lemin Kong, Xiangkun Hu, Tong He, and David Wipf. Common learning constraints alter interpretations of direct preference optimization. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 2215–2223, 2025. URL https://proceedings.mlr.press/v258/kong25a.html.

Yixin Liu, Argyris Oikonomou, Weiqiang Zheng, Yang Cai, and Arman Cohan. COMAL: A Convergent Meta-Algorithm for Aligning LLMs with General Preferences. arXiv:2410.23223, 2024. URL https://arxiv.org/abs/2410.23223v2. Version 2, revised October 2025.

Si Yi Meng, Antonio Orvieto, Daniel Yiming Cao, and Christopher De Sa. Gradient Descent on Logistic Regression with Non-Separable Data and Large Step Sizes. arXiv:2406.05033, 2024. URL https://arxiv.org/abs/2406.05033v3. Version 3, revised September 2026.

Rémi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, Andrea Michi, Marco Selvi, Sertan Girgin, Nikola Momchev, Olivier Bachem, Daniel J. Mankowitz, Doina Precup, and Bilal Piot. Nash Learning from Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 36743–36768, 2024. URL https://proceedings.mlr.press/v235/munos24a.html.

Andi Nika, Debmalya Mandal, Parameswaran Kamalaruban, Georgios Tzannetos, Goran Radanovic, and Adish Singla. Reward Model Learning vs. Direct Policy Optimization: A Comparative Analysis of Learning from Human Preferences. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 38145–38186, 2024. URL https://proceedings.mlr.press/v235/nika24a.html.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7599–7609, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/a85b405ed65c6477a4fe8302b5e06ce7-Paper-Conference.pdf.

Rafael Rafailov, Yaswanth Chittepu, Ryan Park, Harshit Sikchi, Joey Hejna, W. Bradley Knox, Chelsea Finn, and Scott Niekum. Scaling Laws for Reward Model Overoptimization in Direct Alignment Algorithms. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/e45caa3d5273d105b8d045e748636957-Paper-Conference.pdf.

Ruizhe Shi, Minhak Song, Runlong Zhou, Zihan Zhang, Maryam Fazel, and Simon S. Du. Understanding the Performance Gap in Preference Learning: A Dichotomy of RLHF and DPO. arXiv:2505.19770, 2025a. URL https://arxiv.org/abs/2505.19770v5. Version 5, revised May 2026.

Ruizhe Shi, Runlong Zhou, and Simon S. Du. The Crucial Role of Samplers in Online Direct Preference Optimization. In *International Conference on Learning Representations*, 2025b. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/62133b95efaa6f10a5395eb10ca087cd-Abstract-Conference.html.

Yunhao Tang, Zhaohan Daniel Guo, Zeyu Zheng, Daniele Calandriello, Rémi Munos, Mark Rowland, Pierre Harvey Richemond, Michal Valko, Bernardo Avila Pires, and Bilal Piot. Generalized Preference Optimization: A Unified Approach to Offline Alignment. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 47725–47742, 2024. URL https://proceedings.mlr.press/v235/tang24b.html.

Mingzhi Wang, Chengdong Ma, Qizhi Chen, Linjian Meng, Yang Han, Jiancong Xiao, Zhaowei Zhang, Jing Huo, Weijie J. Su, and Yaodong Yang. Magnetic Preference Optimization: Achieving Last-iterate Convergence for Language Model Alignment. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/5833b4daf5b076dd1cdb362b163dff0c-Abstract-Conference.html.

Wei Xiong, Hanze Dong, Chenlu Ye, Ziqi Wang, Han Zhong, Heng Ji, Nan Jiang, and Tong Zhang. Iterative Preference Learning from Human Feedback: Bridging Theory and Practice for RLHF under KL-constraint. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 54715–54754, 2024. URL https://proceedings.mlr.press/v235/xiong24a.html.

The supporting lemmas are in Appendix A. Appendices B and C state all additional results before their proofs in Appendix D. Further calculations and executable rational certificates are in Appendices E and F, respectively.

A LEMMAS

This section collects the facts used in the main theorems. All vector norms are Euclidean, all matrix norms are induced operator norms, and every $O(\cdot)$ statement is local at the point specified, with K, η, β fixed unless a limit is explicitly taken. Empty sums are zero and empty products are one. To avoid confusion between coordinates and iteration indices, (s, t) denotes a policy parameter and n denotes an outer round. The proofs are given in Appendix D in dependency order.

Lemma A.1 (Curvature, existence, and frozen GD)**.** *For the finite minimal family (1),*

$$\nabla L_{\mathrm{D}}(w;\theta) = \mathbb{E}_\theta[d\{\sigma(d^\top w) - q_{IJ}\}], \tag{32}$$

$$\nabla^2 L_{\mathrm{D}}(w;\theta) = \mathbb{E}_\theta[dd^\top \sigma'(d^\top w)], \qquad \nabla L_{\mathrm{Q}}(w;\theta) = -b_\theta + \tfrac{1}{4} M_\theta w. \tag{33}$$

Moreover $M_\theta = 2\nabla^2\psi(\theta) = 2\mathcal{I}(\theta) \succ 0$, $0 \prec \nabla^2 L_{\mathrm{D}} \preceq M_\theta/4$, and both losses have unique finite minimizers. In the model (6),

$$M_{s,t} = \operatorname{diag}\left(\operatorname{sech}^2(s/2), \frac{4(\cosh t + 2)}{(1 + 2\cosh t)^2}\right) \preceq \frac{4}{3}\mathrm{Id}. \tag{34}$$

Consequently, for either loss, $0 < \eta \le 3$ gives

$$L_X(w - \eta\nabla L_X(w;\theta);\theta) \le L_X(w;\theta) - \eta(1 - \eta/6)\left\|\nabla L_X(w;\theta)\right\|^2, \tag{35}$$

and the frozen GD sequence converges to its unique minimizer from every initial $w \in \mathbb{R}^2$.

Lemma A.2 (Calibration and the IPO solution)**.** *Equation (7) has exactly one positive solution. Define*

$$c_x(t) = \mathbb{E}_{z,z'\sim p_t} \tanh\frac{\gamma - \delta(z^2 - z'^2)}{2}, \tag{36}$$

$$c_z(s) = \mathbb{E}_{x,x'\sim p_s} \tanh\frac{\delta - \gamma(x^2 - x'^2)}{2}, \tag{37}$$

with independent draws in each expectation. Then

$$c_x(0) = c = 3/4, \quad 0 < c_x(t) \le \tanh(\gamma/2), \quad 0 < c_z(s) \le \tanh(\delta/2) < 1/40, \tag{38}$$

$$\tfrac{1}{2}\tanh(\gamma/2) \le \tfrac{3}{8} + \tfrac{1}{7200}. \tag{39}$$

The unique IPO minimizer is

$$w_{\mathrm{Q}}^* = 4M_\theta^{-1} b_\theta = \left(-2c_x(t)\tanh(s/2),\ -2c_z(s)\frac{\sinh t}{\cosh t + 2}\right). \tag{40}$$

For $\eta = 1$ and every fixed finite $K \ge 1$,

$$w_{\mathrm{Q},K} = \{\mathrm{Id} - (\mathrm{Id} - M_\theta/4)^K\} w_{\mathrm{Q}}^*. \tag{41}$$

If $\beta > 3/8 + 1/7200$, $F_{\mathrm{Q},K}$ converges globally to zero and strictly increases J outside zero. The same holds for the exact-IPO outer map $\theta \mapsto \theta + w_{\mathrm{Q}}^*/\beta$.

Lemma A.3 (Symmetry and the scalar DPO recursion)**.** *All the population maps commute with reflection of either coordinate. On $t = 0$, the inner iterates satisfy $w_{\mathrm{D},j,z} = 0$. For $\eta = 1$, write $\nu_j(s) = -w_{\mathrm{D},j,x}(s, 0)$. Then*

$$\nu_0(s) = 0, \qquad \nu_{j+1}(s) = \nu_j(s) + f(s, \nu_j(s)), \tag{42}$$

$$f(s, v) = \frac{\frac{3}{2}\sinh s - 2\cosh s \tanh(v/2) - \tanh v}{2(1 + \cosh s)^2}. \tag{43}$$

For $a = \log 3$, $v_j = \nu_j(a)$ is (9), and $0 = v_0 < v_1 < \cdots < v_j < v_\infty < 1$, where $f(a, v_\infty) = 0$. If $u_j = \nu_j'(a)$, $A_j = \tanh(v_j/2)$ and $B_j = \tanh v_j$, then

$$u_0 = 0, \quad u_{j+1} = \left(\frac{13}{16} + \frac{15}{128}A_j^2 + \frac{9}{128}B_j^2\right)u_j + \frac{9}{256} + \frac{3}{64}A_j + \frac{9}{128}B_j. \tag{44}$$

The exact rational bounds are

$$\frac{151}{400} < \beta_{32} < \frac{236}{625}, \qquad -\frac{179}{200} < 1 - \frac{u_{32}}{\beta_{32}} < -\frac{893}{1000}. \tag{45}$$

Lemma A.4 (Transverse attraction). *At $(s, 0)$ let $R_x = -\gamma(x^2 - x'^2)$ and define*

$$\mathcal{N}(s) = \frac{2}{9}\mathbb{E}_{x,x'}[\sigma(R_x + \delta) - \sigma(R_x - \delta)] \in (0, \delta/9]. \tag{46}$$

For the DPO inner iterates at $(\log 3, 0)$ put

$$H_j = \frac{4}{3}\mathbb{E}_{x,x'}[\sigma'(v_j(x - x'))]. \tag{47}$$

The transverse derivative $U_j = \partial_t w_{\mathrm{D},j,z}(\log 3, 0)$ satisfies

$$U_0 = 0, \qquad U_{j+1} = (1 - H_j)U_j - \mathcal{N}(\log 3), \qquad \frac{2}{15} < H_j \le \frac{1}{3}. \tag{48}$$

In particular,

$$0 < \kappa_{32} := -U_{32} = \mathcal{N}(\log 3) \sum_{j=0}^{31} \prod_{\ell=j+1}^{31} (1 - H_\ell) \le \frac{5\delta}{6}. \tag{49}$$

The Jacobian of $F_{\mathrm{D},32}$ at either $a_\pm$ is diagonal, and its second diagonal entry is $1 - \kappa_{32}/\beta_{32} \in (8/9, 1)$.

Lemma A.5 (Empirical map error and concentration). *In the nine-response model with $\eta = 1$, let $\left\|\widehat{P} - P_\theta\right\|_1 \le \epsilon$ for one labeled-pair batch. Then (108) holds for all $K \ge 1$ without requiring $\widehat{M}$ to be invertible. If $\epsilon \le 1/12$, both empirical inner losses satisfy*

$$\widehat{L}_X(w - \nabla\widehat{L}_X(w)) \le \widehat{L}_X(w) - \frac{3}{4}\left\|\nabla\widehat{L}_X(w)\right\|^2. \tag{50}$$

For N independent draws from any distribution P on S states,

$$\mathbb{P}\{\left\|\widehat{P} - P\right\|_1 > \epsilon\} \le 2^S e^{-N\epsilon^2/2}. \tag{51}$$

The same bound holds conditional on the past when the current distribution is chosen from that past.

Lemma A.6 (A common neighborhood for noisy tracking). *Let $G, H_1, H_2 : \mathbb{R}^m \to \mathbb{R}^m$ be C^1 maps. Suppose $G(a_+) = a_-$, $G(a_-) = a_+$, $a_+ \ne a_-$, and $\|DG(a_\pm)\| < 1$. Suppose $H_i(0) = 0$, $\|DH_i(0)\| < 1$ and $H_i^n(a_+) \to 0$ for $i = 1, 2$. Then there exist an open neighborhood U of a_+, $\xi_0 > 0$, $q \in (0, 1)$, $C < \infty$, and $n_0 \ge 1$ such that any three recursions starting at the same $\theta_0 \in U$ and having additive per-round errors of norm at most $\xi \le \xi_0$ satisfy*

$$\left\|\widehat{\theta}_{G,n} - a_{(-1)^n}\right\| \le q^n \|\theta_0 - a_+\| + \frac{\xi}{1 - q}, \qquad n \ge 0, \tag{52}$$

$$\left\|\widehat{\theta}_{H_i,n}\right\| \le Cq^{n-n_0} + \frac{\xi}{1 - q}, \qquad n \ge n_0,\ i = 1, 2. \tag{53}$$

Here $a_{(+1)} = a_+$ and $a_{(-1)} = a_-$. The errors may be arbitrary functions of the preceding iterates. The constants can be chosen uniformly over $\theta_0 \in U$.

Lemma A.7 (Bounded updates and the amplitude curve). *Fix $0 < \eta \le 3$ and finite $K \ge 1$. Each ν_j is odd and real analytic. For every $s > 0$, $f(s, \cdot)$ has a unique root $\nu_\infty(s) \in (0, \log 7)$ and*

$$0 = \nu_0(s) < \nu_1(s) < \cdots < \nu_K(s) < \nu_\infty(s) < \log 7. \tag{54}$$

The amplitude curve extends analytically and evenly at zero, with

$$\mathcal{A}_K(s) = \beta_{c,K} - \tfrac{1}{2}\beta_{c,K}\chi_{\mathrm{D},K}s^2 + O(s^4), \qquad \lim_{s\to\infty} \mathcal{A}_K(s) = 0. \tag{55}$$

Lemma A.8 (Uniform transverse control). *At $(s, 0)$ put $H_j(s) = \frac{4}{3}\mathbb{E}_s\sigma'(\nu_j(s)(x - x'))$ and $\kappa_j(s) = -\partial_t w_{\mathrm{D},j,z}(s, 0)$. For $0 < \eta \le 3$,*

$$\tfrac{1}{9} \le H_j(s) \le \tfrac{1}{3}, \qquad \mathcal{N}(s) = 2c_z(s)/9, \tag{56}$$

$$\kappa_0 = 0, \qquad \kappa_{j+1} = (1 - \eta H_j)\kappa_j + \eta\mathcal{N}(s), \tag{57}$$

$$0 < \kappa_K(s) \le 9\mathcal{N}(s)\{1 - (1 - \eta/9)^K\} < \delta. \tag{58}$$

The axis Jacobian is

$$DF_{\mathrm{D},K}(s, 0) = \mathrm{diag}(g'_{K,\beta}(s), 1 - \kappa_K(s)/\beta). \tag{59}$$

For $\beta > \delta$, its second entry belongs to $(0, 1)$ for every finite s.

Lemma A.9 (Global IPO convergence above the critical scale)**.** *In (6)–(7), for $0 < \eta \le 3$ and finite $L \ge 1$, $\beta > \beta_{c,L} = c\{1-(1-\eta/4)^L\}/2$ implies global convergence of $F_{\mathrm{Q},L}$ to zero from every finite initialization. Its reward tends to J_*. One-step DPO has the same conclusion for $\beta > \beta_{c,1}$.*

Lemma A.10 (Inner flow, exact minimizers, and local coefficients)**.** *For (6), the flows (98) exist globally, converge to their unique minimizers, and depend real analytically on θ at each fixed finite time. Exact minimizers are real analytic at every finite θ. On $t = 0$, put $\nu_\tau(s) = -w_{\mathrm{D},x}(\tau;(s,0))$ and $\nu_\infty(s) = -w^*_{\mathrm{D},x}(s,0)$. For $s > 0$,*

$$\partial_u \nu_u(s) = f(s,\nu_u(s)), \qquad 0 < \nu_u(s) < \nu_v(s) < \nu_\infty(s) < \log 7 \quad (0 < u < v < \infty). \tag{60}$$

For $\mathcal{A}_\tau(s) = \nu_\tau(s)/(2s)$, the analytic extension at zero is

$$\mathcal{A}_\tau(s) = \beta_{c,\tau} - \tfrac{1}{2}\beta_{c,\tau}\chi_{\mathrm{D}}^{\mathrm{flow}}(\tau)s^2 + O(s^4), \qquad \mathcal{A}_\tau(s) \longrightarrow 0 \quad (s \to \infty). \tag{61}$$

For finite τ, the coefficient is (95); for $\tau = \infty$, $\chi_{\mathrm{D}}^{\mathrm{flow}}(\infty) = -13/192$ and $\nu_\infty(s) = cs + (13/512)s^3 + O(s^5)$. The transverse axis derivative of the outer map is $1 - \kappa_\tau(s)/\beta$, where

$$0 < \kappa_\tau(s) < \delta \quad (0 < \tau \le \infty). \tag{62}$$

For each fixed $0 < \tau \le \infty$, the DPO and IPO outer maps have identical Taylor expansions through order two at zero.

Lemma A.11 (C^1 consistency of the inner GD solver)**.** *Fix $0 < \tau < \infty$, $\beta > 0$, and a compact outer-parameter set $\Theta \subset \mathbb{R}^2$. For sufficiently small $\eta > 0$ and positive integers K with $|K\eta - \tau| \le 1$, there is a constant independent of η, K such that*

$$\sup_{\theta\in\Theta}\left\{\|F^{(\eta)}_{\mathrm{D},K}(\theta) - \mathsf{F}_{\mathrm{D},\tau}(\theta)\| + \|DF^{(\eta)}_{\mathrm{D},K}(\theta) - D\mathsf{F}_{\mathrm{D},\tau}(\theta)\|\right\} \le C_\Theta\{\eta + |K\eta - \tau|\}. \tag{63}$$

The superscript displays the GD step size.

Lemma A.12 (Feature moments and scalar curvature)**.** *Under the scalar restriction of Theorem B.1, suppose the centered feature $X = e^\top\phi(Y) - \mathbb{E}_0[e^\top\phi(Y)]$ has a reflection-symmetric distribution at θ_*. Set $\mu_2 = \mathbb{E}_0 X^2 > 0$ and $\mu_4 = \mathbb{E}_0 X^4$. For independent draws, $d_e = X - X'$, and*

$$H = \frac{\mu_2}{2}, \qquad H_2 = \frac{\mu_4 - 3\mu_2^2}{4}, \qquad Q_4 = 2\mu_4 + 6\mu_2^2. \tag{64}$$

With $\kappa = \mu_4/\mu_2^2$, condition (74) is equivalent to (76).

Lemma A.13 (The fifth coefficient and a uniform analytic remainder)**.** *For the nine-response scalar DPO flow, put $x = \tau/4$ and $E = e^{-x}$. Then, locally and uniformly for τ in a neighborhood of τ_*,*

$$\nu_\tau(s) = a(x)s + b(x)s^3 + d(x)s^5 + O(s^7), \tag{65}$$

where

$$a(x) = \frac{3}{4}(1-E), \qquad b(x) = \frac{26 + (199 - 462x)E - 270E^2 + 45E^3}{1024}, \tag{66}$$

$$d(x) = \frac{1}{2621440}\sum_{k=0}^{5} P_k(x)E^k, \tag{67}$$

and

$$\begin{aligned} P_0 &= -38840, & P_1 &= -227929 + 838380x - 355740x^2,\\ P_2 &= 636120 - 831600x, & P_3 &= -529650 + 207900x,\\ P_4 &= 181440, & P_5 &= -21141. \end{aligned}$$

Consequently $\chi(\tau) = -2b(x)/a(x)$ and $\psi(\tau) = -2d(x)/a(x)$. The rational inequalities (3.5) hold, as does

$$\frac{855}{10^7} < \frac{\beta_c(\tau_*)\kappa^2}{8\psi_*} < \frac{857}{10^7}. \tag{68}$$

There are $r, h, C > 0$ such that the even analytic extension $\mathcal{H}(z,\tau) = \mathcal{A}_\tau(\sqrt{z})$ satisfies

$$\mathcal{H}(z,\tau) = \beta_c(\tau)\left[1 - \frac{\chi(\tau)}{2}z - \frac{\psi(\tau)}{2}z^2 + z^3R(z,\tau)\right], \tag{69}$$

for $|z| \le r$, $|\tau - \tau_*| \le h$, with $\max_{i+j\le 2}|\partial_z^i\partial_\tau^j R(z,\tau)| \le C$. The notation at negative z means this analytic extension, not a real square root.

Lemma A.14 (A common amplitude-curve criterion)**.** *Let $\nu : \mathbb{R} \to \mathbb{R}$ be odd and real analytic, with $\nu(s) > 0$ for $s > 0$, $\nu'(0) = 2\beta_c > 0$, and $\nu(s)/s \to 0$ as $s \to \infty$. Extend $\mathcal{A}(s) = \nu(s)/(2s)$ analytically to zero and assume $\mathcal{A}''(0) > 0$. Then $m = \max_{s \geq 0} \mathcal{A}(s) > \beta_c$ has a positive finite maximizer a_*. There is $e > 0$ such that every $\beta \in (m - e, m)$ satisfies $\beta > \beta_c$ and the odd map $g_\beta(s) = s - \nu(s)/\beta$ has a unique, locally stable fixed point at zero and two cycles with amplitudes $a_- < a_* < a_+$. The smaller cycle is repelling and the larger attracting on $\mathbb{R}$. If $\mathcal{A}''(a_*) < 0$, their collision is a nondegenerate local fold. At a cycle amplitude,*

$$g'_\beta(a) = -1 - 2a\mathcal{A}'(a)/\beta. \tag{70}$$

If a C^1 two-dimensional extension commutes with both coordinate reflections, restricts to g_β on the first axis, and has transverse derivative of modulus less than one at these points, the larger cycle attracts an open set in $\mathbb{R}^2$ and the smaller is a saddle. No nondegeneracy or uniqueness of the maximizer is required.

Lemma A.15 (Uniform smooth control of the inner solvers)**.** *The following bounds hold on a sufficiently small fixed interval $|s| \leq s_0$, with $\|\cdot\|_{C^9}$ denoting the maximum of the supremum norms of derivatives of orders zero through nine.*

1. *In the nine-response DPO model, fix $0 < T_0 < T_1 < \infty$. There are $C, \eta_0 > 0$ such that, for $0 < \eta \leq \eta_0$ and $K\eta \in [T_0, T_1]$,*

$$\|\nu_{\eta,K} - \nu_{K\eta}\|_{C^9} \leq C\eta. \tag{71}$$

   The inner displacements have uniformly bounded C^9 norms there. The full two-dimensional inner maps also have a C^1 error at most $C\eta$ on a fixed compact neighborhood of zero.

2. *In the three-response mixture model, fix $0 < c < 1$ and $\vartheta > 0$, put $\alpha = 1 + \vartheta\lambda$, and use unit GD. There are $\lambda_0, C > 0$ and $q \in (0, 1)$ such that for $0 \leq \lambda \leq \lambda_0$ and every $K \geq 0$,*

$$\|\nu_{\lambda,K} - \nu_{\lambda,\infty}\|_{C^9} \leq C(K + 1)^9 q^K. \tag{72}$$

   All these displacements have uniformly bounded C^9 norms, and the minimizer $\nu_{\lambda,\infty}(s)$ is smooth jointly in (s, λ).

For positive depths and times bounded away from zero, their even amplitude functions have uniform expansions $\mathcal{A}(s) = \beta_c[1 - \chi s^2/2 - \psi s^4/2] + O(s^6)$. In $z = s^2$, the remainder and its first two derivatives are respectively $O(z^3)$, $O(z^2)$, and $O(z)$, uniformly over the stated parameters.

B Additional Statements and Their Scope

B.1 The DPO criterion and its feature interpretation

This subsection retains the DPO specialization of Theorem 3.2, together with its detailed moment interpretation and the explicit model region at logistic weight one.

The general finite-family identity in Theorem B.4 gives the *leading possible cubic difference* between equal-depth DPO and IPO. It is a sum of comparison-feature fourth moments transported through the linearized inner dynamics; it can vanish, and does vanish at depth one. The following criterion turns that identity into a sign test applicable beyond the nine-response construction.

Consider a unit vector e such that, at outer parameters $\theta_* + se$, the inner gradients at we stay parallel to e. Suppose the restricted loss satisfies $L_X(-w; -s) = L_X(w; s)$, up to a term independent of w. This includes reflection-symmetric scalar exponential families; an independent product extension is sufficient but not required. For $d_e = e^\top d$, use the restricted local quantities in (13), with $A, H > 0$. These coefficients are derivatives of finite comparison expectations: $A = -b'_e(0)$, $B = b'''_e(0)/6$, and $H_2 = (\mathbb{E}_s d_e^2)''|_{s=0}/8$.

Theorem B.1 (A transferable criterion for a depth-induced sign reversal)**.** *Let $0 < \eta H < 1$, $r = 1 - \eta H$, and $d_K = 1 - r^K$. At the common longitudinal critical scale $\beta_{c,K} = Ad_K/(2H)$, the*

restricted maps are $-s + \chi_{X,K} s^3 + O(s^5)$*, where*

$$\chi_{\mathrm{Q},K} = \frac{2B}{A} + \frac{2H_2}{H}(1-u_K), \qquad \chi_{\mathrm{D},K} = \chi_{\mathrm{Q},K} - \Lambda W_K, \tag{73}$$
$$u_K = \frac{K(1-r)r^{K-1}}{d_K}, \qquad W_K = \frac{1-r}{d_K}\sum_{j=0}^{K-1} r^{K-1-j}(1-r^j)^3, \qquad \Lambda = \frac{Q_4 A^2}{24H^3}.$$

If

$$H_2 \le 0, \qquad 0 < \chi_{\mathrm{Q},\infty} := \frac{2B}{A} + \frac{2H_2}{H} < \Lambda, \tag{74}$$

then every IPO coefficient is positive, whereas the DPO coefficients strictly decrease from $2B/A > 0$ *to* $\chi_{\mathrm{Q},\infty} - \Lambda < 0$*. Thus they become negative at a finite depth.*

Condition (74) compares the limiting IPO cubic restoring term with the accumulated logistic fourth-moment correction. It is a local sufficient condition, not a general cycle theorem: finite-amplitude coexistence additionally needs global amplitude control and transverse stability. In our example, $(A, B, H, H_2, Q_4) = (3/16, 1/16, 1/4, -1/16, 5/2)$, so $\chi_{\mathrm{Q},\infty} = 1/6 < 15/64 = \Lambda$. The exact model-specific calculation in Theorem B.7 places the sign change between depths 15 and 16 at $\eta = 1$. This is a threshold for local criticality, not the first depth admitting a cycle or a failure threshold at one fixed β.

Reading the criterion from moments. The coefficient roles are described in Section 3; here we record the equivalent moment test. Center the symmetric scalar feature $X = e^\top \phi(Y)$ at the fixed point, let $\mu_j = \mathbb{E}_0 X^j$, and write $\kappa = \mu_4/\mu_2^2$ for its kurtosis. For independent comparisons, Lemma A.12 gives

$$H = \frac{\mu_2}{2}, \qquad H_2 = \frac{\mu_4 - 3\mu_2^2}{4}, \qquad Q_4 = 2\mu_4 + 6\mu_2^2. \tag{75}$$

Consequently, the sufficient condition becomes the dimensionless test

$$\kappa \le 3, \qquad 0 < \frac{2B}{A\mu_2} + \kappa - 3 < \frac{\kappa+3}{6}\left(\frac{A}{H}\right)^2. \tag{76}$$

The middle expression is IPO's remaining cubic restoring margin in feature-variance units; the right side is the logistic correction. Kurtosis alone does not determine the comparison, since it changes both H_2 and Q_4, while A, B depend on preferences. Statistical guarantees for estimating these local quantities are not asserted.

An explicit region of three-response models. To illustrate the criterion without the transverse coordinate, use the family (15) on $x \in \{-1, 0, 1\}$. Here α controls the central response's base weight and $c = \tanh(g/2) \in (0, 1)$ controls the strength of its BT preference advantage. Proposition B.5 proves that the criterion holds throughout the explicit parameter region

$$1 < \alpha \le 4, \qquad c^2 > \frac{(\alpha-1)(\alpha+2)^2}{\alpha^2(\alpha+8)}. \tag{77}$$

Figure 3 shows an open interval of preference strengths for every $1 < \alpha < 4$, with no reward-calibration equation. For example, $\alpha = 3$ and $g = \log 9$ give $c = 4/5$ and

$$\chi_{\mathrm{Q},\infty} = \frac{4}{15}, \qquad \Lambda = \frac{1056}{3125}, \qquad \chi_{\mathrm{D},\infty} = -\frac{668}{9375} < 0.$$

The displacement bound $|w_{\mathrm{D},K}(s)| < g$ supplies the global amplitude control needed by Lemma A.14. The resulting scale interval depends on the model and depth. Transverse attraction, the common numerical interval, and the safeguard are separate nine-response guarantees.

Why this feature and reward geometry matters. A linear policy log-ratio cannot favor zero over both endpoints, although the central response is preferred to each. The expected reward and central-response probability,

$$J_g(s) = -\frac{2g\cosh s}{\alpha + 2\cosh s}, \qquad p_s(0) = \frac{\alpha}{\alpha + 2\cosh s} \tag{78}$$

are uniquely maximized at $s = 0$. For $s > 0$, sampling favors positive endpoints and $b(s) = -2\alpha c \sinh s/(\alpha + 2\cosh s)^2 < 0$; this imbalance reverses after an overshoot to the reflected policy.

The roles of the parameters are explicit: $\Lambda \propto c^2$, whereas $\chi_{\mathrm{Q},\infty}$ is independent of c; $\kappa = (\alpha+2)/2$ makes $\alpha \le 4$ equivalent to $\kappa \le 3$; and $\alpha > 1$ keeps the limiting IPO coefficient positive. These are sufficient conditions, not a classification outside (77).

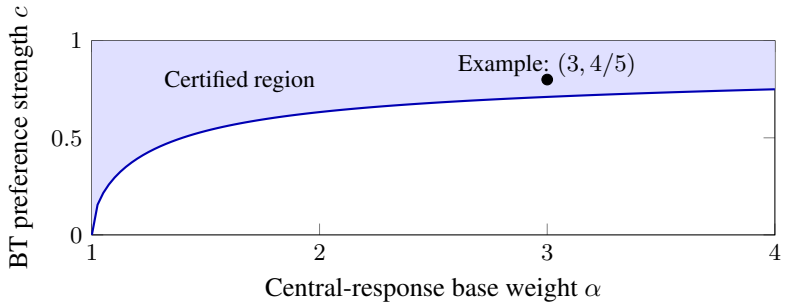


Figure 3: An explicit region of model parameters. Above the boundary $c = \sqrt{(\alpha-1)(\alpha+2)^2/[\alpha^2(\alpha+8)]}$, with $1 < \alpha \le 4$ and $c < 1$, Proposition B.5 guarantees a depth-induced sign reversal and, at every sufficiently large finite depth, an open scale interval of scalar attracting coexistence. The curve is the analytic boundary of the sufficient condition, not a numerically estimated failure boundary.

How to use the local test. First verify invariance and reflection symmetry; reward symmetry alone need not imply an invariant direction. Then compute the five local coefficients and apply Theorem B.1, or its equivalent form (76). Certifying an attracting cycle further requires global amplitude control, and a multidimensional conclusion requires transverse stability. These checks distinguish the reusable local diagnostic from a certificate of long-run failure.

B.2 The complete finite-depth coexistence conclusion

For $0 < \eta \le 3$, the scalar inner displacement obeys

$$\nu_0(s) = 0, \qquad \nu_{j+1}(s) = \nu_j(s) + \eta f(s, \nu_j(s)), \tag{79}$$

where f is (43). This is the finite-GD amplitude recurrence used by the following theorem.

Theorem B.2 (Coexistence from negative cubic criticality)**.** *In (6)–(7), fix $0 < \eta \le 3$ and K with $\chi_{\mathrm{D},K} < 0$. Then $\beta_{m,K} := \max_{s\ge 0} \mathcal{A}_K(s) > \beta_{c,K}$, with a positive finite maximizer $a_\star$. There is $\varepsilon_K \in (0, \beta_{m,K} - \beta_{c,K})$ such that, for every $\beta \in \mathcal{B}_K := (\beta_{m,K} - \varepsilon_K, \beta_{m,K})$:*

1. *Zero is the unique fixed point of $F_{\mathrm{D},K}$, is reward-optimal, and is locally asymptotically stable. The equal-depth DPO and IPO maps agree through second order there.*
2. *There are amplitudes $0 < r_-(\beta) < a_\star < r_+(\beta)$ defining two-cycles $\mathcal{C}_\pm(\beta) = \{(r_\pm(\beta), 0), (-r_\pm(\beta), 0)\}$. $\mathcal{C}_+$ attracts an open set in $\mathbb{R}^2$. At $\mathcal{C}_-$ the second-iterate derivative has one eigenvalue greater than one and one in $(0, 1)$.*
3. *Every $F_{\mathrm{Q},L}$ with fixed $1 \le L \le K$, and $F_{\mathrm{D},1}$, converges globally to zero at these same η, β.*
4. *The attracting cycle has reward gap*

$$\Delta_K(\beta) = \frac{\gamma}{2}\tanh^2(r_+(\beta)/2) \ge \frac{\gamma}{2}\tanh^2(a_\star/2) > 0. \tag{80}$$

In a frozen round initialized at either attracting cycle point, the last inner step strictly decreases both DPO loss and candidate-policy reward. Every frozen inner GD sequence nevertheless converges globally.

No nondegeneracy or uniqueness of the amplitude-curve maximum is assumed. The lower reward-gap bound is uniform over $\mathcal{B}_K$ for fixed (η, K); the scale interval may depend on both parameters.

B.3 A COMMON SCALE INTERVAL ACROSS DEPTHS

Corollary B.3 (One certified interval for four depths). *At $\eta = 1$, every $K \in \{31, 32, 33, 34\}$ and every*

$$\beta \in \left(\frac{377}{1000}, \frac{189}{500}\right) \tag{81}$$

have a stable reward-optimal fixed point at zero and an attracting DPO two-cycle $\{(a, 0), (-a, 0)\}$ with $1 < a < 6/5$ and reward gap greater than $\gamma/10$. The amplitude is unique in $(1, 6/5)$; the cycle multipliers satisfy $-1 < \lambda_x < -3/4$ and $0 < \lambda_z < 1$. DPO-1 and every fixed finite depth of IPO converge globally at the same parameters. The interval includes β_{32}.

B.4 THE LEADING POSSIBLE DIFFERENCE BEYOND QUADRATIC GEOMETRY

The next result applies to every family (1), not just the construction. Fix θ_* with $b_{\theta_*} = 0$ and write

$$M_* = M_{\theta_*}, \quad B_* = D_\theta b(\theta_*), \quad R = \mathrm{Id} - \eta M_*/4, \quad V_j = \eta \sum_{\ell=0}^{j-1} R^\ell B_*, \quad V_0 = 0. \tag{82}$$

Define the vector-valued cubic map $\mathcal{T}_*(v) = \mathbb{E}_{\theta_*}[d(d^\top v)^3]$.

Theorem B.4 (Finite-depth fourth-moment formula). *For fixed $K \geq 1$, $\eta > 0$ and $\beta > 0$, as $h \to 0$,*

$$F_{\mathrm{D},K}(\theta_* + h) - F_{\mathrm{Q},K}(\theta_* + h) = \frac{\eta}{48\beta} \sum_{j=0}^{K-1} R^{K-1-j} \mathcal{T}_*(V_j h) + O(\|h\|^4). \tag{83}$$

In particular, their derivatives of orders zero, one, and two agree at θ_*.

The shared quadratic curvature is $M_*/4 = \mathcal{I}(\theta_*)/2$. The leading possible difference uses comparison-feature fourth moments, accumulated through the inner linearized operator R. This is an information-geometric description in the stated natural coordinates; the fourth moment is not identified with Riemannian curvature, and ordinary GD is not invariant under arbitrary reparameterization. The theorem concerns the *same* depth in both methods: changing K generally changes their Jacobian. It also gives no universal sign for the cubic correction in arbitrary multidimensional families.

B.5 AN EXPLICIT MODEL REGION FOR THE SCALAR MECHANISM

Consider the scalar family (15) with $g > 0$, and let $c = \tanh(g/2)$. It uses the same BT rule, independent on-policy comparisons, matched losses, and reference refresh as (5). Its comparison features are $d = x - x'$.

Proposition B.5 (A two-parameter region of scalar coexistence). *Suppose (77) holds and set $\eta = 1$. The coefficients in (13) are*

$$A = \frac{2\alpha c}{(\alpha+2)^2}, \qquad \frac{B}{A} = \frac{10-\alpha}{6(\alpha+2)}, \qquad H = \frac{1}{\alpha+2},$$
$$H_2 = \frac{\alpha-4}{2(\alpha+2)^2}, \qquad Q_4 = \frac{4(\alpha+8)}{(\alpha+2)^2}. \tag{84}$$

In particular, $\chi_{\mathrm{Q},K} > 0$ for every depth, and $\chi_{\mathrm{D},K} < 0$ at every sufficiently large finite depth. For each depth K with $\chi_{\mathrm{D},K} < 0$, there is a nonempty open interval $\mathcal{B}_{\alpha,c,K} \subset (\beta_{c,K}, \infty)$ such that every β in it has all the following properties:

1. *DPO-K has zero as its unique fixed point. It is locally asymptotically stable and uniquely maximizes the true expected reward.*
2. *An attracting two-cycle $\{a, -a\}$ coexists with a smaller repelling two-cycle. The attracting cycle has an open basin in $\mathbb{R}$ and the strictly positive reward gap*

$$J_g(0) - J_g(a) = \frac{2\alpha g(\cosh a - 1)}{(\alpha+2)(\alpha+2\cosh a)}. \tag{85}$$

3. *Every fixed IPO depth $1 \le L \le K$, as well as DPO-1, converges to zero from every finite initial state, with a geometric rate depending on its initial interval. Every nonstationary outer step strictly increases reward.*

For every fixed outer state, both inner losses have unique finite minimizers and their unit-step GD iterations converge globally. The region (77) has nonempty interior in (α, c) and imposes no equality constraint on the reward.

This proposition proves a scalar attracting cycle and does not assert stability of an arbitrary multi-dimensional embedding. Its model region is explicit; its coexistence interval is supplied separately for each model and depth. Corollary B.3 gives a common numerical scale interval for the specified nine-response construction.

B.6 How the failure changes as the loss approaches IPO

Proposition B.6 (Depth and local coexistence near IPO)**.** *Fix $c \in (0,1)$ and $\vartheta \in (0, c^2)$. In the three-response model, set $\alpha_\lambda = 1 + \vartheta\lambda$, use $\Phi_\lambda = (1-\lambda)\Phi_{\mathrm{Q}} + \lambda\Phi_{\mathrm{D}}$ and unit inner GD, and let $\lambda \downarrow 0$. Define*

$$C_\lambda = \frac{4-\alpha_\lambda}{\alpha_\lambda + 2}, \qquad \Lambda_\lambda = \frac{2\alpha_\lambda^2(\alpha_\lambda+8)c^2}{3(\alpha_\lambda+2)^3},$$
$$\Delta_\lambda = \lambda\Lambda_\lambda - \frac{2(\alpha_\lambda - 1)}{3(\alpha_\lambda+2)}, \qquad D_0 = \frac{2(c^2-\vartheta)}{9} > 0. \tag{86}$$

Here $r_\lambda = 1 - 1/(\alpha_\lambda + 2)$ and u_K, W_K use this r_λ in (14). For every $0 < \lambda \le 1$, $\Delta_\lambda > 0$ and the finite-depth bounds are

$$-\Delta_\lambda + (C_\lambda + \lambda\Lambda_\lambda)u_K \le \chi_{\lambda,K} \le -\Delta_\lambda + (C_\lambda + 3\lambda\Lambda_\lambda)u_K. \tag{87}$$

The first depth with a strictly negative cubic coefficient satisfies

$$K_{\mathrm{sgn}}(\lambda) := \min\{K : \chi_{\lambda,K} < 0\} = \frac{\log(1/\lambda) + \log\log(1/\lambda)}{\log(3/2)} + O(1). \tag{88}$$

This is a cubic-sign threshold, not the minimum depth admitting a cycle.

For any choice of finite depths K_λ with $u_{K_\lambda} = o(\lambda)$, there is a local fold of two-cycles. Its amplitude and scale obey

$$a_{\mathrm{f}}^2 = 10(c^2-\vartheta)\lambda + o(\lambda),$$
$$\beta_{\mathrm{f}} - \beta_{c,K_\lambda} = \frac{5c(c^2-\vartheta)^2}{27}\lambda^2 + o(\lambda^2). \tag{89}$$

Such depth choices exist; for example $K_\lambda = \lceil 2\log(1/\lambda)/\log(3/2)\rceil$ has the required property for sufficiently small λ. For fixed $\rho \in (0,1)$, set $\beta = \beta_{c,K_\lambda} + \rho(\beta_{\mathrm{f}} - \beta_{c,K_\lambda})$. Zero is the stable, unique reward-optimal fixed point, and the attracting and repelling cycle amplitudes satisfy

$$a_\pm^2 = 10(c^2-\vartheta)(1 \pm \sqrt{1-\rho})\lambda + o(\lambda),$$
$$J_g(0) - J_g(a_\pm) = \frac{10g(c^2-\vartheta)}{9}(1 \pm \sqrt{1-\rho})\lambda + o(\lambda), \qquad g = 2\operatorname{arctanh} c. \tag{90}$$

The signs $+$ and $-$ denote the attracting and repelling branches, respectively. Matched IPO converges globally at these same depths and scales. Remainders are uniform when ρ stays in a compact subset of $(0,1)$. The fold is local; the result does not classify all scale intervals or retain a fixed reward gap as $\lambda \to 0$.

The finite test gives a directly computable sufficient condition $u_K < \Delta_\lambda/(C_\lambda + 3\lambda\Lambda_\lambda)$ and a necessary condition $u_K < \Delta_\lambda/(C_\lambda + \lambda\Lambda_\lambda)$ for a negative coefficient. For $c = 2/3$, $\vartheta = 1/12$, exact rational evaluation gives

λ	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
$K_{\mathrm{sgn}}(\lambda)$	18	24	30	37	43.

The certificate checks both adjacent signs; strict monotonicity then identifies the first negative depth. These numbers do not certify the onset of every possible cycle.

B.7 Depth changes the local bifurcation type

Return to (6) and allow $0 < \eta \le 3$. Let

$$c = \frac{3}{4}, \quad \rho = 1 - \frac{\eta}{4}, \quad d_K = 1 - \rho^K, \quad \beta_{c,K} = \frac{c d_K}{2}. \tag{91}$$

At zero, the first-coordinate eigenvalue of either depth-K outer map is $1 - c d_K/\beta$, which crosses -1 at $\beta_{c,K}$. The other eigenvalue lies in $(0,1)$ there. Reflection symmetry makes $t = 0$ invariant and eliminates even powers on this line. Define

$$u_K = \frac{(1-\rho)K\rho^{K-1}}{1-\rho^K}, \qquad B_K = \frac{1-\rho}{1-\rho^K} \sum_{j=0}^{K-1} \rho^{K-1-j}(1-\rho^j)^3. \tag{92}$$

Theorem B.7 (A depth threshold for flip criticality). *At $\beta = \beta_{c,K}$,*

$$F_{X,K,x}(s,0) = -s + \chi_{X,K} s^3 + O(s^5), \qquad \chi_{\mathrm{Q},K} = \frac{1}{6} + \frac{1}{2}u_K, \quad \chi_{\mathrm{D},K} = \frac{1}{6} + \frac{1}{2}u_K - \frac{15}{64}B_K. \tag{93}$$

For every fixed $\eta \in (0,3]$, $\chi_{\mathrm{D},K}$ strictly decreases from $2/3$ to $-13/192$ as K increases, whereas $\chi_{\mathrm{Q},K} > 0$ for all K. When $\eta = 1$, $\chi_{\mathrm{D},K} > 0$ for $K \le 15$ and $\chi_{\mathrm{D},K} < 0$ for $K \ge 16$.

With $\mu = c d_K/\beta - 2$, the local symmetric two-cycle satisfies $s^2 = \mu/\chi_{X,K} + O(\mu^2)$. For $\chi_{X,K} > 0$, it emerges on the $\mu > 0$ side and is attracting; for $\chi_{X,K} < 0$, it emerges on the $\mu < 0$ side and is unstable. Thus the sign change is from a supercritical to a subcritical flip. The attracting finite-amplitude cycle in Theorem 3.1 is proved separately; it cannot be inferred from the existence of the unstable small cycle.

This change has a continuous inner-time limit. If $\eta \downarrow 0$ and $K\eta \to \tau > 0$, put $x = \tau/4$. Then

$$\chi_{\mathrm{Q}}^{\mathrm{flow}}(\tau) = \frac{1}{6} + \frac{x}{2(e^x - 1)}, \tag{94}$$

$$\chi_{\mathrm{D}}^{\mathrm{flow}}(\tau) = -\frac{13}{192} + \frac{77}{64}\frac{x}{e^x - 1} - \frac{75}{128}e^{-x} + \frac{15}{128}e^{-2x}. \tag{95}$$

The latter function strictly decreases and has a unique zero $\tau_* \in (16, 17)$ (approximately 16.8361). If $K_{\mathrm{crit}}(\eta) = \min\{K : \chi_{\mathrm{D},K} < 0\}$, then $\eta K_{\mathrm{crit}}(\eta) \to \tau_*$. The limit keeps the outer update discrete; it does not assert a cycle of the inner continuous-time gradient flow.

B.8 Depth ranges and matched longitudinal gain

Corollary B.8 (Coexistence across optimization depths). *Theorem B.2 applies to every $K \ge 16$ at $\eta = 1$, and to all sufficiently large finite K for each fixed $\eta \in (0,3]$.*

Thus the local fourth-moment mechanism yields a finite-amplitude attracting failure throughout a depth-indexed family of scale intervals. To control the linear gain in the critical direction, set

$$d_1 = \eta/4, \qquad \beta_1 = \beta d_1/d_K, \qquad G_1(\theta) = \theta + \eta b_\theta/\beta_1. \tag{96}$$

Here G_1 is one-step DPO at scale β_1.

Corollary B.9 (Separation with matched longitudinal gain). *Under Theorem B.2, G_1 converges globally to zero and*

$$\partial_s G_{1,x}(0,0) = 1 - c d_1/\beta_1 = 1 - c d_K/\beta = \partial_s F_{\mathrm{D},K,x}(0,0). \tag{97}$$

From the attracting basin of $\mathcal{C}_+$, its limiting reward gap is zero, whereas that of $F_{\mathrm{D},K}$ is $\Delta_K(\beta) > 0$.

This comparison matches the first-coordinate derivative. Changing β also changes the learning rate η/β^2 in policy coordinates. The equal-depth DPO–IPO comparison retains identical η, β and matches the complete second-order expansion.

B.9 Coexistence under gradient flow and exact minimization

We next remove the finite inner step size from the construction. Keep the outer reference/sampling refresh discrete and replace the inner GD loop by

$$\frac{d}{du} w_X(u;\theta) = -\nabla_w L_X(w_X(u;\theta);\theta), \qquad w_X(0;\theta) = 0, \tag{98}$$

with $X \in \{\mathrm{D}, \mathrm{Q}\}$. For $0 < \tau < \infty$, define $\mathsf{F}_{X,\tau}(\theta) = \theta + w_X(\tau;\theta)/\beta$. For $\tau = \infty$, define $\mathsf{F}_{X,\infty}(\theta) = \theta + w_X^*(\theta)/\beta$ using the unique exact inner minimizer. Put

$$d_\tau = 1 - e^{-\tau/4}, \quad \beta_{c,\tau} = c d_\tau/2 \quad (0 < \tau < \infty), \qquad d_\infty = 1, \quad \beta_{c,\infty} = 3/8. \tag{99}$$

Theorem B.10 (Coexistence without inner discretization)**.** *In (6)–(7), for every $\tau \in (\tau_*, \infty) \cup \{\infty\}$ there is a nonempty open interval $\mathcal{B}_\tau \subset (\beta_{c,\tau}, \infty)$ such that, for every $\beta \in \mathcal{B}_\tau$:*

1. *$\mathsf{F}_{\mathrm{D},\tau}$ has the unique, locally asymptotically stable, reward-optimal fixed point zero and an attracting two-cycle $\{(a_\tau(\beta), 0), (-a_\tau(\beta), 0)\}$, with $a_\tau(\beta) > 0$ and an open basin in $\mathbb{R}^2$.*
2. *$\mathsf{F}_{\mathrm{Q},\tau}$ converges globally to zero. Its Taylor expansion through order two at zero agrees with that of $\mathsf{F}_{\mathrm{D},\tau}$.*
3. *The DPO cycle has reward gap $\Delta_\tau(\beta) = \frac{\gamma}{2}\tanh^2(a_\tau(\beta)/2) > 0$. For a frozen DPO round starting at a cycle point, every candidate at inner time $0 < u < \tau$ has strictly greater reward than the returned candidate. For $\tau = \infty$, this comparison holds for every finite $u > 0$. The inner loss decreases along the flow, which converges to its unique minimizer.*

The exact-minimization case shows that eliminating inner optimization error can retain the low-reward attractor. For finite τ, the coefficient in (95) controls coexistence of the outer map itself.

Corollary B.11 (Persistence at arbitrarily small GD steps)**.** *Fix finite $\tau > \tau_*$ and $\beta \in \mathcal{B}_\tau$. If $\eta_n \downarrow 0$ and integers K_n satisfy $K_n\eta_n \to \tau$, then, for all sufficiently large n, ordinary DPO-K_n at the same β has a stable optimal fixed point and an attracting symmetric two-cycle. Its positive amplitude a_n and reward gap Δ_n satisfy*

$$|a_n - a_\tau(\beta)| + |\Delta_n - \Delta_\tau(\beta)| \le C\{\eta_n + |K_n\eta_n - \tau|\}. \tag{100}$$

Equal-depth IPO converges globally to zero at the same η_n, β.

Here β is held fixed as the GD step tends to zero. The result follows from C^1 approximation of the outer map and a transverse intersection of its amplitude curve, rather than from coefficient convergence alone.

A narrow window and slow attraction. Figure 2 separates cycle amplitude from window width. At fixed ρ, the axial basin radius tends to zero as $\varepsilon \downarrow 0$ and grows like $\sqrt{\varepsilon}$ as excess inner time increases. This does not imply monotone basin shrinkage at fixed β.

The attraction itself weakens near onset. Put $q = \sqrt{1-\rho}$ and let μ_+ be the longitudinal multiplier of the second iterate at the larger cycle. The proof gives

$$\mu_+ = 1 - \frac{2\kappa^2}{\psi_*}(1+q)q\,\varepsilon^2 + O(\varepsilon^3). \tag{101}$$

For fixed ρ, a constant-factor reduction of a linearized perturbation therefore takes order ε^{-2} outer rounds. Window width and longitudinal stability margin vanish at the same order. This local time scale does not bound transients from arbitrary initial states; its vanishing margin requires parameter-uniform remainder control.

B.10 Finite GD at a vanishing cycle-stability margin

Proposition B.12 (Joint limits of inner time and GD step size)**.** *In the nine-response model, let $\eta \downarrow 0$, $K \in \mathbb{N}$, and $\varepsilon = K\eta - \tau_* > 0$ tend to zero. Put*

$$\beta_c^{\eta,K} = \frac{3}{8}\{1 - (1-\eta/4)^K\}, \qquad C_* = \frac{\beta_c(\tau_*)\kappa^2}{8\psi_*}, \qquad B_* = \frac{3\tau_* e^{-\tau_*/4}}{256} > 0. \tag{102}$$

The following assertions are uniform for the relative positions specified below in compact subsets of $(0,1)$.

1. *If $\eta = o(\varepsilon)$, the finite-GD amplitude curve has a local fold with*

$$(a_{\mathrm{f}}^{\eta,K})^2 = \frac{\kappa}{2\psi_*}\varepsilon + o(\varepsilon),$$
$$\beta_{\mathrm{f}}^{\eta,K} - \beta_c^{\eta,K} = C_*\varepsilon^2 + o(\varepsilon^2). \tag{103}$$

At $\beta = \beta_c^{\eta,K} + \rho C_*\varepsilon^2$, for fixed $\rho \in (0,1)$, the unique stable optimum coexists with an attracting cycle and a saddle cycle. Equations (27) and (28) hold with $o(\varepsilon)$ remainders. The connected axial basin is $(-a_-, a_-)$, and equal-depth IPO converges globally.

2. *Suppose instead that the scale is left at its flow value, $\beta = \beta_c(K\eta) + \rho\{\beta_{\mathrm{f}}(K\eta) - \beta_c(K\eta)\}$, and $\eta/\varepsilon^2 \to h \in [0,\infty)$. If $h < \rho C_*/B_*$, the same conclusions hold with*

$$\rho_{\mathrm{eff}} = \rho - B_* h/C_* \in (0,1) \tag{104}$$

in place of ρ. In particular, $\eta = o(\varepsilon^2)$ preserves the flow's leading branch and reward constants at its unmodified scale. If $h > \rho C_*/B_*$, zero is longitudinally unstable at that scale. The same instability holds if $\eta/\varepsilon^2 \to \infty$. No assertion is made at equality.

In the attracting cases the longitudinal second-iterate stability margin is of order ε^2, and the transverse multiplier has a strict limit in $(0,1)$. The rational certificates give

$$0.002931 < B_* < 0.002933, \qquad 0.0291 < C_*/B_* < 0.0293. \tag{105}$$

These are joint asymptotic guarantees, not explicit upper bounds on admissible ε or full-dimensional basin radii.

For a prescribed flow horizon $\tau_* + e$ and $K = \lfloor(\tau_* + e)/\eta\rfloor$, the effective excess satisfies $|\varepsilon - e| < \eta$. Thus $\eta = o(e^2)$ implies $\varepsilon/e \to 1$ and preserves the flow conclusions when its scale is evaluated at the prescribed horizon as well. The time-rounding error is another $O(\eta)$ perturbation. Recentring at the exact discrete critical scale removes the leading linear mismatch and permits $\eta = o(e)$ instead. Neither rate is a universal step-size rule for arbitrary feature models.

B.11 Robustness to model perturbations

The depth-32 separation also persists when the features, base weights, and BT reward are perturbed. Write $\xi = ((\ell_i, \phi_i, r_i))_{i=1}^9$ for these data, with $h_i = e^{\ell_i}$, and let ξ_0 denote the construction. Set

$$e(\xi) = \max_i\{|\ell_i - \ell_i^0| + \|\phi_i - \phi_i^0\| + |r_i - r_i^0|\}. \tag{106}$$

Use superscript ξ for the corresponding maps and reward, at the fixed values $\eta = 1$, $\beta = \beta_{32}$.

Theorem B.13 (Robust reward separation)**.** *There are $\epsilon_0, C > 0$ and a nonempty open set $U \subset \mathbb{R}^2$ such that every $e(\xi) < \epsilon_0$ has the following properties.*

1. *DPO-1, DPO-32, and IPO-32 have a common locally asymptotically stable fixed point θ_ξ, unique in a fixed neighborhood of zero. DPO-32 and IPO-32 agree through second order there.*

2. *DPO-32 has an attracting two-cycle $p_{+,\xi} \leftrightarrow p_{-,\xi}$, with $\|\theta_\xi\| + \sum_\pm \|p_{\pm,\xi} - (\pm\log 3, 0)\| \le Ce(\xi)$. From every initialization in U, DPO-32 converges alternately to these points, whereas both comparison procedures converge to θ_ξ.*

3. *At both phases,*

$$J_\xi(\theta_\xi) - J_\xi(p_{\pm,\xi}) \ge \gamma/16. \tag{107}$$

A nearby strict local reward maximum θ_ξ^r satisfies $0 \le J_\xi(\theta_\xi^r) - J_\xi(\theta_\xi) \le Ce(\xi)^2$.

4. *Every frozen population inner GD problem remains globally convergent; each step decreases its loss by at least one quarter of the squared gradient norm.*

The cycles need not remain symmetric, and their two reward values need not coincide; (107) therefore holds phase by phase. Within a six-parameter product family, exact global optimality at zero and global IPO convergence also persist, without reimposing the calibration equation (Corollary C.2). Under unrestricted feature perturbations, global reward optimality can fail even as local reward regret vanishes (Proposition C.3). The persistence argument uses stability of contracting neighborhoods; its role is to remove exact parameter tuning as an explanation of the separation.

B.12 ADAPTIVE FINITE-SAMPLE REWARD SEPARATION

In an empirical round, draw N conditionally independent triples (I, J, A) at the current policy. Replace each expectation in (3)–(4) by that empirical average, use the same batch for all K steps, and then refresh the reference and sampling policy. Denote the resulting outer trajectory by $\widehat{\theta}_{X,K,n}$, where n indexes rounds. Each procedure samples from its own current policy.

Let $D = 2\sqrt{2}$ and let $\widehat{P}$ and P_θ be the empirical and population distributions on the $9^2 \cdot 2 = 162$ labeled ordered pairs. A deterministic bound, valid at every finite θ, is

$$\left\|\widehat{P} - P_\theta\right\|_1 \le \epsilon \quad \Longrightarrow \quad \begin{cases} \left\|\widehat{F}_{\mathrm{D},K}(\theta) - F_{\mathrm{D},K}(\theta)\right\| \le KD\epsilon/\beta, \\ \left\|\widehat{F}_{\mathrm{Q},K}(\theta) - F_{\mathrm{Q},K}(\theta)\right\| \le 5KD\epsilon/(2\beta). \end{cases} \tag{108}$$

No lower eigenvalue bound on an empirical Fisher matrix is required. The bound is uniform in the inner iterate, so batch reuse causes no independence gap in the argument.

Theorem B.14 (Finite-sample and growing-batch separation)**.** *Fix the model, $\eta = 1$ and $\beta = \beta_{32}$ of Theorem 3.1. There are an open neighborhood U of a_+, constants $C, \epsilon_0 > 0$, $q \in (0,1)$, and an integer n_0, independent of N, T, α, with the following property. Initialize DPO-1, DPO-32 and IPO-32 at the same $\theta_0 \in U$. For $T \ge n_0$, $0 < \alpha < 1$ and*

$$\epsilon_{N,T,\alpha} = \sqrt{\frac{2}{N}\left(162\log 2 + \log\frac{3T}{\alpha}\right)} \le \epsilon_0, \tag{109}$$

with probability at least $1-\alpha$, simultaneously for $n_0 \le n \le T$,

$$0 \le J_* - J(\widehat{\theta}_{X,K,n}) \le C\{q^{2(n-n_0)} + \epsilon^2_{N,T,\alpha}\}, \qquad (X,K) \in \{(\mathrm{D},1),(\mathrm{Q},32)\}, \tag{110}$$

$$\left|J_* - J(\widehat{\theta}_{\mathrm{D},32,n}) - \gamma/8\right| \le C\{q^n \left\|\theta_0 - a_+\right\| + \epsilon_{N,T,\alpha}\}. \tag{111}$$

For each horizon T, use batch size N_T at every round of that run. If $N_T/\log T \to \infty$, then the two gaps in (110) converge to zero in probability and expectation at round T, while the DPO-32 gap converges to $\gamma/8$ in both senses.

The growing-batch assertion is a triangular-array limit, not permanent trapping with a fixed batch size. The finite-time bound follows by a conditional concentration inequality and deterministic tracking of contracting neighborhoods. We provide the full construction of the shared initial neighborhood and finite burn-in in Appendix D.35. For sufficiently small ϵ, each empirical inner DPO step also decreases its own empirical loss, with decrease coefficient $3/4$.

C SUPPLEMENTARY RESULTS

Table 1: Scope of the population guarantees and joint limits. Local coexistence scalings fix an interior relative scale $\rho \in (0, 1)$.

Result	Setting	Guaranteed conclusion
Loss coefficient (Theorem 3.2)	Symmetric invariant scalar direction	Depth-dependent cubic sign test
Loss-class cycles (Theorem 3.2)	Three-response model; global curvature and tail-slope conditions	An open failure region for every positive logistic weight
Near-IPO limit (Proposition B.6)	Three responses; $\alpha_\lambda = 1 + \vartheta\lambda$, $0 < \vartheta < c^2$; unit GD step; $\lambda \downarrow 0$	$K_{\mathrm{sgn}} = [L + \log L]/\log(3/2) + O(1)$, $L = \log(1/\lambda)$; deeper choices give width $\asymp \lambda^2$ and reward gap $\asymp \lambda$
Local fold (Theorem 3.3)	Nine-response DPO; inner gradient flow, $\tau = \tau_* + \varepsilon$	Width $\asymp \varepsilon^2$; reward gap $\asymp \varepsilon$; axial radius $\asymp \sqrt{\varepsilon}$
Finite GD near onset (Proposition B.12)	Nine-response DPO; $\varepsilon = K\eta - \tau_* \downarrow 0$	Fold scalings and attraction persist: $\eta = o(\varepsilon^2)$ at the flow scale, or $\eta = o(\varepsilon)$ after discrete recentering
Safeguard (Theorem 3.4)	Nine-response population DPO; product structure	Geometric convergence for arbitrary inner depths

For Proposition B.6, K_{sgn} is the first depth with a negative *cubic coefficient*, not the first depth with any cycle; the stated width and reward scalings require $u_{K_\lambda} = o(\lambda)$. Proposition B.12 also gives the finite-ratio boundary when $\eta/\varepsilon^2 \to h$, including the loss of stability of zero above that boundary. The table records sufficient rates preserving the flow's leading constants. Neither result supplies an explicit onset neighborhood or a full-dimensional basin radius.

The proof route is shared across losses: the scalar calculation in Appendix D.7 yields the cubic formula, Lemma A.14 converts global amplitude control into coexistence, and Appendix D.10 supplies the loss-independent scalar IPO comparison. The DPO specializations substitute their moments and separately verify transverse attraction where needed.

Table 2: Population behavior at the same $\eta = 1$ and $\beta = \beta_{32}$. The IPO loss is normalized as in (4).

Procedure	Outer limiting behavior	Limiting reward gap
DPO, $K = 1$	Global convergence to zero	0
IPO, any fixed $K \geq 1$	Global convergence to zero	0
DPO, $K = 32$	Attracting two-cycle from an open set	$\gamma/8$

C.1 WHAT THE ENVELOPE ADDS TO CONSTANT DAMPING

The safeguard is a diagnostic intervention in the failure model. Exact IPO is already globally convergent there, and we prove no computational, reward, or convergence-speed advantage for running DPO first. Constant damping can also work. The following statement makes its relationship with the envelope explicit.

Proposition C.1 (Constant damping and the maximal envelope fraction)**.** *Use the model and parameter range of Theorem 3.4.*

1. *Fix a compact initial rectangle C and use its constants in (158). Every constant $0 < a_0 \leq 3h_*$ gives global convergence within C for arbitrary positive inner depths, including exact minimization, under $\theta^+ = \theta + a_0 w_{\mathrm{D},K}(\theta)/\beta$. The rate (31) holds with $\rho_C = \min\{1/2, a_0 m_*/\beta, \varepsilon_*\}$.*

2. *For a fixed finite K with $\chi_{\mathrm{D},K} < 0$, let $A_{\max} = \max_{s\geq 0} \mathcal{A}_K(s)$. A constant $a_0 \in (0,1]$ strictly decreases $|s|$ at every nonzero point on $t = 0$ if and only if $a_0 < \beta/A_{\max}$. If $a_0 = \beta/A_{\max} \leq 1$, a positive maximizer yields a two-cycle. If $a_0 > \beta/A_{\max}$, strict absolute-coordinate decrease fails at a maximizer.*

3. *At every state and inner depth, (30) is the largest $a \in [0,1]$ satisfying $|aw_i| \leq |z_{\theta,i}|$ in both coordinates. It therefore needs no initial rectangle or uniform curvature bound as an input, and it accepts every one-step DPO update in full.*

Part 1 is a sufficient compact-set bound, not an optimal constant damping threshold in two dimensions. Part 2 is sharp only for strict decrease of $|s|$ on the invariant line; it is not a complete characterization of all convergent two-dimensional trajectories. Part 3 is pointwise maximality subject to the specified envelope, not optimality for reward.

C.2 Structured perturbations and the scope of global optimality

Consider the six-parameter family

$$p_{s,t}(x,z) = \frac{h_\alpha(x)e^{usx}}{\alpha + 2\cosh(us)} \frac{k_\zeta(z)e^{vtz}}{\zeta + 2\cosh(vt)}, \qquad r(x,z) = -\gamma' x^2 - \delta' z^2, \tag{112}$$

where $h_\alpha(0) = \alpha$, $k_\zeta(0) = \zeta$, their endpoint weights are one, and the sufficient statistic is (ux, vz). The base point is $(\alpha, \zeta, u, v, \gamma', \delta') = (2, 1, 1, 1, \gamma, \delta)$.

Corollary C.2 (An open structured family with a globally optimal fixed point)**.** *There is an open neighborhood of this six-parameter base point such that, at the unchanged $\eta = 1$, $\beta = \beta_{32}$, zero remains uniquely globally reward-optimal and locally stable for DPO-32, while an attracting cycle $\{(a_\xi, 0), (-a_\xi, 0)\}$ has reward gap at least $\gamma/16$. DPO-1, every fixed finite depth of IPO, and exact IPO converge globally to zero and strictly increase reward at every nonstationary outer step. No calibration constraint is imposed on γ', δ'.*

Proposition C.3 (An obstruction to unrestricted global optimality)**.** *Keep all base weights and rewards unchanged and replace only the feature of response $(0,1)$ by $(0, 1+\epsilon)$, with $\epsilon > 0$. Let θ_ϵ be the fixed point from Theorem B.13, and set $J_\epsilon^* = \sup_{\theta\in\mathbb{R}^2} J_\epsilon(\theta)$. Then*

$$\liminf_{\epsilon\downarrow 0} \{J_\epsilon^* - J_\epsilon(\theta_\epsilon)\} \geq \gamma/2 - \delta/3 > 0. \tag{113}$$

Thus the local reward bound in Theorem B.13 cannot be replaced by vanishing global reward regret for arbitrary feature perturbations on the unbounded parameter space.

C.3 The final inner step lowers loss and reward

Proposition C.4 (Strict opposition at the cycle)**.** *At $\theta = a_+$, $\eta = 1$ and $\beta = \beta_{32}$, the DPO inner iterates satisfy*

$$L_{\mathrm{D}}(w_{\mathrm{D},32}; a_+) < L_{\mathrm{D}}(w_{\mathrm{D},31}; a_+), \qquad J(a_+ + w_{\mathrm{D},32}/\beta_{32}) < J(a_+ + w_{\mathrm{D},31}/\beta_{32}).$$

Both strict inequalities persist for sufficiently small perturbations of the outer state and of the labeled-pair distribution. They also persist under sufficiently small additive errors in the finitely many inner steps.

C.4 Persistence in the scale parameter

Proposition C.5 (An open interval of coexistence)**.** *There is an open interval $\mathcal{B}$ containing β_{32} such that, for every $\beta \in \mathcal{B}$, DPO-32 has a nearby attracting period-two orbit with strictly positive reward gap at both points. Throughout $\mathcal{B}$, zero is still the only fixed point and is locally stable, and every finite-depth IPO map and DPO-1 converge globally to zero. The cycle's reward gap need not equal $\gamma/8$ away from β_{32}.*

C.5 A frozen-Fisher update and exact minimization

This supplement records a complementary separation for a classical quadratic-majorization update. To distinguish it from ordinary GD, define

$$\bar{w}_0 = 0, \qquad \bar{w}_{j+1} = \bar{w}_j - 4M_\theta^{-1}\nabla L_{\mathrm{D}}(\bar{w}_j; \theta), \qquad \bar{F}_j(\theta) = \theta + \bar{w}_j(\theta)/\beta. \tag{114}$$

The matrix is frozen throughout a round. Since $M_\theta = 2\mathcal{I}(\theta)$, this uses $2\mathcal{I}(\theta)^{-1}$ as a preconditioner. The Hessian upper bound is the familiar logistic quadratic-majorization construction (Böhning & Lindsay, 1988).

Proposition C.6 (Loss decrease and local order of approximation)**.** *For every finite regular family (1), a step $w^+ = w - 4M_\theta^{-1}g$, $g = \nabla L_{\mathrm{D}}(w;\theta)$, satisfies*

$$L_{\mathrm{D}}(w^+;\theta) \le L_{\mathrm{D}}(w;\theta) - 2g^\top M_\theta^{-1} g. \tag{115}$$

Its first iterate is exactly the IPO minimizer, $\bar{w}_1 = w^*_{\mathrm{Q}}$. If $b_{\theta_*} = 0$ and $w^*_{\mathrm{D}}(\theta)$ is the exact DPO minimizer, then for each fixed $K \ge 0$,

$$\bar{w}_K(\theta_* + h) - w^*_{\mathrm{D}}(\theta_* + h) = O(\|h\|^{2K+1}). \tag{116}$$

Consequently $\bar{F}_K$ agrees with the exact-DPO outer map through order $2K$, and all $K \ge 2$ have the same cubic term as exact DPO.

Proposition C.7 (A one-step versus two-step separation)**.** *In (6)–(7), let*

$$a = \log 2, \qquad \bar{\beta} = \frac{1 - \frac{10}{9}\tanh(1/4) - \frac{4}{9}\tanh(1/2)}{2\log 2}. \tag{117}$$

Then $\bar{F}_1$ converges globally to zero, whereas $\bar{F}_2$ has an attracting two-cycle $(a,0) \leftrightarrow (-a,0)$ with reward gap $\gamma/18$. Both maps have zero as their only fixed point, and it is locally stable. On the invariant line at the flip value $\beta = 3/8$, the cubic coefficient is $1/6$ for $\bar{F}_1$ and $-13/192$ for $\bar{F}_K$ with every $K \ge 2$, as well as for exact DPO.

D Proofs

D.1 Curvature and convergence of each frozen inner problem

Proof of Lemma A.1. Because the support is finite, differentiation and expectation commute. Differentiating $\log(1+e^u)$ gives $\sigma(u)$, and differentiating $\sigma(u)$ gives $\sigma(u)(1-\sigma(u)) \in (0, 1/4]$. This gives (32)–(33). If $\bar{\phi}_\theta = \mathbb{E}_\theta \phi(Y)$, then independence gives

$$\begin{aligned} M_\theta &= \mathbb{E}[(\phi(I) - \phi(J))(\phi(I) - \phi(J))^\top] \\ &= 2\mathbb{E}[\phi(Y)\phi(Y)^\top] - 2\bar{\phi}_\theta \bar{\phi}_\theta^\top = 2\,\mathrm{Cov}_\theta(\phi(Y)) = 2\nabla^2\psi(\theta). \end{aligned}$$

The score is $\nabla_\theta \log p_\theta(Y) = \phi(Y) - \bar{\phi}_\theta$, so this covariance is precisely $\mathcal{I}(\theta)$. For any nonzero u, the spanning assumption provides a pair with $u^\top d \ne 0$ and positive probability. Therefore $u^\top M_\theta u > 0$; the same argument with the strictly positive factor $\sigma'(d^\top w)$ proves strict positivity of the logistic Hessian.

We prove coercivity, including a uniform bound over directions. Set $q_{ij} = \sigma(r_i - r_j)$ and $q_{\min} = \min_{i,j} \min(q_{ij}, 1 - q_{ij}) > 0$. For any $z \in \mathbb{R}$,

$$\begin{aligned} z \ge 0: &\quad \log(1+e^z) - q_{ij} z \ge (1 - q_{ij}) z, \\ z < 0: &\quad \log(1+e^z) - q_{ij} z \ge -q_{ij} z. \end{aligned}$$

Thus $L_{\mathrm{D}}(w;\theta) \ge q_{\min}\mathbb{E}_\theta|d^\top w|$. On the unit sphere the continuous function $u \mapsto \mathbb{E}_\theta|d^\top u|$ is positive. Its minimum $c_\theta > 0$ is attained, and hence $L_{\mathrm{D}}(w;\theta) \ge q_{\min} c_\theta \|w\| \to \infty$. Strict convexity and coercivity prove existence and uniqueness of a finite minimizer. The quadratic loss has these properties because $M_\theta \succ 0$.

For the model, let $C = \cosh s$, $S = \sinh s$. Its x marginal has

$$\mathbb{E}x = \frac{S}{1+C}, \qquad \mathbb{E}x^2 = \frac{C}{1+C}, \qquad \mathrm{Var}(x) = \frac{C(1+C) - S^2}{(1+C)^2} = \frac{1}{1+C}.$$

For $C_t = \cosh t$, $S_t = \sinh t$ the z marginal gives

$$\mathbb{E}z = \frac{2S_t}{1+2C_t}, \quad \mathbb{E}z^2 = \frac{2C_t}{1+2C_t}, \quad \mathrm{Var}(z) = \frac{2C_t(1+2C_t) - 4S_t^2}{(1+2C_t)^2} = \frac{2(C_t+2)}{(1+2C_t)^2}.$$

The coordinates are independent. Twice these variances gives (34), using $1 + C = 2\cosh^2(s/2)$. The first entry is at most one. For $C_t \geq 1$, $3(C_t+2) \leq (1+2C_t)^2$ is equivalent to $4C_t^2 + C_t - 5 \geq 0$, which holds because it is zero at one and has positive derivative there and thereafter. This bounds the second entry by $4/3$.

Let $g = \nabla L_X(w;\theta)$, keeping θ fixed. Taylor's formula with integral remainder, together with the Hessian bound, gives

$$\begin{aligned} L_X(w - \eta g;\theta) &= L_X(w;\theta) - \eta \|g\|^2 + \eta^2 \int_0^1 (1-u) g^\top \nabla^2 L_X(w - u\eta g;\theta) g \, du \\ &\leq L_X(w;\theta) - \eta \|g\|^2 + \frac{\eta^2}{6} \|g\|^2 . \end{aligned}$$

For $0 < \eta \leq 3$, the decrease coefficient is positive. Summing the decrease over all iterates and using the lower bound on the loss yields $\sum_j \|g_j\|^2 < \infty$, so $g_j \to 0$. The iterates remain in a compact sublevel set by coercivity. Every cluster point is stationary by continuity of the gradient and is thus the unique minimizer. A bounded sequence with a unique cluster point converges to it. This proves the asserted global convergence without any claim of uniform strong convexity over all θ, w. □

D.2 Calibration, pair enumeration, and global IPO convergence

Proof of Lemma A.2. Let $T(u) = \tanh(u/2)$ and $g(u) = 5T(u)/9 + 2T(u-\delta)/9 + 2T(u+\delta)/9$. Then $g(0) = 0$, $g'(u) > 0$ for every finite u, and $g(u) \to 1$ as $u \to \infty$. The intermediate value theorem and strict monotonicity give exactly one positive solution of $g(\gamma) = 3/4$. At $t = 0$, z is uniform on $\{-1, 0, 1\}$; the probabilities of $z^2 - z'^2 = 0, 1, -1$ are respectively $5/9, 2/9, 2/9$. This proves $c_x(0) = 3/4$.

We use the following identity, obtained by putting the two hyperbolic tangents over a common denominator:

$$\begin{aligned} \frac{\tanh((u+v)/2) + \tanh((u-v)/2)}{2} &= \frac{\sinh u}{2\cosh((u+v)/2)\cosh((u-v)/2)} \\ &= \frac{\sinh u}{\cosh u + \cosh v}. \end{aligned}$$

For $u > 0$ this is strictly positive and at most $\sinh u/(\cosh u + 1) = \tanh(u/2)$. The difference of two iid random variables has a symmetric distribution. Applying this observation to both squared-coordinate differences proves (38). Also $\tanh(\delta/2) < \delta/2 = 1/40$.

For completeness, $T'(u) = \frac{1}{2}\operatorname{sech}^2(u/2)$ and $T''(u) = -\frac{1}{2}\operatorname{sech}^2(u/2)\tanh(u/2)$, so $|T''(u)| \leq 1/2$. Taylor's integral formula on the two sides gives

$$T(\gamma+\delta) + T(\gamma-\delta) - 2T(\gamma) = \int_0^\delta (\delta - v)[T''(\gamma+v) + T''(\gamma-v)]\, dv.$$

Its absolute value is at most $\delta^2/2$. Consequently

$$\frac{3}{4} = T(\gamma) + \frac{2}{9}[T(\gamma+\delta) + T(\gamma-\delta) - 2T(\gamma)] \geq T(\gamma) - \frac{\delta^2}{9}.$$

Dividing by two proves (39).

We now compute b_θ by ordered-pair enumeration. Write the x probabilities as p_-, p_0, p_+ and the z probabilities as q_-, q_0, q_+ (these marginal symbols are distinct from the pairwise preference q_{IJ}). Conditional on $x^2 - x'^2 = 1$, averaging over z, z' gives $\mathbb{E}(q_{IJ} - 1/2) = -c_x(t)/2$. For squared difference -1 the average is $c_x(t)/2$; for squared difference zero it is zero by symmetry. For the four ordered comparisons between zero and an endpoint, the contributions to $\mathbb{E}[(x-x')(q_{IJ}-1/2)]$ are

(x, x')	$(1,0)$	$(0,1)$	$(-1,0)$	$(0,-1)$
contribution	$-p_+p_0c_x/2$	$-p_0p_+c_x/2$	$p_-p_0c_x/2$	$p_0p_-c_x/2.$

All other contributions vanish after averaging. Therefore

$$b_x = -c_x p_0(p_+ - p_-) = -\frac{c_x \sinh s}{(1+\cosh s)^2}, \qquad b_z = -c_z q_0(q_+ - q_-) = -\frac{2c_z \sinh t}{(1+2\cosh t)^2}. \quad (118)$$

The second equality follows from the identical enumeration with the coordinates interchanged. Solving $-b + Mw/4 = 0$ using (34) now gives (40).

For the finite GD iterate, subtract w_{Q}^* from the affine update:

$$w_{\mathrm{Q},j+1} - w_{\mathrm{Q}}^* = (\mathrm{Id} - M_\theta/4)(w_{\mathrm{Q},j} - w_{\mathrm{Q}}^*).$$

Starting from $w_{\mathrm{Q},0} = 0$ yields (41). Each diagonal entry of M_θ lies in $(0, 4/3]$, so the coordinate multipliers $a_{i,K} = 1 - (1 - M_{ii}/4)^K$ lie strictly between zero and one.

For $s \neq 0$, the exact-IPO displacement is toward zero and its relative magnitude is

$$\frac{|w_{\mathrm{Q},x}^*|}{\beta|s|} = \frac{2c_x(t)}{\beta}\frac{\tanh(|s|/2)}{|s|} \leq \frac{c_x(t)}{\beta} < 2,$$

where the last inequality uses (39) and the hypothesis on β. For $u = |t| > 0$, $\sinh u = \int_0^u \cosh v\, dv \leq u \cosh u$, hence

$$\frac{|w_{\mathrm{Q},z}^*|}{\beta|t|} \leq \frac{2c_z(s)}{\beta}\frac{\cosh u}{\cosh u + 2} < 2,$$

since $c_z(s) < 1/40 < \beta$. A move toward zero with relative magnitude strictly in $(0, 2)$ strictly decreases the coordinate's absolute value. Multiplying the move by $a_{i,K} \in (0, 1)$ preserves that property. Zero coordinates remain zero. Thus $V(s, t) = |s| + |t|$ strictly decreases at every nonzero outer state and all trajectories remain in their initial closed coordinate rectangle.

To justify convergence rather than just decrease, let $V_n \downarrow \ell$. If a convergent subsequence of states has limit $\bar{\theta}$, continuity implies $V(\bar{\theta}) = \ell$ and $V(F_{\mathrm{Q},K}(\bar{\theta})) = \ell$. Strict decrease forces $\bar{\theta} = 0$. Compactness then gives convergence of the entire trajectory to zero. This argument also applies to exact IPO. Finally, $C \mapsto C/(1 + C)$ and $C \mapsto 2C/(1 + 2C)$ have positive derivatives for $C \geq 1$. Formula (8) therefore increases strictly whenever a nonzero coordinate's absolute value decreases, and has its unique maximum at zero. □

D.3 The scalar recursion and certified longitudinal attraction

Proof of Lemma A.3. Let $S_x = \mathrm{diag}(-1, 1)$. Relabeling responses by $(x, z) \mapsto (-x, z)$ maps the law at θ to that at $S_x\theta$, preserves rewards, and maps feature differences to $S_x d$. Thus $L_X(S_x w; S_x\theta) = L_X(w; \theta)$, and differentiation gives $\nabla_w L_X(S_x w; S_x\theta) = S_x \nabla_w L_X(w; \theta)$. Induction from zero proves equivariance of every inner and outer iterate. The same proof applies to $S_z = \mathrm{diag}(1, -1)$. At $t = 0$ it forces $w_{\mathrm{D},j,z} = 0$ and makes $\nu_j(s)$ odd.

For an x-coordinate increment $w_x = -v$, let $C = \cosh s$, $S = \sinh s$, $A = \tanh(v/2)$, $B = \tanh v$. The total probabilities of $|x - x'| = 1$ and $|x - x'| = 2$ are

$$2p_0(p_+ + p_-) = \frac{2C}{(1 + C)^2}, \qquad 2p_+p_- = \frac{1}{2(1 + C)^2}.$$

For differences ± 1, $(x - x')\{\sigma(-v(x - x')) - 1/2\} = -A/2$. For differences ± 2 the same expression is $-B$. Hence

$$\mathbb{E}[d_x\{\sigma(-vd_x) - 1/2\}] = -\frac{CA}{(1 + C)^2} - \frac{B}{2(1 + C)^2},$$

$$\nabla_x L_{\mathrm{D}}((-v, 0); (s, 0)) = \frac{cS - CA - B/2}{(1 + C)^2} = f(s, v).$$

Here $-b_x = cS/(1 + C)^2$ and $c = 3/4$ were used. Since $\nu_{j+1} = -w_{x,j+1} = \nu_j + \nabla_x L_{\mathrm{D}}$, this proves (42)–(43). At $a = \log 3$, $C = 5/3$, $S = 4/3$, $1 + C = 8/3$, giving (9).

Direct differentiation, holding v or s fixed respectively, gives

$$f_v(s, v) = -\frac{C(1 - A^2) + 1 - B^2}{2(1 + C)^2}, \tag{119}$$

$$f_s(s, v) = \frac{\frac{3}{2}(2 - C)(1 + C) + 2S\{(C - 1)A + B\}}{2(1 + C)^3}. \tag{120}$$

For clarity, if $N = \frac{3}{2}S - 2CA - B$ is the numerator of f, then $N_s = \frac{3}{2}C - 2SA$, and the quotient rule gives

$$f_s = \frac{(\frac{3}{2}C - 2SA)(1+C) - 2S(\frac{3}{2}S - 2CA - B)}{2(1+C)^3}.$$

Using $S^2 = C^2 - 1$ reduces the terms without A, B to $\frac{3}{2}(2-C)(1+C)$ and the A, B terms to $2S\{(C-1)A + B\}$. Substituting $C = 5/3$, $S = 4/3$ in $u_{j+1} = (1+f_v)u_j + f_s$ yields (44).

The update $T(v) = v + f(a, v)$ is strictly increasing because $T'(v) \geq 13/16$. Also $f_v(a, v) < 0$ for finite v and $f(a, 0) = 9/64 > 0$. The elementary bounds

$$\frac{8}{3} < e < \frac{11}{4} \tag{121}$$

follow from the exponential series: the lower bound is its partial sum through degree three, with a strictly positive tail; the partial sum through degree four is $65/24$, and the remaining tail is at most $(1/120)/(1 - 1/6) = 1/100$, whose sum is below $11/4$. It follows that $\tanh(1/2) > 5/11$ and $\tanh 1 > 55/73$. In particular, $\frac{10}{3}\tanh(1/2) + \tanh 1 > 50/33 + 55/73 = 5465/2409 > 2$, so $f(a, 1) < 0$. There is therefore a unique $v_\infty \in (0, 1)$ with $f(a, v_\infty) = 0$. If $0 \leq v < v_\infty$, then $f(a, v) > 0$ and the monotonicity of T gives $v < T(v) < T(v_\infty) = v_\infty$. This proves the asserted strict increase and bound by induction.

Finally, (45) follows from the rational interval evaluation in Appendix F, applied to the exact recurrences (9) and (44). That appendix proves the enclosure rules, exponential and logarithm remainder bounds, and gives the full program and its rational output intervals. Thus the two finite inequalities are reproducible exact-arithmetic parts of this proof, not floating-point estimates. □

D.4 Transverse attraction, with the distribution derivative included

Proof of Lemma A.4. At $t = 0$ and $w_z = 0$, the partial derivative of the z gradient with respect to t must include the changing sampling law. The z-marginal score at zero is z, so the score of the ordered pair is $z + z'$. Holding w fixed and using $d_z(z + z') = z^2 - z'^2$ gives

$$\partial_t \nabla_z L_{\mathrm{D}}((w_x, 0); (s, 0)) = \mathbb{E}\big[(z^2 - z'^2)\{\sigma(w_x(x - x')) - \sigma(R_x - \delta(z^2 - z'^2))\}\big].$$

The first term vanishes because $\mathbb{E}(z^2 - z'^2) = 0$ and the two coordinates are independent. The probabilities of squared difference 1 and -1 are each $2/9$. The remaining term is exactly

$$-\frac{2}{9}\mathbb{E}\sigma(R_x - \delta) + \frac{2}{9}\mathbb{E}\sigma(R_x + \delta) = \mathcal{N}(s).$$

It is independent of w_x. Since σ is strictly increasing, $\mathcal{N}(s) > 0$, and the bound $\sigma' \leq 1/4$ gives $\sigma(R_x + \delta) - \sigma(R_x - \delta) \leq \delta/2$. This proves (46).

At $w_z = 0, t = 0$, the mixed Hessian is

$$\mathbb{E}[d_x d_z \sigma'(w_x d_x)] = \mathbb{E}_{x,x'}[d_x \sigma'(w_x d_x)]\mathbb{E}_{z,z'} d_z = 0.$$

The zz Hessian is $\mathbb{E}d_z^2\, \mathbb{E}_{x,x'}\sigma'(w_x d_x)$. Uniform z has variance $2/3$, so $\mathbb{E}d_z^2 = 4/3$ and this is H_j at the jth iterate. Differentiating the inner recursion, including both its iterate and outer-parameter arguments, yields $U_{j+1} = U_j - H_j U_j - \mathcal{N}(s)$, as asserted.

At a cycle point, Lemma A.3 gives $|v_j d_x| < 2$. Since $\sigma'(u) = 1/(e^u + 2 + e^{-u})$ is even and decreasing on $[0, \infty)$, (121) implies

$$\sigma'(2) = \frac{1}{e^2 + 2 + e^{-2}} > \frac{1}{121/16 + 2 + 9/64} = \frac{64}{621} > \frac{1}{10}.$$

Therefore $2/15 < H_j \leq 1/3$. Solving the linear recursion gives (49); its upper bound follows from

$$\kappa_{32} \leq \mathcal{N}(s) \sum_{r=0}^{31} (13/15)^r \leq \frac{15}{2}\mathcal{N}(s) \leq \frac{15}{2}\frac{\delta}{9} = \frac{5\delta}{6}.$$

It is positive because every summand is positive. The outer derivative in the z direction is $1 + U_{32}/\beta_{32}$. Since $\beta_{32} > 3/8$ and $\delta = 1/20$,

$$0 < \frac{\kappa_{32}}{\beta_{32}} < \frac{(5/6)(1/20)}{3/8} = \frac{1}{9}.$$

Reflection in z makes the first outer coordinate even in t and the second coordinate identically zero along $t = 0$. Both off-diagonal entries of the Jacobian thus vanish. Reflection in x transfers the same diagonal entries to a_-. This completes the proof. □

D.5 The general finite-depth expansion

Proof of Theorem B.4. The sigmoid satisfies $\sigma(u) - 1/2 = \frac{1}{2}\tanh(u/2)$. Differentiation at zero gives

$$\sigma(0) = \frac{1}{2}, \quad \sigma'(0) = \frac{1}{4}, \quad \sigma''(0) = 0, \quad \sigma'''(0) = -\frac{1}{8}, \quad \sigma^{(4)}(0) = 0.$$

Taylor's theorem on a fixed neighborhood of zero therefore gives $\sigma(u) = 1/2 + u/4 - u^3/48 + O(|u|^5)$. Finiteness of the support makes this remainder uniform over the feature differences, and over θ in a compact neighborhood of θ_*. Thus

$$\nabla L_{\mathrm{D}}(w;\theta) = -b_\theta + \frac{1}{4}M_\theta w - \frac{1}{48}\mathcal{T}_\theta(w) + O(\|w\|^5), \tag{122}$$
$$\mathcal{T}_\theta(w) = \mathbb{E}_\theta[d(d^\top w)^3].$$

Since $b_{\theta_*} = 0$, both sequences start and remain at zero when $\theta = \theta_*$. For $\theta = \theta_* + h$, all maps are smooth and K is finite, so induction in the recursion gives $w_{X,j} = O(\|h\|)$ for each $j \le K$. More explicitly, $b_{\theta_*+h} = B_* h + O(\|h\|^2)$ and $M_{\theta_*+h} = M_* + O(\|h\|)$. If $w_{X,j} = V_j h + O(\|h\|^2)$, then either loss gives

$$w_{X,j+1} = (RV_j + \eta B_*)h + O(\|h\|^2).$$

The initial condition $V_0 = 0$ solves this recursion as $V_j = \eta\sum_{\ell=0}^{j-1} R^\ell B_*$.

Put $e_j = w_{\mathrm{D},j} - w_{\mathrm{Q},j}$ at the same outer parameter. Subtracting updates and using (122) yields

$$e_{j+1} = (\mathrm{Id} - \eta M_\theta/4)e_j + \frac{\eta}{48}\mathcal{T}_\theta(w_{\mathrm{D},j}) + O(\|h\|^5). \tag{123}$$

Starting with $e_0 = 0$ and using $w_{\mathrm{D},j} = O(\|h\|)$ shows by induction that $e_j = O(\|h\|^3)$. In the first term of (123), replacing M_θ by M_* changes the result by $O(\|h\|)O(\|h\|^3) = O(\|h\|^4)$. In the cubic term, the change of distribution from θ to θ_* contributes $O(\|h\|^4)$, because the probabilities change by $O(\|h\|)$. Replacing $w_{\mathrm{D},j} = V_j h + O(\|h\|^2)$ by $V_j h$ also contributes $O(\|h\|^4)$: in the scalar identity $(a+b)^3 - a^3 = 3a^2b + 3ab^2 + b^3$, here $a = O(\|h\|)$ and $b = O(\|h\|^2)$. Consequently

$$e_{j+1} = Re_j + \frac{\eta}{48}\mathcal{T}_*(V_j h) + O(\|h\|^4).$$

Repeated substitution, with fixed finite K, gives

$$e_K = \frac{\eta}{48}\sum_{j=0}^{K-1} R^{K-1-j}\mathcal{T}_*(V_j h) + O(\|h\|^4).$$

Dividing by β proves (83). The functions are smooth; the displayed expansion therefore also proves equality of their Taylor coefficients through degree two. □

D.6 Interpreting scalar curvature through feature moments

Proof of Lemma A.12. Centering the sufficient statistic does not change its exponential tilt or any comparison difference. The cumulant generating function of the centered statistic is $\psi_e(s) = \log \mathbb{E}_0 e^{sX}$. Symmetry gives $\mathbb{E}_0 X = \mathbb{E}_0 X^3 = 0$, so finite support and the power series for the exponential and logarithm yield

$$\mathbb{E}_0 e^{sX} = 1 + \frac{\mu_2}{2}s^2 + \frac{\mu_4}{24}s^4 + O(s^6),$$
$$\psi_e(s) = \frac{\mu_2}{2}s^2 + \left(\frac{\mu_4}{24} - \frac{\mu_2^2}{8}\right)s^4 + O(s^6),$$
$$\psi_e''(s) = \mu_2 + \frac{\mu_4 - 3\mu_2^2}{2}s^2 + O(s^4).$$

The derivatives of ψ_e are the tilted mean and variance. Independence gives $\mathbb{E}_s(X - X')^2 = 2\operatorname{Var}_s X = 2\psi_e''(s)$. Dividing by four proves the expressions for H and H_2. At zero, expanding the fourth power and using centering gives

$$\begin{aligned}\mathbb{E}_0(X - X')^4 &= \mathbb{E}_0X^4 - 4\mathbb{E}_0X^3\mathbb{E}_0X' + 6\mathbb{E}_0X^2\mathbb{E}_0(X')^2 - 4\mathbb{E}_0X\mathbb{E}_0(X')^3 + \mathbb{E}_0(X')^4 \\ &= 2\mu_4 + 6\mu_2^2.\end{aligned}$$

Finally $\mu_2 > 0$ implies

$$H_2 \le 0 \iff \kappa \le 3, \qquad \frac{\chi_{\mathrm{Q},\infty}}{\mu_2} = \frac{2B}{A\mu_2} + \kappa - 3.$$

For the other side of the comparison,

$$\frac{\Lambda}{\mu_2} = \frac{Q_4}{24H\mu_2}\left(\frac{A}{H}\right)^2 = \frac{2\mu_2^2(\kappa+3)}{24(\mu_2/2)\mu_2}\left(\frac{A}{H}\right)^2 = \frac{\kappa+3}{6}\left(\frac{A}{H}\right)^2.$$

Dividing the two strict inequalities in (74) by μ_2 proves the equivalence. □

D.7 A SCALAR CRITERION BEYOND THE CONSTRUCTED PRODUCT FAMILY

Proof of Theorem B.1 and the local part of Theorem 3.2. We give one calculation for the even potential in (12), under its invariant-direction and reflection assumptions. DPO and IPO are obtained by substituting their fourth derivatives. Write $\phi_4 = \Phi^{(4)}(0)$ and $h(s) = H + H_2s^2 + O(s^4)$. Analyticity and evenness give

$$\Phi'(u) = u/4 + \phi_4u^3/6 + O(u^5), \qquad \partial_w L_\Phi(w;s) = -b_e(s) + h(s)w + \frac{\phi_4}{6}\mathbb{E}_s d_e^4\, w^3 + O(w^5).$$

Finite support justifies termwise expectation. Reflection and initialization at zero make every iterate odd in s. The restricted loss symmetry also makes $\phi_4\mathbb{E}_s d_e^4$ even in s, so its variation contributes only at order s^5. For each fixed inner index write

$$w_{\Phi,j}(s) = -v_j s + c_{\Phi,j}s^3 + O(s^5), \qquad v_0 = c_{\Phi,0} = 0.$$

Substitution in a GD step and collection of linear and cubic powers give

$$\begin{aligned}v_{j+1} &= rv_j + \eta A, \\ c_{\Phi,j+1} &= rc_{\Phi,j} + \eta B + \eta H_2 v_j + \frac{\eta\phi_4 Q_4}{6}v_j^3.\end{aligned}$$

Indeed, $w_j^3 = -v_j^3s^3 + O(s^5)$, and GD subtracts that gradient term. For DPO, $\phi_4 = -1/8$, giving the negative coefficient $-\eta Q_4 v_j^3/48$; IPO has $\phi_4 = 0$. The linear recursion yields $v_j = (A/H)(1-r^j)$. Summing the cubic recursion first for IPO and then for its difference from Φ gives

$$\begin{aligned}c_{\mathrm{Q},K} &= \eta\sum_{j=0}^{K-1} r^{K-1-j}\left\{B + \frac{H_2A}{H}(1-r^j)\right\} \\ &= \frac{B}{H}d_K + \frac{H_2A}{H^2}\{d_K - K(1-r)r^{K-1}\}, \\ c_{\Phi,K} - c_{\mathrm{Q},K} &= \frac{\eta\phi_4Q_4A^3}{6H^3}\sum_{j=0}^{K-1} r^{K-1-j}(1-r^j)^3.\end{aligned}$$

Here $\sum r^{K-1-j} = d_K/(1-r)$, $\sum r^{K-1-j}r^j = Kr^{K-1}$, and $\eta H = 1 - r$. The common outer linear coefficient $1 - v_K/\beta$ equals -1 at $\beta_{c,K} = Ad_K/(2H)$. Dividing the cubic coefficients by this scale,

$$\begin{aligned}\chi_{\mathrm{Q},K} &= \frac{2B}{A} + \frac{2H_2}{H}(1-u_K), \\ \chi_{\Phi,K} - \chi_{\mathrm{Q},K} &= \frac{\phi_4Q_4A^2}{3H^3}\frac{1-r}{d_K}\sum_{j=0}^{K-1} r^{K-1-j}(1-r^j)^3 = 8\phi_4\Lambda W_K = -\omega_\Phi\Lambda W_K.\end{aligned}$$

This proves (16), and its DPO specialization is (73).

We give the monotonicity details, since a negative limiting coefficient alone does not identify a single sign transition. The identity

$$\frac{1}{u_K} = \frac{1}{K}\sum_{j=0}^{K-1} r^{-j}$$

shows that u_K strictly decreases: adding the next term, which exceeds the preceding average, strictly increases this average. Also $u_1 = 1$ and $u_K \to 0$. Let J_K have distribution

$$\mathbb{P}(J_K = j) = \frac{(1-r)r^{K-1-j}}{1-r^K}, \qquad 0 \le j < K.$$

For $0 \le m < K$ its distribution functions obey

$$\mathbb{P}(J_{K+1} \le m) = \frac{r(1-r^K)}{1-r^{K+1}}\mathbb{P}(J_K \le m) < \mathbb{P}(J_K \le m).$$

The prefactor is less than one because $r - r^{K+1} < 1 - r^{K+1}$. For $g(j) = (1-r^j)^3$, the finite summation identity

$$\mathbb{E}g(J) = g(0) + \sum_{m=0}^{K-1}(g(m+1) - g(m))\mathbb{P}(J > m)$$

implies that $W_K = \mathbb{E}g(J_K)$ strictly increases. For every fixed m, $\mathbb{P}(J_K \le m) \to 0$. Given $\epsilon > 0$, choose m with $g(m) > 1 - \epsilon$; then $W_K \ge g(m)\mathbb{P}(J_K \ge m)$ gives $\liminf W_K \ge 1 - \epsilon$. Since $W_K < 1$, this proves $W_K \to 1$.

Under the sign condition of Theorem 3.2, $\omega_\Phi > 0$ and $\chi_{Q,K} = \chi_{Q,\infty} - 2H_2 u_K/H \ge \chi_{Q,\infty} > 0$. This coefficient is nonincreasing, whereas $\omega_\Phi \Lambda W_K$ strictly increases. Consequently $\chi_{\Phi,K}$ strictly decreases from $2B/A > 0$ to $\chi_{Q,\infty} - \omega_\Phi\Lambda < 0$. There is a finite first negative coefficient, with at most one exactly zero coefficient preceding it. Setting $\omega_\Phi = 1$ and using (74) proves all assertions of Theorem B.1. These are longitudinal statements; other directions and finite-amplitude cycles require separate checks. □

D.8 The common amplitude-curve argument

Proof of Lemma A.14. Oddness of ν gives an even analytic extension of $\mathcal{A}$ at zero with value β_c. Its positive second derivative implies $\mathcal{A}(s_0) > \beta_c$ for some $s_0 > 0$. The assumed tail limit provides $R > s_0$ such that $\mathcal{A}(s) < \beta_c$ for $s \ge R$. The maximum on $[0, R]$ is therefore global and occurs at a point $a_* \in (0, R)$, with value $m > \beta_c$. The function is nonconstant because its tail limit differs from its value at zero. Analyticity on the connected domain excludes a flat Taylor series at a_*. The first nonconstant term at a maximum must have even order and a negative coefficient. Thus, for some integer $j \ge 1$ and $b > 0$,

$$\mathcal{A}(a_* + v) = m - bv^{2j} + O(v^{2j+1}), \qquad \mathcal{A}'(a_* + v) = -2jbv^{2j-1} + O(v^{2j}).$$

Choose $r \in (0, a_*/2)$ small enough that the derivative is positive on $(a_* - r, a_*)$, negative on $(a_*, a_* + r)$, and

$$\mathcal{A}(s) > \beta_c, \qquad |s\mathcal{A}'(s)|/\mathcal{A}(s) < 1 \quad (|s - a_*| \le r). \tag{124}$$

The latter inequalities follow from continuity, $m > \beta_c$, and $\mathcal{A}'(a_*) = 0$. Set

$$e = \frac{1}{2}\min\{m - \beta_c, m - \mathcal{A}(a_* - r), m - \mathcal{A}(a_* + r)\} > 0.$$

Every $\beta \in (m - e, m)$ has exactly one intersection with the amplitude curve in each of these two intervals, by strict monotonicity and the intermediate value theorem. Denote them a_-, a_+. Both tend to a_* as $\beta \uparrow m$. At either intersection, $\nu(a) = 2\beta a$, hence $g_\beta(a) = -a$ and oddness gives $g_\beta(-a) = a$. Since $a > 0$, these are genuine two-cycles. Differentiating $\nu(s) = 2s\mathcal{A}(s)$ gives

$$g'_\beta(a) = 1 - \frac{2\mathcal{A}(a) + 2a\mathcal{A}'(a)}{\beta} = -1 - \frac{2a\mathcal{A}'(a)}{\beta}. \tag{125}$$

This is less than -1 at a_- and in $(-1,1)$ at a_+ by (124). The derivative is even, so the second-iterate multipliers are the corresponding squares. A continuous derivative bounded below one in modulus on a neighborhood gives local attraction by integration along segments; a multiplier greater than one gives local repulsion. Positivity of $\nu(s)$ on $s>0$ and oddness make zero the only fixed point. Its derivative $1-2\beta_c/\beta$ belongs to $(-1,1)$.

For the two-dimensional extension, reflection in the second coordinate makes the first component even in that coordinate, and invariance of the first axis makes the other mixed derivative zero. The cycle Jacobians are diagonal. Reflection in the first coordinate makes their diagonal entries agree at reflected cycle points. The eigenvalues of the second iterate are therefore the squares of the longitudinal and transverse entries. This gives a saddle at a_- and attraction at a_+. More explicitly, both derivative norms at the attracting cycle points p_+, p_- are below one. Choose disjoint equal-radius closed balls around them where these norms are at most $q<1$. Then $\|F(\theta)-p_\mp\| \le q\|\theta-p_\pm\|$ in the respective ball. The balls map alternately into each other and their interiors form an open attracting set. No assumption on the order $2j$ was used beyond its finiteness.

If $\mathcal{A}''(a_*)<0$, the collision at $\beta=m$ is nondegenerate. At that point $g(a_*)=-a_*$, $g'(\pm a_*)=-1$, and $g''(a_*)=-2a_*\mathcal{A}''(a_*)/\beta$. Oddness gives $g''(-a_*)=-g''(a_*)$, while $\partial_\beta g(\pm a_*)=\pm 2a_*/\beta$. The chain rule yields

$$\partial_\beta\{g(g(a_*))-a_*\}=-4a_*/\beta\neq 0, \qquad \partial_s^2\{g(g(s))-s\}|_{s=a_*}=4a_*\mathcal{A}''(a_*)/\beta\neq 0.$$

These are the two nondegeneracy conditions for the fixed-point fold of the second iterate. This calculation is local and also applies to a strictly concave local maximum that is not a global maximum. □

D.9 Uniform smooth approximation and deep inner limits

Proof of Lemma A.15. We first give the derivative bounds needed for finite-time Euler approximation. Write $v'=f(s,v)$ for the scalar inner flow. For $Y_m=\partial_s^m v$, the chain rule gives

$$\partial_s^m f(s,v(s))=\sum_{p+\sum_{j=1}^m jn_j=m}\frac{m!}{p!\prod_{j=1}^m n_j!(j!)^{n_j}}(\partial_s^p\partial_v^{\sum n_j}f)(s,v)\prod_{j=1}^m Y_j^{n_j}. \tag{126}$$

For $m\ge 1$, the term $p=0, n_m=1$ is f_vY_m; every other term depends only on $Y_0=v, Y_1,\dots,Y_{m-1}$. Call their sum P_m. Thus

$$Y_0'=f(s,Y_0), \qquad Y_m'=f_v(s,Y_0)Y_m+P_m(s,Y_0,\dots,Y_{m-1}), \qquad Y_m(0)=0.$$

Differentiating the Euler recurrence yields precisely the Euler recurrence for this finite triangular system, with the same step size.

On a fixed time interval $[0,T_1+1]$, the DPO gradient is bounded by the feature-difference bound, so Y_0 and its discrete counterpart stay in a fixed bounded interval. All derivatives of the analytic field are bounded on this interval and $|s|\le s_0$. Suppose that the derivatives of orders below m have already been bounded uniformly. Formula (126) gives a bound $|P_m|\le P_m^*$ and $|f_v|\le L_0$. Discrete induction and the integrating-factor inequality give, respectively,

$$|Y_{m,j}|\le \eta P_m^*\sum_{i=0}^{j-1}(1+\eta L_0)^i\le P_m^*(T_1+1)e^{L_0(T_1+1)},$$

$$|Y_m(t)|\le P_m^*\int_0^t e^{L_0(t-u)}du\le P_m^*(T_1+1)e^{L_0(T_1+1)}.$$

These expressions also apply when $L_0=0$. Induction through order nine bounds every coordinate of the augmented state Y. Choose a compact box containing these trajectories and let B, L bound the norm and Lipschitz constant of its smooth augmented field $G(s,Y)$. Enlarging the box slightly also contains each exact short flow segment starting on a trajectory. For such a segment,

$$\begin{aligned}\|Y(t+\eta)-Y(t)-\eta G(s,Y(t))\| &\le \int_0^\eta L\|Y(t+u)-Y(t)\|du\\ &\le \int_0^\eta LBu\,du=LB\eta^2/2.\end{aligned}$$

The error e_j of the augmented Euler scheme therefore obeys

$$e_{j+1} \le (1+\eta L)e_j + LB\eta^2/2, \qquad e_K \le \frac{B\eta}{2}\{(1+\eta L)^K - 1\} \le \frac{B\eta}{2}(e^{LT_1} - 1).$$

This proves (71), uniformly in s and $K\eta \in [T_0, T_1]$. Including the first derivatives in both outer coordinates gives the asserted two-dimensional C^1 bound by the same argument. These constants are uniform in time on the compact interval; they are not constants for just one chosen horizon.

For the mixture, choose $\lambda_0 < 1/2$ sufficiently small that $1 \le \alpha_\lambda < 4$. On a compact interval of s, its inner Hessian is bounded below by $(1-\lambda_0)\min_{|s|\le s_0} h_{\alpha_\lambda}(s) > 0$ and above by $1/2$, uniformly in the inner coordinate. Shrinking λ_0 if needed makes the minimum uniform in λ. The scalar update $T_\lambda(s,v) = v + f_{\Phi_\lambda}(s,v)$ then satisfies

$$0 < \partial_v T_\lambda \le q < 1.$$

A common V bounds its fixed point: for example, since $\Phi'_\lambda(V) \ge (1-\lambda_0)V/4$, any $V > 2c/(1-\lambda_0)$ works in the sign argument for (134). The zero-initialized iterates stay between zero and that fixed point, and hence in $[-V, V]$. The implicit function theorem applies to $f_{\Phi_\lambda}(s,v) = 0$ with a derivative bounded away from zero; its unique solution is jointly smooth and all of its first nine s derivatives are bounded on a smaller compact parameter rectangle.

Differentiate the iterated update with (126), now applied to T_λ. For derivative order m, the highest derivative is multiplied by $\partial_v T_\lambda \le q$, and all other terms are bounded functions of lower derivatives. Induction therefore gives $|Y_{m,K}| \le q|Y_{m,K-1}| + P^*_m$ and a uniform bound $P^*_m/(1-q)$, after enlarging P^*_m to include the zero-order case. Compare these recurrences with the derivatives of the fixed point equation. Let $E_{m,K}$ be the supremum difference at order m. Bounded derivatives on the compact boxes and the mean value inequality give

$$E_{0,K} \le 2Vq^K, \qquad E_{m,K+1} \le qE_{m,K} + C_m \sum_{i=0}^{m-1} E_{i,K}.$$

If $E_{i,K} \le C(K+1)^{m-1}q^K$ for $i < m$, summing this last recurrence yields

$$E_{m,K} \le q^K E_{m,0} + C'_m \sum_{j=0}^{K-1} q^{K-1-j}(j+1)^{m-1}q^j \le C''_m (K+1)^m q^K.$$

Induction through $m = 9$ proves (72).

For the final assertion, oddness and the uniform derivative bounds give $\nu(s) = v_1 s + v_3 s^3 + v_5 s^5 + R_7(s)$ with $R_7^{(j)}(s) = O(|s|^{7-j})$ for $j = 0, 1, 2$, uniformly. Dividing by $2s$ gives the amplitude expansion with $\beta_c = v_1/2$, $\chi = -2v_3/v_1$, and $\psi = -2v_5/v_1$. In the uses here the positive depths/times stay away from zero, so v_1 has a uniform positive lower bound. If $E(s) = R_7(s)/(2s)$, then $E = O(s^6)$, $E_s = O(s^5)$, and $E_{ss} = O(s^4)$. For $z = s^2 > 0$,

$$E_z = E_s/(2s) = O(z^2), \qquad E_{zz} = E_{ss}/(4s^2) - E_s/(4s^3) = O(z).$$

The limits at zero follow by the analytic even extension. This proves the uniform differentiated remainder estimates. □

D.10 An explicit scalar model region and its attracting cycles

Proof of Proposition B.5. Write $C = \cosh s$, $S = \sinh s$, and $Z = \alpha + 2C$. The three probabilities are $e^{-s}/Z, \alpha/Z, e^s/Z$. Since $\sigma(g) - 1/2 = c/2$, the two orientations of each pair with one zero and one endpoint contribute equally to the preference gradient. The zero–positive pair contributes $-\alpha c e^s/Z^2$, and the zero–negative pair contributes $\alpha c e^{-s}/Z^2$. Pairs between endpoints have equal rewards. Consequently,

$$b(s) = -\frac{2\alpha c S}{Z^2}, \qquad \mathbb{E}_s X = \frac{2S}{Z}, \qquad \mathrm{Var}_s\, X = \frac{2C}{Z} - \frac{4S^2}{Z^2} = \frac{2(\alpha C + 2)}{Z^2}. \tag{127}$$

Here $C^2 - S^2 = 1$ was used in the final equality. Thus the scalar inner curvature at zero is $h(s) = \mathbb{E}_s d^2/4 = (\alpha C + 2)/Z^2$. Expanding $S = s + s^3/6 + O(s^5)$ and $Z^{-2} = (\alpha+2)^{-2}\{1 - 2s^2/(\alpha+2) + O(s^4)\}$ gives

$$b(s) = -\frac{2\alpha c}{(\alpha+2)^2}s + \frac{2\alpha c}{(\alpha+2)^2}\left(\frac{2}{\alpha+2} - \frac{1}{6}\right)s^3 + O(s^5).$$

At zero, $\mu_2 = \mu_4 = 2/(\alpha+2)$, so Lemma A.12 supplies H, H_2, Q_4. This proves every coefficient in (84). Direct substitution gives

$$\chi_{\mathrm{Q},\infty} = \frac{10-\alpha}{3(\alpha+2)} + \frac{\alpha-4}{\alpha+2} = \frac{2(\alpha-1)}{3(\alpha+2)},$$
$$\Lambda = \frac{4(\alpha+8)}{(\alpha+2)^2}\,\frac{4\alpha^2c^2}{(\alpha+2)^4}\,\frac{(\alpha+2)^3}{24} = \frac{2\alpha^2(\alpha+8)c^2}{3(\alpha+2)^3}. \tag{128}$$

For $1 < \alpha \le 4$, $H_2 \le 0$ and $\chi_{\mathrm{Q},\infty} > 0$. The inequality $\chi_{\mathrm{Q},\infty} < \Lambda$ is exactly (77). Its lower bound on c^2 is strictly between zero and one, since

$$\alpha^2(\alpha+8) - (\alpha-1)(\alpha+2)^2 = 5\alpha^2 + 4 > 0.$$

This also proves the nonempty-interior assertion. Reflection of x preserves both its base weight and its reward, so it gives the symmetry required by Theorem B.1. Moreover $0 < H = 1/(\alpha+2) < 1$. That theorem applies at $\eta = 1$ and proves the assertions about the cubic coefficients.

We next obtain global control of the frozen DPO displacement. The probabilities of absolute comparison differences one and two are respectively $4\alpha C/Z^2$ and $2/Z^2$. The remaining probability has zero difference. Since $\sigma(u) - 1/2 = \frac{1}{2}\tanh(u/2)$, the centered logistic gradient evaluated at $w = -v$ equals $-\{2\alpha C\tanh(v/2) + 2\tanh v\}/Z^2$. It follows that $\nu_j(s) = -w_{\mathrm{D},j}(s)$ satisfies

$$\nu_0 = 0, \qquad \nu_{j+1} = \nu_j + f_{\alpha,c}(s,\nu_j), \qquad f_{\alpha,c}(s,v) = \frac{2\{\alpha cS - \alpha C\tanh(v/2) - \tanh v\}}{Z^2}. \tag{129}$$

Its derivative in the inner coordinate is

$$\partial_v f_{\alpha,c}(s,v) = -\frac{\alpha C\,\mathrm{sech}^2(v/2) + 2\,\mathrm{sech}^2 v}{Z^2} < 0.$$

Because $X \in [-1,1]$, $\mathrm{Var}_s X \le 1$, and hence $-\partial_v f_{\alpha,c} \le h(s) = \mathrm{Var}_s X/2 \le 1/2$. The unit-step map $v \mapsto v + f_{\alpha,c}(s,v)$ is therefore strictly increasing with derivative at least $1/2$. For $s > 0$, $f_{\alpha,c}(s,0) > 0$, whereas

$$f_{\alpha,c}(s,g) = \frac{2\{\alpha c(S-C) - \tanh g\}}{Z^2} < 0,$$

using $\tanh(g/2) = c$ and $S - C = -e^{-s} < 0$. Strict decrease of $f_{\alpha,c}$ gives exactly one zero $\nu_\infty(s) \in (0,g)$. The increasing unit-step map fixes this zero and maps every $0 \le v < \nu_\infty(s)$ strictly between v and $\nu_\infty(s)$. Induction proves

$$0 < \nu_j(s) < \nu_{j+1}(s) < \nu_\infty(s) < g \quad (s > 0,\ j \ge 1). \tag{130}$$

In particular, the finite-depth maps have no nonzero fixed point. All finite-depth displacements are analytic and odd in s.

For completeness, global convergence of each frozen inner problem holds beyond these zero-initialized iterates. The support spans the scalar feature space and every BT probability is strictly between zero and one. Lemma A.1's general finite-family assertion gives strict convexity, coercivity, and a unique finite minimizer. Here both Hessians are at most $1/2$, so the descent lemma at unit step gives $L_X(w - L'_X(w); s) \le L_X(w;s) - 3|L'_X(w;s)|^2/4$. The iterates stay in a compact sublevel set and their gradients tend to zero. Every cluster point is the unique minimizer, which proves convergence of the whole sequence from any finite inner initialization.

Fix a depth with $\chi_{\mathrm{D},K} < 0$. Define $\mathcal{A}(s) = \nu_K(s)/(2s)$ for $s > 0$ and extend it to zero. The linear recursion in Theorem B.1 and its cubic expansion give

$$\mathcal{A}(s) = \beta_{c,K} - \frac{\beta_{c,K}\chi_{\mathrm{D},K}}{2}s^2 + O(s^4), \qquad \beta_{c,K} = \frac{\alpha c}{\alpha+2}\left(1 - \left(1 - \frac{1}{\alpha+2}\right)^K\right). \tag{131}$$

The expansion gives $\mathcal{A}''(0) > 0$, and (130) gives $\nu_K(s) > 0$ and $\nu_K(s)/s \to 0$. All hypotheses of Lemma A.14 hold. It supplies the two cycles, their scalar stability, a stable unique fixed point, and a nonempty scale interval $\mathcal{B}_{\alpha,c,K}$ above $\beta_{c,K}$. No transverse claim is needed for this scalar proposition.

A loss-independent IPO comparison. The following argument uses only $1 \le \alpha \le 4$, $0 < c < 1$, and $\beta > \beta_{c,K} = Ad_K/(2H)$; it does not use the DPO cubic sign condition. It will therefore also supply the comparison for other losses. From (127), its exact inner minimizer is

$$w_{\mathrm{Q}}^*(s) = \frac{b(s)}{h(s)} = -\frac{2\alpha c \sinh s}{\alpha \cosh s + 2}. \tag{132}$$

For $s > 0$ and $\alpha \ge 1$,

$$\frac{\sinh s}{s} \le \frac{\alpha \cosh s + 2}{\alpha + 2}. \tag{133}$$

To prove it, multiply by the positive denominator and define $E_\alpha(s) = \alpha s \cosh s + 2s - (\alpha+2)\sinh s$. Then

$$E_\alpha(s) = E_1(s) + (\alpha - 1)(s\cosh s - \sinh s), \qquad E_1(s) = s\cosh s + 2s - 3\sinh s.$$

The last parenthesis is zero at zero and has derivative $s \sinh s \ge 0$. Also $E_1(0) = E_1'(0) = E_1''(0) = 0$ and $E_1'''(s) = s\sinh s \ge 0$. Three integrations on $[0, s]$ yield $E_1(s) \ge 0$. This proves (133), and therefore $|w_{\mathrm{Q}}^*(s)|/|s| \le 2\alpha c/(\alpha + 2) = A/H$.

Viewing h as a function of $C \ge 1$ gives

$$\frac{d}{dC}\frac{\alpha C + 2}{(\alpha + 2C)^2} = \frac{\alpha^2 - 2\alpha C - 8}{(\alpha + 2C)^3} \le 0,$$

because the numerator is at most $\alpha^2 - 2\alpha - 8 = (\alpha - 4)(\alpha + 2) \le 0$. Thus $0 < h(s) \le H < 1$. The linear IPO recursion can be summed exactly:

$$w_{\mathrm{Q},L}(s) = \{1 - (1 - h(s))^L\} w_{\mathrm{Q}}^*(s).$$

For $1 \le L \le K$, $d_L = 1 - (1 - H)^L \le d_K$, so

$$0 < \frac{|w_{\mathrm{Q},L}(s)|}{\beta|s|} \le \frac{Ad_L}{H\beta} \le \frac{2\beta_{c,K}}{\beta} < 2 \quad (s \ne 0).$$

The displacement points toward zero. Define the continuous positive ratio $q_L(s) = |w_{\mathrm{Q},L}(s)|/(\beta|s|)$ for $s \ne 0$ and extend it by $q_L(0) = Ad_L/(H\beta) > 0$. For an initial bound $R < \infty$, continuity on $[-R, R]$ and the strict upper bound above imply $\rho_L := \max_{|s| \le R} |1 - q_L(s)| < 1$. Every IPO outer step satisfies $|F_{\mathrm{Q},L}(s)| \le \rho_L |s|$ in that interval, so induction proves invariance and geometric convergence to zero. DPO-1 equals IPO-1, giving its claimed convergence as well. Finally

$$J_g(s) = -\frac{2g\cosh s}{\alpha + 2\cosh s}, \qquad \frac{d}{dC}\left(-\frac{2gC}{\alpha + 2C}\right) = -\frac{2\alpha g}{(\alpha + 2C)^2} < 0.$$

It is uniquely maximized at zero and strictly improves when $|s|$ strictly decreases. Subtracting its value at $s = a$ from that at zero gives (85), completing all assertions. □

D.11 Loss potentials, depth, and scalar coexistence

Proof of Theorem 3.2. The common calculation in Appendix D.7 proves the local formula, the signs, and monotonicity for every potential under this theorem's hypotheses. We now verify the additional global conditions in the three-response model; none is supplied by the local calculation alone.

For the three-response family, write $S = \sinh s$, $C = \cosh s$, and $Z = \alpha + 2C$. Enumeration of the zero–endpoint comparisons gives $b(s) = -2\alpha c S/Z^2$, independently of the fitting loss. The moments and local preference coefficients are therefore (84). In particular,

$$H_2 = \frac{\alpha - 4}{2(\alpha + 2)^2} \le 0, \quad \chi_{\mathrm{Q},\infty} = \frac{2(\alpha - 1)}{3(\alpha + 2)} > 0, \quad \Lambda = \frac{2\alpha^2(\alpha + 8)c^2}{3(\alpha + 2)^3}.$$

The condition $\chi_{\mathrm{Q},\infty} < \omega_\Phi \Lambda$ becomes (17) after multiplication by the positive factor $3(\alpha + 2)^3/2$. At $\eta = 1$, $H = 1/(\alpha + 2) \in (0, 1)$, so the local result supplies a finite first negative coefficient.

To convert it into a cycle, we verify the required global facts for this potential. The probabilities of $|d| = 1$ and $|d| = 2$ are $4\alpha C/Z^2$ and $2/Z^2$. Oddness of Φ' implies

$$\mathbb{E}_s[d\Phi'(dw)] = \frac{4}{Z^2}\{\alpha C\Phi'(w) + \Phi'(2w)\}.$$

Consequently $\nu_j = -w_{\Phi,j}$ satisfies

$$\nu_{j+1} = \nu_j + f_\Phi(s, \nu_j), \quad \nu_0 = 0, \qquad f_\Phi(s, v) = \frac{2\alpha cS - 4\alpha C\Phi'(v) - 4\Phi'(2v)}{Z^2}. \tag{134}$$

Its inner derivative obeys

$$0 < -\partial_v f_\Phi(s, v) = \frac{4}{Z^2}\{\alpha C\Phi''(v) + 2\Phi''(2v)\} \le \frac{\alpha C + 2}{Z^2} = h(s) = \frac{1}{2}\operatorname{Var}_s X \le \frac{1}{2}.$$

Thus $v \mapsto v + f_\Phi(s, v)$ has derivative in $[1/2, 1)$. Choose a finite $V > 0$ with $\Phi'(V) > c/2$, which exists by the assumed limiting slope. For $s > 0$, $f_\Phi(s, 0) > 0$, while

$$Z^2 f_\Phi(s, V) < 2\alpha c(S - C) - 4\Phi'(2V) < 0.$$

There is exactly one root $\nu_\infty(s) \in (0, V)$, by strict decrease of $f_\Phi(s, \cdot)$. The increasing unit-step map fixes this root, so induction gives

$$0 < \nu_j(s) < \nu_{j+1}(s) < \nu_\infty(s) < V \qquad (j \ge 1).$$

Reflection gives the corresponding negative-state assertions and makes every finite-depth displacement analytic and odd. In particular, the outer map has no nonzero fixed point.

Each frozen loss is strictly convex, since its scalar Hessian is positive. Its derivative has opposite strict signs at $-V$ and V, and is increasing; integrating these bounds outside $[-V, V]$ proves coercivity. Its Hessian is at most $1/2$. Unit-step GD therefore satisfies $L(w - L'(w)) \le L(w) - 3|L'(w)|^2/4$. The iterates remain in a compact sublevel set, and the sum of squared gradients is finite. Every cluster point is the unique zero of the derivative. Hence the whole sequence converges to the unique finite minimizer. This argument also applies to IPO, whose positive Hessian is the constant $h(s) \le 1/2$.

Fix any depth with $\chi_{\Phi,K} < 0$. Its amplitude function $\mathcal{A}(s) = \nu_K(s)/(2s)$ has analytic extension

$$\mathcal{A}(s) = \beta_{c,K} - \frac{\beta_{c,K}\chi_{\Phi,K}}{2}s^2 + O(s^4).$$

The bounded positive displacement gives the tail and sign conditions of Lemma A.14, and the negative cubic coefficient gives $\mathcal{A}''(0) > 0$. That lemma supplies an attracting cycle, a smaller repelling cycle, and a stable unique fixed point throughout a nonempty open interval above $\beta_{c,K}$. The model-specific work here is the global displacement bound for Φ and the comparison with IPO, not a new maximum argument.

The loss-independent comparison in Appendix D.10 applies: $1 < \alpha \le 4$, $0 < c < 1$, and the scales just obtained satisfy $\beta > \beta_{c,K}$. In particular, for every $1 \le L \le K$, it gives

$$0 < \frac{|w_{Q,L}(s)|}{\beta|s|} \le \frac{Ad_L}{H\beta} \le \frac{2\beta_{c,K}}{\beta} < 2 \qquad (s \ne 0),$$

and a continuous, strictly positive extension of this ratio at zero. The compact-interval contraction proved there yields global IPO convergence. Its reward calculation is also loss-independent: zero is uniquely optimal, and either cycle has the positive gap (85). This proves the comparison and reward claims.

Finally, the mixture potential is

$$\Phi_\lambda(u) = (1 - \lambda)u^2/8 + \lambda \log(2\cosh(u/2)).$$

Its curvature is $(1 - \lambda)/4 + \lambda \operatorname{sech}^2(u/2)/4 \in (0, 1/4]$, and $\Phi_\lambda^{(4)}(0) = -\lambda/8$. Its limiting positive slope is infinite for $\lambda < 1$ and $1/2$ for $\lambda = 1$, exceeding $c/2$ for every $c < 1$. For $\alpha > 1$, the identity

$$\alpha^2(\alpha + 8) - (\alpha + 2)^2 = (\alpha - 1)(\alpha^2 + 8\alpha + 4) > 0$$

implies $T(\alpha) < \alpha - 1$. Hence the explicit open rectangle

$$1 + \frac{\lambda}{16} < \alpha < 1 + \frac{\lambda}{8}, \qquad \frac{1}{2} < c < \frac{3}{4} \tag{135}$$

satisfies $T(\alpha) < \lambda/8 < \lambda c^2$ and $1 < \alpha < 4$. It is nonempty for every $0 < \lambda \le 1$. For any compact logit interval and any fixed derivative order m, $\|\Phi_\lambda - \Phi_Q\|_{C^m} = \lambda\|\Phi_D - \Phi_Q\|_{C^m} \to 0$. Thus this convergence of losses does not give safety uniformly over the constructed model regions and inner depths. □

D.12 Quantifying the limit of a small logistic component

Proof of Proposition B.6. The loss-class formula specializes to

$$\chi_{\lambda,K} = -\Delta_\lambda + C_\lambda u_K + \lambda\Lambda_\lambda(1 - W_K).$$

For $0 < \lambda \le 1$, $1 < \alpha_\lambda < 2$ and the identity used in (135) gives $T(\alpha_\lambda) < \alpha_\lambda - 1 = \vartheta\lambda < c^2\lambda$. Thus $\Delta_\lambda > 0$ and $C_\lambda > 0$ throughout this full range. Write $p_{K,j} = (1-r)r^{K-1-j}/(1-r^K)$ for $0 \le j < K$; these nonnegative weights sum to one. For $t \in [0,1]$, $t \le 1-(1-t)^3 = 3t - 3t^2 + t^3 \le 3t$. Taking $t = r^j$ and averaging gives

$$u_K = \sum_{j=0}^{K-1} p_{K,j} r^j \le 1 - W_K \le 3 \sum_{j=0}^{K-1} p_{K,j} r^j = 3u_K.$$

Here the sum is exactly $K(1-r)r^{K-1}/(1-r^K)$ because each unnormalized summand equals $(1-r)r^{K-1}$. Multiplication by $\lambda\Lambda_\lambda > 0$ proves (87). Expansion at $\alpha = 1$ gives

$$r_\lambda = \frac{2}{3} + O(\lambda), \qquad C_\lambda = 1 + O(\lambda), \qquad \Lambda_\lambda = \frac{2c^2}{9} + O(\lambda),$$
$$\frac{2(\alpha_\lambda - 1)}{3(\alpha_\lambda + 2)} = \frac{2\vartheta}{9}\lambda + O(\lambda^2), \qquad \Delta_\lambda = D_0\lambda + O(\lambda^2) > 0.$$

Since $\vartheta > 0$ and $\alpha_\lambda < 4$ for small λ, the monotonicity part of Theorem 3.2 applies: the cubic coefficients strictly decrease from a positive value to $-\Delta_\lambda < 0$.

Put $L = \log(1/\lambda)$ and $\ell_0 = \log(3/2)$. For integers $K = (L + \log L)/\ell_0 + t$ with bounded t, one has $K = O(L)$ and

$$\frac{r_\lambda^K}{(2/3)^K} = \exp\{K \log(r_\lambda/(2/3))\} = 1 + o(1), \qquad 1 - r_\lambda^K \to 1.$$

Consequently

$$\frac{u_K}{\lambda} = \frac{K(1-r_\lambda)}{r_\lambda(1-r_\lambda^K)} \frac{r_\lambda^K}{\lambda}$$
$$= \left(\frac{K}{2} + o(K)\right) \frac{e^{-\ell_0 K}}{\lambda}(1 + o(1)) = \frac{e^{-\ell_0 t}}{2\ell_0} + o(1).$$

For a ceiling or floor in this expression, t changes by less than one. Choose a fixed negative offset with $e^{-\ell_0(t+1)}/(2\ell_0) > 2D_0$, and a fixed positive offset with $e^{-\ell_0 t}/(2\ell_0) < D_0/2$. The two finite-depth bounds then give a positive cubic coefficient at the lower integer and a negative one at the upper integer for small λ. Strict monotonicity places K_{sgn} between them, proving (88), including its bounded integer-rounding error.

For $u_{K_\lambda} = o(\lambda)$, the same bounds give

$$\chi_{\lambda,K_\lambda} = -D_0\lambda + o(\lambda). \tag{136}$$

The example depth in the statement has $r_\lambda^{K_\lambda} = O(\lambda^2)$, because $K_\lambda O(\lambda) \to 0$, and hence $u_{K_\lambda} = O(L\lambda^2) = o(\lambda)$. Every such choice tends to infinity. The second part of Lemma A.15 implies convergence of its fifth coefficient to that of the exact IPO map at $\lambda = 0, \alpha = 1$. At that endpoint, $\beta_0 = c/3$ and

$$F_{0,\infty}(s) = s - \frac{6\sinh s}{\cosh s + 2} = -s + \frac{s^5}{90} + O(s^7).$$

For the coefficient, use $\sinh s = s + s^3/6 + s^5/120 + O(s^7)$ and $\cosh s + 2 = 3 + s^2/2 + s^4/24 + O(s^6)$. If their ratio is $q_1 s + q_3 s^3 + q_5 s^5 + O(s^7)$, multiplication gives $3q_1 = 1$, $3q_3 + q_1/2 = 1/6$, and $3q_5 + q_3/2 + q_1/24 = 1/120$. Thus $q_1 = 1/3$, $q_3 = 0$, $q_5 = -1/540$, proving the display. It follows that

$$\beta_{c,K_\lambda} \to \beta_0 = c/3, \qquad \psi_{\lambda,K_\lambda} \to \psi_0 = 1/90 > 0.$$

Write $\mathcal{H}_\lambda(z) = \mathcal{A}_{\lambda,K_\lambda}(\sqrt{z})$. The uniform differentiated remainder in the lemma and (136) show that, on every bounded interval of $y \geq 0$,

$$Q_\lambda(y) := \frac{\mathcal{H}_\lambda(\lambda y) - \beta_{c,K_\lambda}}{\lambda^2} \longrightarrow Q_0(y) = \frac{\beta_0}{2}(D_0 y - \psi_0 y^2) \quad \text{in } C^2. \tag{137}$$

Indeed the linear-in-z term contributes $-(\beta_{c,K_\lambda}/2)(\chi_{\lambda,K_\lambda}/\lambda)y$, the quadratic term contributes $-(\beta_{c,K_\lambda}/2)\psi_{\lambda,K_\lambda}y^2$, and the remainder and its first two y derivatives are $O(\lambda)$ on bounded intervals. No differentiability of the integer choice K_λ as a function of λ is assumed.

The limiting quadratic is strictly concave, with its unique maximum at $y_{\mathrm{f}}^0 = D_0/(2\psi_0) > 0$ and value $\beta_0 D_0^2/(8\psi_0) > 0$. On an interval containing zero and $2D_0/\psi_0$, C^2 convergence makes $Q_\lambda'' < 0$ and gives opposite signs of Q_λ' on either side of y_{f}^0. Thus there is a unique local maximum $y_{\mathrm{f}} \to y_{\mathrm{f}}^0$, with value tending to the stated positive limit. Setting $a_{\mathrm{f}}^2 = \lambda y_{\mathrm{f}}$ and $\beta_{\mathrm{f}} = \mathcal{H}_\lambda(a_{\mathrm{f}}^2)$ yields

$$\frac{D_0}{2\psi_0} = 10(c^2 - \vartheta), \qquad \frac{\beta_0 D_0^2}{8\psi_0} = \frac{c}{3}\frac{4(c^2-\vartheta)^2}{81}\frac{90}{8} = \frac{5c(c^2-\vartheta)^2}{27}.$$

This proves (89). The maximum has strictly negative second derivative for each small positive λ. The local derivative identities in the proof of Lemma A.14 therefore verify that it is a nondegenerate fold of cycles.

At a relative level ρ of this maximum, the two roots of the limiting quadratic have values $y_\pm^0 = D_0(1 \pm \sqrt{1-\rho})/(2\psi_0)$. Their derivatives are nonzero; monotonicity and C^1 convergence give roots $y_\pm = y_\pm^0 + o(1)$. Compact subsets of $(0,1)$ keep these roots separated and their derivatives bounded away from zero, which gives the stated uniformity. Set $a_\pm^2 = \lambda y_\pm$. At these points the longitudinal derivative equals $-1 - 4z\mathcal{H}_\lambda'(z)/\beta$. With $q_\rho = \sqrt{1-\rho}$, substitution gives

$$F'_{\lambda,K_\lambda}(a_\pm) = -1 \pm \frac{D_0^2}{\psi_0}(1 \pm q_\rho)q_\rho\lambda^2 + o(\lambda^2).$$

The larger cycle is attracting, the smaller repelling. The fixed point is stable because $\beta > \beta_{c,K_\lambda}$, and uniqueness and global IPO convergence were proved in the three-response part of Theorem 3.2 under precisely these inequalities. Finally, direct expansion of (85) gives

$$J_g(0) - J_g(a) = \frac{\alpha_\lambda g}{(\alpha_\lambda + 2)^2}a^2 + O(a^4) = \left(\frac{g}{9} + O(\lambda)\right)a^2 + O(a^4).$$

Substitution of the branch amplitudes proves (90). The rational depth table evaluates the exact formula at each reported K and $K-1$; the executable certificate supplies both strict signs. This completes the quantitative statements without asserting that the local fold is a global maximum. □

D.13 Completion of the population separation theorem

Proof of Theorem 3.1. Lemma A.1 proves the first assertion, including (10). The lower bound for β_{32} in (45) exceeds the IPO sufficient bound because

$$\frac{151}{400} - \left(\frac{3}{8} + \frac{1}{7200}\right) = \frac{17}{7200} > 0.$$

Lemma A.2 therefore gives global convergence and strict reward increase for every finite IPO depth. At $w = 0$, both gradients equal $-b_\theta$, so with $\eta = 1$ we have $w_{\mathrm{D},1} = w_{\mathrm{Q},1} = b_\theta$. This proves the DPO-1 statement.

Lemma A.3 and the definition of β_{32} give

$$F_{\mathrm{D},32}(a_+) = \left(\log 3 - \frac{v_{32}}{\beta_{32}}, 0\right) = (-\log 3, 0) = a_-.$$

Oddness in s gives the reverse equality. Lemmas A.3 and A.4 give (11) and a diagonal Jacobian at each point, so $\|DF_{\mathrm{D},32}(a_\pm)\| < 1$. By continuity there are disjoint closed balls $B_\pm$ of the same sufficiently small radius $r > 0$ and a number $q_D < 1$ with $\sup_{B_+ \cup B_-} \|DF_{\mathrm{D},32}\| \leq q_D$. Integrating the Jacobian on a segment inside either ball gives

$$\|F_{\mathrm{D},32}(\theta) - a_\mp\| \leq q_D \|\theta - a_\pm\|, \qquad \theta \in B_\pm.$$

The balls map alternately into one another, and the distance to the appropriate cycle point is at most q_D^n times its initial value. Their interiors give a nonempty open two-dimensional basin.

If $b_\theta = 0$, every inner iterate equals zero and θ is an outer fixed point. Conversely, if $b_\theta \neq 0$, the first inner step strictly lowers the loss by (10), and later steps cannot increase it. Thus $w_{\mathrm{D},32}$ cannot equal its starting value zero. Formula (118), together with $c_x, c_z > 0$, shows that $b_\theta = 0$ holds precisely at $s = t = 0$. This proves uniqueness of the fixed point.

To prove its local stability, we compute its Jacobian instead of inferring it from global convergence alone. Reflection makes $c_x(t)$ even in t and $c_z(s)$ even in s. From (40) and (41), for general $0 < \eta \leq 3$,

$$DF_{\mathrm{Q},K}(0) = \operatorname{diag}\left(1 - \frac{c}{\beta}\{1-(1-\eta/4)^K\},\ 1 - \frac{2c_z(0)}{3\beta}\{1-(1-\eta/3)^K\}\right). \tag{138}$$

Indeed the first derivatives of w_{Q}^* at zero are $-c$ and $-2c_z(0)/3$, while the corresponding entries of M_0 are 1 and $4/3$. Terms differentiating the damping factor vanish because $w_{\mathrm{Q}}^*(0) = 0$. At $\eta = 1$, $\beta = \beta_{32} > 3/8$, the first eigenvalue is strictly between -1 and 1 because $0 < c\{1-(3/4)^K\} < 3/4 < 2\beta$. The second is strictly between zero and one, since $2c_z(0)/(3\beta) < (1/60)/(3/8) < 1$. Theorem B.4 gives equality of the DPO and IPO Jacobians and second derivatives for the same K. Local contraction then proves local asymptotic stability of zero for DPO-32 as well as DPO-1. Finally, $\cosh(\log 3) = 5/3$ in (8) yields

$$J_* - J(a_\pm) = \gamma\left(\frac{5/3}{1+5/3} - \frac{1}{2}\right) = \gamma\left(\frac{5}{8} - \frac{1}{2}\right) = \frac{\gamma}{8}.$$

□

D.14 Cubic coefficients and monotonicity in depth

Proof of Theorem B.7. The first eigenvalue in (138) is $1 - cd_K/\beta$ and its derivative with respect to β is $cd_K/\beta^2 > 0$, so the crossing at $\beta_{c,K}$ is nondegenerate. Put $a_\eta = \eta/4$ and $b_\eta = \eta/3$. Since $0 < a_\eta \leq b_\eta \leq 1$,

$$\frac{1-(1-b_\eta)^K}{1-(1-a_\eta)^K} = \frac{b_\eta}{a_\eta}\frac{\sum_{j=0}^{K-1}(1-b_\eta)^j}{\sum_{j=0}^{K-1}(1-a_\eta)^j} \leq \frac{4}{3}.$$

At $\beta = cd_K/2$, the second eigenvalue therefore satisfies

$$1 > \lambda_z = 1 - \frac{4c_z(0)}{3c}\frac{1-(1-\eta/3)^K}{1-(1-\eta/4)^K} \geq 1 - \frac{16c_z(0)}{9c} > 0.$$

The invariant line $t = 0$ is consequently an invariant manifold tangent to the sole critical eigendirection.

We specialize the common scalar formula to this line, where $c_x(0) = c = 3/4$. Lemma A.2 gives

$$\begin{aligned} h(s) &= M_{xx}(s,0)/4 = \tfrac{1}{4}\operatorname{sech}^2(s/2) = \tfrac{1}{4} - \tfrac{1}{16}s^2 + O(s^4), \\ w_{\mathrm{Q},x}^*(s,0) &= -2c\tanh(s/2) = -cs + \tfrac{c}{12}s^3 + O(s^5), \\ b_x(s,0) &= h(s)w_{\mathrm{Q},x}^*(s,0) = -\tfrac{c}{4}s + \left(\tfrac{c}{48} + \tfrac{c}{16}\right)s^3 + O(s^5) = -\tfrac{3}{16}s + \tfrac{1}{16}s^3 + O(s^5). \end{aligned}$$

At zero the x marginal is $(1/4, 1/2, 1/4)$, so its odd moments vanish and $\mathbb{E}x^2 = \mathbb{E}x^4 = 1/2$. Independence gives

$$Q_4 = \mathbb{E}(x-x')^4 = 2\mathbb{E}x^4 + 6(\mathbb{E}x^2)^2 = 1 + 6/4 = 5/2.$$

Thus $(A, B, H, H_2, Q_4) = (3/16, 1/16, 1/4, -1/16, 5/2)$ and

$$\frac{2B}{A} = \frac{2}{3}, \qquad \frac{2H_2}{H} = -\frac{1}{2}, \qquad \Lambda = \frac{(5/2)(3/16)^2}{24(1/4)^3} = \frac{15}{64}.$$

In (14), $r = \rho = 1 - \eta/4$ and $W_K = B_K$ of (92). Substituting these values and $\omega_{\mathrm{Q}} = 0, \omega_{\mathrm{D}} = 1$ in (16) gives

$$\chi_{\mathrm{Q},K} = \frac{2}{3} - \frac{1}{2}(1-u_K) = \frac{1}{6} + \frac{1}{2}u_K, \qquad \chi_{\mathrm{D},K} = \frac{1}{6} + \frac{1}{2}u_K - \frac{15}{64}B_K,$$

which is (93); odd symmetry gives the $O(s^5)$ remainder. For $0 < \eta \le 3$, $\rho \in [1/4, 1)$, so the monotonicity argument in Appendix D.7 applies. It proves that u_K strictly decreases from one to zero and B_K strictly increases from zero to one. Hence $\chi_{\mathrm{Q},K} > 0$, while $\chi_{\mathrm{D},K}$ strictly decreases from $2/3$ to $1/6 - 15/64 = -13/192$. For $\eta = 1$, every expression in (92) is rational. The exact arithmetic in Appendix F proves

$$\frac{3}{1000} < \chi_{\mathrm{D},15} < \frac{1}{250}, \qquad -\frac{11}{1000} < \chi_{\mathrm{D},16} < -\frac{1}{100}.$$

Strict monotonicity gives the sign statements for every other integer depth. □

Justification of the bifurcation interpretation. Fix K, η with $\chi = \chi_{X,K} \neq 0$ and use $\mu = cd_K/\beta - 2$ as the parameter. The invariant scalar map is analytic, odd, and has the expansion

$$f_\mu(s) = -(1+\mu)s + \{\chi + O(\mu)\}s^3 + O(s^5),$$

with an analytic remainder near $(s, \mu) = (0, 0)$. Solving $f_\mu(s) = -s$ for $s \neq 0$ is equivalent to solving, with $z = s^2$,

$$0 = -\mu + \{\chi + O(\mu)\}z + O(z^2).$$

This is an analytic equation in z, μ; its derivative in z at zero is $\chi \neq 0$. The implicit function theorem gives a unique small branch $z(\mu) = \mu/\chi + O(\mu^2)$. Where $z > 0$, the distinct points $\pm\sqrt{z}$ are exchanged by the map and form a two-cycle. At either point,

$$f'_\mu(\sqrt{z}) = -(1+\mu) + 3\{\chi + O(\mu)\}z + O(z^2) = -1 + 2\mu + O(\mu^2).$$

The derivative of the second iterate in this direction is its square, $1 - 4\mu + O(\mu^2)$. For small $\mu > 0$ it lies in $(0, 1)$; for small $\mu < 0$ it exceeds one. The transverse multiplier remains inside $(0, 1)$ by continuity of the Jacobian along $t = 0$, and mixed derivatives vanish by reflection. At zero the critical multiplier has modulus $1 + \mu$, so zero is stable on the $\mu < 0$ side near zero. This proves the supercritical/subcritical statements made in the main text. No assertion is made here about the degenerate case $\chi = 0$.

D.15 The continuous inner-time threshold

Suppose $\eta \downarrow 0$ and $K\eta \to \tau > 0$. Set $x = \tau/4$. Since $\log(1-\eta/4) = -\eta/4 + O(\eta^2)$ and $K\eta^2 \to 0$,

$$\rho^K \to e^{-x}, \qquad u_K = \frac{(\eta/4)K\rho^{K-1}}{1-\rho^K} \longrightarrow \frac{x}{e^x - 1}.$$

Expanding the cubic inside the sum gives the exact identity

$$\begin{aligned}\sum_{j=0}^{K-1} \rho^{K-1-j}(1-\rho^j)^3 &= \sum_{j=0}^{K-1} \rho^{K-1-j} - 3\sum_{j=0}^{K-1} \rho^{K-1} + 3\sum_{j=0}^{K-1} \rho^{K-1+j} - \sum_{j=0}^{K-1} \rho^{K-1+2j} \\ &= \frac{1-\rho^K}{1-\rho} - 3K\rho^{K-1} + 3\rho^{K-1}\frac{1-\rho^K}{1-\rho} - \rho^{K-1}\frac{1-\rho^{2K}}{1-\rho^2}.\end{aligned}$$

Multiplying by $(1-\rho)/(1-\rho^K)$ simplifies this further to

$$B_K = 1 - 3u_K + 3\rho^{K-1} - \rho^{K-1}\frac{1+\rho^K}{1+\rho}.$$

Its limit is

$$B(x) = 1 - \frac{3x}{e^x - 1} + \frac{5}{2}e^{-x} - \frac{1}{2}e^{-2x}. \tag{139}$$

Substitution into (93) gives (94) and (95), including $1/2 + 45/64 = 77/64$ as the coefficient of $x/(e^x - 1)$.

For strict monotonicity, the derivative of $u(x) = x/(e^x - 1)$ has numerator $e^x - 1 - xe^x = e^x(1-x) - 1$. It is negative for $x > 0$, since its value at zero is zero and its derivative is $-xe^x < 0$. The second term has the integral representation

$$\begin{aligned}B(x) &= \frac{1}{e^x - 1}\int_0^x e^v(1-e^{-v})^3\,dv \\ &= \frac{e^x - 3x - 3e^{-x} + \frac{1}{2}e^{-2x} + \frac{3}{2}}{e^x - 1}.\end{aligned}$$

Algebraically dividing the numerator verifies (139). The integrand represents the expectation of the strictly increasing function $g(v) = (1 - e^{-v})^3$ under a random variable supported on $[0, x]$ with density $e^v/(e^x - 1)$. For $0 < v < x$, its distribution function is $(e^v - 1)/(e^x - 1)$, which strictly decreases as x increases. The expectation $B(x)$ therefore strictly increases. It follows that $1/6 + u(x)/2 - 15B(x)/64$ strictly decreases. As $x \downarrow 0$, $u(x) \to 1$ and $0 \le B(x) \le g(x) \to 0$, so the coefficient tends to $2/3$. As $x \to \infty$, it tends to $-13/192$. There is exactly one zero. The exact sign certificates at $\tau = 16$ and $\tau = 17$ put this zero in $(16, 17)$.

For the threshold limit, fix $0 < \varepsilon < \tau_*$. Let $K_-(\eta) = \lfloor(\tau_* - \varepsilon)/\eta\rfloor$ and $K_+(\eta) = \lfloor(\tau_* + \varepsilon)/\eta\rfloor$. Both are positive for small η. The just-proved limits imply $\chi_{\mathrm{D},K_-} > 0$ and $\chi_{\mathrm{D},K_+} < 0$ for all sufficiently small η. Strict decrease in K gives $K_- < K_{\mathrm{crit}}(\eta) \le K_+$. Multiplying by η, taking lower and upper limits, and then letting $\varepsilon \downarrow 0$ proves $\eta K_{\mathrm{crit}}(\eta) \to \tau_*$. This argument allows a particular finite-η coefficient to be exactly zero.

D.16 Bounded updates and the amplitude curve

Proof of Lemma A.7. For $C = \cosh s$, differentiation gives

$$f_v(s, v) = -\frac{C\,\mathrm{sech}^2(v/2) + \mathrm{sech}^2 v}{2(1 + C)^2}, \qquad -\tfrac{1}{4} \le -\frac{1}{2(1 + C)} \le f_v(s, v) < 0.$$

For $s > 0$, $f(s, 0) = 3\sinh s/[4(1 + C)^2] > 0$. Since $\tanh(\log 7/2) = 3/4$ and $\tanh(\log 7) = 24/25$,

$$f(s, \log 7) = \frac{-\frac{3}{2}e^{-s} - 24/25}{2(1 + C)^2} < 0.$$

Strict decrease and continuity give a unique root in $(0, \log 7)$. The update $T_s(v) = v + \eta f(s, v)$ satisfies $T_s'(v) \ge 1 - \eta/4 \ge 1/4 > 0$. If $0 \le v < \nu_\infty(s)$, then $f(s, v) > 0$ and $v < T_s(v) < T_s(\nu_\infty(s)) = \nu_\infty(s)$. Induction proves (54). The denominator of f is positive on $\mathbb{R}^2$, so f and its finite compositions are real analytic. The identity $f(-s, -v) = -f(s, v)$ gives oddness by induction. Dividing an odd analytic function by s gives an even analytic extension at zero. At $\beta = \beta_{c,K}$, Theorem B.7 gives

$$s - \nu_K(s)/\beta_{c,K} = -s + \chi_{\mathrm{D},K} s^3 + O(s^5), \qquad \nu_K(s) = 2\beta_{c,K} s - \beta_{c,K}\chi_{\mathrm{D},K} s^3 + O(s^5).$$

Division by $2s$ proves the expansion. Finally $0 < \mathcal{A}_K(s) < \log 7/(2s)$ proves the tail limit. □

D.17 Uniform transverse control

Proof of Lemma A.8. For any outer (s, t) and inner $(w_x, 0)$, independence gives

$$\mathbb{E}[(z - z')\sigma(w_x(x - x'))] = \mathbb{E}[z - z']\mathbb{E}[\sigma(w_x(x - x'))] = 0.$$

Since $\mathbb{E}[z - z'] = 0$, (118) yields

$$\nabla_z L_{\mathrm{D}}((w_x, 0); (s, t)) = -b_z(s, t) = \frac{2c_z(s)\sinh t}{(1 + 2\cosh t)^2}.$$

Its partial t derivative at zero is $2c_z(s)/9 = \mathcal{N}(s)$. This includes the sampling-law derivative and does not depend on w_x. Moreover $0 < c_z(s) < \delta/2$, so $0 < \mathcal{N}(s) < \delta/9$. At $t = w_z = 0$ the mixed Hessian is $\mathbb{E}[d_x\sigma'(w_x d_x)]\mathbb{E}[d_z] = 0$, and its zz entry is $\mathbb{E}[d_z^2]\mathbb{E}[\sigma'(w_x d_x)] = \frac{4}{3}\mathbb{E}[\sigma'(w_x d_x)]$. Differentiating GD and putting $U_j = \partial_t w_{\mathrm{D},j,z}$ gives $U_{j+1} = (1 - \eta H_j)U_j - \eta\mathcal{N}(s)$, $U_0 = 0$. Thus $\kappa_j = -U_j$ obeys (57). Let p_-, p_0, p_+ be the x marginal. The event $x = x'$ gives

$$H_j \ge \frac{4}{3}\sigma'(0)\mathbb{P}(x = x') = \tfrac{1}{3}(p_-^2 + p_0^2 + p_+^2) \ge \tfrac{1}{9},$$

where $1 = (p_- + p_0 + p_+)^2 \le 3(p_-^2 + p_0^2 + p_+^2)$. Also $H_j \le 1/3$ by $\sigma' \le 1/4$. Hence $0 \le 1 - \eta H_j \le 1 - \eta/9 < 1$. The recurrence gives positivity for $j \ge 1$ and

$$\kappa_K \le \eta\mathcal{N}\sum_{j=0}^{K-1}(1 - \eta/9)^j = 9\mathcal{N}\{1 - (1 - \eta/9)^K\} < \delta.$$

Reflection makes the first outer coordinate even in t and the second identically zero on $t = 0$, so both mixed derivatives vanish. The diagonal entries are $g'_{K,\beta}(s)$ and $1 - \kappa_K/\beta$. For $\beta > \delta$, $0 < 1 - \delta/\beta < 1 - \kappa_K/\beta < 1$. □

D.18 Global IPO convergence at the critical scale

Proof of Lemma A.9. Put $d_L = 1-(1-\eta/4)^L$, $e_L = 1-(1-\eta/3)^L$. Solving the affine inner recursion gives $w_{\mathrm{Q},L} = \{\mathrm{Id}-(\mathrm{Id}-\eta M/4)^L\}w_{\mathrm{Q}}^*$. Define $a_x(s) = 1-(1-\eta M_{xx}/4)^L$ and $a_z(t) = 1-(1-\eta M_{zz}/4)^L$. Then $0 < a_x \le d_L$, $0 < a_z \le e_L$, and

$$s^+ = s - \frac{2c_x(t)a_x(s)}{\beta}\tanh(s/2), \tag{140}$$

$$t^+ = t - \frac{2c_z(s)a_z(t)}{\beta}\frac{\sinh t}{\cosh t + 2}. \tag{141}$$

The geometric sums satisfy

$$\frac{e_L}{d_L} = \frac{4}{3}\frac{\sum_{j=0}^{L-1}(1-\eta/3)^j}{\sum_{j=0}^{L-1}(1-\eta/4)^j} \le \frac{4}{3}.$$

For $u > 0$, $\sinh u = \int_0^u \cosh v\,dv \le u\cosh u$ gives $0 < \sinh u/[u(\cosh u+2)] < 1$. Thus for $t \neq 0$,

$$0 < \frac{|t-t^+|}{|t|} = \frac{2c_z(s)a_z(t)}{\beta}\frac{\sinh|t|}{|t|(\cosh|t|+2)} < \frac{\delta e_L}{\beta} < \frac{2\delta e_L}{cd_L} \le \frac{8\delta}{3c} = \frac{8}{45} < 1.$$

The transverse coordinate keeps its sign and its absolute value strictly decreases outside zero. The s displacement points toward zero and has magnitude $m \le 2/\beta$, since $c_x < 1$, $a_x \le 1$, $|\tanh| \le 1$. Consequently $|s^+| = ||s|-m| \le \max(|s|, 2/\beta)$. Every trajectory stays in the compact rectangle $|s_n| \le \max(|s_0|, 2/\beta)$, $|t_n| \le |t_0|$. Monotonicity gives a signed limit $t_n \to t_\infty$. For a convergent subsequence $s_{n_j} \to \bar{s}$, continuity gives

$$0 = \frac{2c_z(\bar{s})a_z(t_\infty)}{\beta}\frac{\sinh t_\infty}{\cosh t_\infty + 2}.$$

Both coefficients are positive, so $t_\infty = 0$. Hence $c_x(t_n) \to c$. Since $\beta > cd_L/2$, there is n_0 such that $c_x(t_n)d_L/\beta < 2$ for $n \ge n_0$. For nonzero s_n after this time,

$$0 < \frac{|s_n - s_{n+1}|}{|s_n|} = \frac{2c_x(t_n)a_x(s_n)}{\beta}\frac{\tanh(|s_n|/2)}{|s_n|} \le \frac{c_x(t_n)d_L}{\beta} < 2.$$

Thus $|s_n|$ eventually decreases to a limit ℓ. Along a subsequence $(s_{n_j}, t_{n_j}) \to (\bar{s}, 0)$, continuity yields $|\bar{s}| = \ell = |F_{\mathrm{Q},L,x}(\bar{s}, 0)|$. Strict decrease on the axis excludes $\bar{s} \neq 0$, so $\ell = 0$. This proves convergence to zero and, by continuity, convergence of reward to J_*. Finally $\nabla L_{\mathrm{D}}(0;\theta) = \nabla L_{\mathrm{Q}}(0;\theta) = -b_\theta$ gives $F_{\mathrm{D},1} = F_{\mathrm{Q},1}$. □

D.19 Coexistence from negative cubic criticality

Proof of Theorem B.2. Fix η, K as stated and abbreviate $\mathcal{A} = \mathcal{A}_K$, $\beta_c = \beta_{c,K}$, $\beta_m = \beta_{m,K}$. The weights defining B_K sum to one, and $1-\rho^j < d_K$ for $j < K$. Thus $B_K < d_K^3$. Negativity of the cubic coefficient implies

$$\frac{15}{64}B_K > \frac{1}{6} + \frac{1}{2}u_K > \frac{1}{6}, \qquad d_K^3 > B_K > \frac{32}{45} > \left(\frac{2}{3}\right)^3.$$

Therefore

$$\beta_c = 3d_K/8 > 1/4 > \delta. \tag{142}$$

Lemma A.7 gives the analytic odd displacement, its positivity for $s > 0$, its sublinear tail, and $\mathcal{A}''(0) > 0$. Apply Lemma A.14 to obtain the maximum $a_\star$, interval $\mathcal{B}_K$, and the two longitudinal cycle branches. The transverse condition required by that lemma follows from Lemma A.8 and $\beta > \beta_c > \delta$. Its two-dimensional conclusion therefore gives the attracting cycle and saddle, including an open attracting set.

Formula (118), with $c_x, c_z > 0$, gives $b_\theta = 0$ exactly at zero. There all inner iterates vanish. Elsewhere the first GD step satisfies

$$L_{\mathrm{D}}(w_1;\theta) \le L_{\mathrm{D}}(0;\theta) - \eta(1-\eta/6)\|b_\theta\|^2 < L_{\mathrm{D}}(0;\theta),$$

and later steps cannot increase the loss. Hence $w_K \neq 0$ and no other outer fixed point exists. At zero,

$$DF_{\mathrm{D},K}(0) = \mathrm{diag}(1 - 2\beta_c/\beta, 1 - \kappa_K(0)/\beta).$$

Both entries have modulus below one, proving local contraction. Equation (8) makes zero uniquely reward-optimal, and Theorem B.4 gives second-order agreement. For $L \leq K$, $\beta_{c,L} \leq \beta_c < \beta$; global convergence of all comparison procedures follows from Lemma A.9. For $a > 0$, the reward gap is

$$J_* - J(\pm a, 0) = \gamma \left(\frac{\cosh a}{1 + \cosh a} - \frac{1}{2} \right) = \frac{\gamma}{2} \frac{\cosh a - 1}{\cosh a + 1} = \frac{\gamma}{2} \tanh^2(a/2).$$

As $r_+ > a_\star$, this proves the uniform lower bound. Continuity of J and attraction imply convergence to this reward gap throughout the basin. Finally $\chi_{\mathrm{D},1} = 2/3$ implies $K \geq 2$. At a cycle amplitude $a = r_+$, $0 < \nu_{K-1}(a) < \nu_K(a) = 2\beta a$, so the preceding candidate has coordinate $a - \nu_{K-1}(a)/\beta \in (-a, a)$ and strictly larger reward than the final candidate $(-a, 0)$. Since $\nu_{K-1} < \nu_\infty$, its gradient is nonzero, and the descent inequality makes the last loss decrease strict. Reflection handles the other cycle point. Frozen GD convergence is Lemma A.1. □

D.20 Depth ranges and matched longitudinal gain

Proof of Corollary B.8. Theorem B.7 gives negativity for all $K \geq 16$ at $\eta = 1$. For fixed $0 < \eta \leq 3$ its coefficient decreases to $-13/192 < 0$, so it is negative at all sufficiently large finite depths. Theorem B.2 applies in each case. □

Proof of Corollary B.9. $\beta_1 = \beta d_1/d_K > (cd_K/2)d_1/d_K = cd_1/2 = \beta_{c,1}$, so Lemma A.9 gives global convergence of G_1. Equation (118) gives $\partial_s b_x(0,0) = -c/4$; hence

$$\partial_s G_{1,x}(0,0) = 1 - \eta c/(4\beta_1) = 1 - cd_1/\beta_1 = 1 - cd_K/\beta.$$

This is the derivative of $F_{\mathrm{D},K}$ at zero. Its attracting basin gives the positive limiting gap, while global convergence of G_1 gives gap zero. □

D.21 Inner flow and exact-minimization geometry

Proof of Lemma A.10. For fixed θ, both gradients are globally Lipschitz with constant at most $1/3$, by Lemma A.1. Thus the ODEs have unique global solutions. Along either solution,

$$\frac{d}{du} L_X(w_X(u); \theta) = -\|\nabla L_X(w_X(u); \theta)\|^2 \leq 0.$$

The coercivity proved in Lemma A.1 makes the initial sublevel set compact. It is convex and contains the unique minimizer w_X^*. The positive-definite continuous Hessian has a positive minimum eigenvalue m_θ on this set. Hence, integrating the Hessian on the segment from w_X^* to $w_X(u)$,

$$\begin{aligned} \frac{d}{du} \|w_X(u) - w_X^*\|^2 &= -2(w_X(u) - w_X^*)^\top \nabla L_X(w_X(u); \theta) \\ &\leq -2m_\theta \|w_X(u) - w_X^*\|^2. \end{aligned}$$

Multiplying by $e^{2m_\theta u}$ and integrating proves convergence. Analytic dependence of an analytic ODE on its parameters, continued along the solution over a fixed finite time interval, gives analyticity in θ. The analytic implicit function theorem applied to $\nabla L_X(w_X^*; \theta) = 0$ gives analytic minimizers: its derivative in w is the invertible inner Hessian, and uniqueness identifies the local branches with the global minimizer.

Reflection equivariance follows from the same relabeling as in Lemma A.3 and uniqueness of ODE solutions or minimizers. Thus the axis is invariant, and its scalar flow is $\dot{\nu} = f(s, \nu)$, $\nu(0) = 0$. For $s > 0$ the function $f(s, \cdot)$ is positive below its root $\nu_\infty(s) \in (0, \log 7)$. The solution cannot meet this equilibrium at finite time: local uniqueness backward from a meeting point would identify it with the constant solution. It therefore increases strictly while staying below the root. Its limit is the root, since a smaller limit would leave its derivative bounded below by a positive constant. This proves (60). The same bound implies $0 < \mathcal{A}_\tau(s) < \log 7/(2s)$ for finite or infinite τ, proving the limit at infinity.

We derive the local coefficients directly, rather than exchanging a time limit with a Taylor remainder. Expansion of (43) gives

$$f(s,v) = \frac{3s}{16} - \frac{v}{4} - \frac{s^3}{16} + \frac{s^2 v}{16} + \frac{5v^3}{96} + O((|s|+|v|)^5). \tag{143}$$

Indeed $\sinh s = s + s^3/6 + O(s^5)$, $\cosh s = 1 + s^2/2 + O(s^4)$, $\tanh(v/2) = v/2 - v^3/24 + O(v^5)$, and $\tanh v = v - v^3/3 + O(v^5)$. The numerator is $3s/2 - 2v + s^3/4 - s^2 v/2 + 5v^3/12 + O((|s| + |v|)^5)$, and the inverse denominator is $(1/8)(1 - s^2/2 + O(s^4))$. Multiplication yields (143).

Write $\nu_u(s) = a(u)s + b(u)s^3 + O(s^5)$. Differentiating the ODE at $s = 0$, or matching these analytic coefficients, gives

$$\begin{aligned} a' &= 3/16 - a/4, & a(0) &= 0, \\ b' &= -b/4 - 1/16 + a/16 + 5a^3/96, & b(0) &= 0. \end{aligned}$$

Thus $a(u) = c(1 - e^{-u/4})$. In IPO the cubic v^3 term is absent, while the remaining terms agree. Its cubic coefficient is consequently

$$\begin{aligned} b_{\mathrm{Q}}(\tau) &= \int_0^\tau e^{-(\tau-u)/4} \left\{ -\frac{1}{16} + \frac{c(1-e^{-u/4})}{16} \right\} du \\ &= \frac{c-1}{4}(1 - e^{-\tau/4}) - \frac{c\tau e^{-\tau/4}}{16} \\ &= -\frac{d_\tau}{16} - \frac{c\tau e^{-\tau/4}}{16}, \\ b(\tau) - b_{\mathrm{Q}}(\tau) &= \frac{5c^3}{96} \int_0^\tau e^{-(\tau-u)/4}(1 - e^{-u/4})^3 \, du. \end{aligned}$$

At $\beta_{c,\tau} = c d_\tau/2$, the outer cubic coefficient is $-b(\tau)/\beta_{c,\tau}$. Put $x = \tau/4$. Changing variables $u = 4v$ in the last integral yields

$$\begin{aligned} \chi_{\mathrm{D}}^{\mathrm{flow}}(\tau) &= \frac{1}{6} + \frac{x}{2(e^x - 1)} - \frac{15}{64} B(x), \\ B(x) &= \frac{1}{e^x - 1} \int_0^x e^v (1 - e^{-v})^3 \, dv \\ &= 1 - \frac{3x}{e^x - 1} + \frac{5}{2} e^{-x} - \frac{1}{2} e^{-2x}. \end{aligned}$$

This is exactly (95). Division of $\nu_\tau(s) = c d_\tau s + b(\tau) s^3 + O(s^5)$ by $2s$ proves the finite-time assertion in (61).

For exact minimization, use the analytic equation $f(s, \nu_\infty(s)) = 0$. The linear coefficient is $a = c$. Its cubic coefficient satisfies

$$0 = -b/4 - 1/16 + c/16 + 5c^3/96, \qquad b = -1/4 + c/4 + 5c^3/24 = 13/512.$$

Thus $\nu_\infty(s) = cs + (13/512)s^3 + O(s^5)$ and the critical outer coefficient is $-(13/512)/(3/8) = -13/192$.

For transverse derivatives, the calculation in Lemma A.8 gives, along the flow,

$$\dot{\kappa}_u = \mathcal{N}(s) - H(u,s)\kappa_u, \qquad \kappa_0 = 0, \quad H(u,s) = \tfrac{4}{3}\mathbb{E}_s \sigma'(\nu_u(s)(x - x')) \in [1/9, 1/3].$$

The integrating factor formula gives

$$0 < \kappa_\tau(s) = \mathcal{N}(s) \int_0^\tau \exp\left\{ -\int_u^\tau H(v,s)\, dv \right\} du \le 9\mathcal{N}(s)(1 - e^{-\tau/9}) < \delta.$$

At the exact minimizer, implicit differentiation of the z gradient, whose mixed Hessian is zero, gives $\kappa_\infty(s) = \mathcal{N}(s)/H(\infty, s) \le 9\mathcal{N}(s) < \delta$. Reflection eliminates mixed outer derivatives, completing the axis bound.

It remains to verify the full second-order comparison. Write

$$\mathcal{R}_\theta(w) = \mathbb{E}_\theta\left[d\left\{\frac{d^\top w}{4} - \sigma(d^\top w) + \frac{1}{2}\right\}\right] = O(\|w\|^3).$$

The DPO gradient is $-b_\theta + M_\theta w/4 - \mathcal{R}_\theta(w)$. Near $\theta = 0$, smooth dependence and $b_0 = 0$ give $w_X(u;h) = O(\|h\|)$ uniformly on each fixed finite time interval. Subtracting the flow equations and solving their linear difference equation,

$$w_{\mathrm{D}}(\tau;h) - w_{\mathrm{Q}}(\tau;h) = \int_0^\tau e^{-(\tau-u)M_h/4}\mathcal{R}_h(w_{\mathrm{D}}(u;h))\,du = O(\|h\|^3).$$

For exact minimizers, analytic dependence gives $w^*_{\mathrm{D}}(h) = O(\|h\|)$, and the stationary equations yield

$$w^*_{\mathrm{D}}(h) - w^*_{\mathrm{Q}}(h) = 4M_h^{-1}\mathcal{R}_h(w^*_{\mathrm{D}}(h)) = O(\|h\|^3),$$

since M_h^{-1} is locally bounded. Analyticity and these remainders give equality of all terms of degree at most two after division by β. □

D.22 Fifth-order expansion and rational sign certification

Proof of Lemma A.13. We expand the scalar vector field (43) through degree five. The requisite series are

$$\begin{aligned}
\sinh s &= s + s^3/6 + s^5/120 + O(s^7),\\
\cosh s &= 1 + s^2/2 + s^4/24 + O(s^6),\\
\tanh(v/2) &= v/2 - v^3/24 + v^5/240 + O(v^7),\\
\tanh v &= v - v^3/3 + 2v^5/15 + O(v^7).
\end{aligned}$$

The numerator of (43) consequently equals

$$\begin{aligned}
N(s,v) = \frac{3}{2}s - 2v + \frac{1}{4}s^3 - \frac{1}{2}s^2v + \frac{5}{12}v^3\\
+ \frac{1}{80}s^5 - \frac{1}{24}s^4v + \frac{1}{24}s^2v^3 - \frac{17}{120}v^5 + O((|s|+|v|)^7).
\end{aligned}$$

The denominator and its inverse are

$$\begin{aligned}
2(1+\cosh s)^2 &= 8\{1 + s^2/2 + 5s^4/48 + O(s^6)\},\\
\{2(1+\cosh s)^2\}^{-1} &= \frac{1}{8}\{1 - s^2/2 + 7s^4/48 + O(s^6)\}.
\end{aligned}$$

For clarity, the five-degree coefficients in their product are

$$\begin{aligned}
[s^5]f &= \frac{1}{8}\left(\frac{1}{80} - \frac{1}{8} + \frac{7}{32}\right) = \frac{17}{1280},\\
[s^4v]f &= \frac{1}{8}\left(-\frac{1}{24} + \frac{1}{4} - \frac{7}{24}\right) = -\frac{1}{96},\\
[s^2v^3]f &= \frac{1}{8}\left(\frac{1}{24} - \frac{5}{24}\right) = -\frac{1}{48}, \qquad [v^5]f = -\frac{17}{960}.
\end{aligned}$$

Together with (143), this proves

$$\begin{aligned}
f(s,v) = \frac{3s}{16} - \frac{v}{4} - \frac{s^3}{16} + \frac{s^2v}{16} + \frac{5v^3}{96}\\
+ \frac{17s^5}{1280} - \frac{s^4v}{96} - \frac{s^2v^3}{48} - \frac{17v^5}{960} + O((|s|+|v|)^7).
\end{aligned} \tag{144}$$

Set $x = \tau/4$ and expand $\nu_{4x}(s) = a(x)s + b(x)s^3 + d(x)s^5 + O(s^7)$. Primes in the next display mean differentiation in x, so that $\partial_x\nu_{4x} = 4f(s,\nu_{4x})$. Matching coefficients gives

$$\begin{aligned}
a' &= 3/4 - a,\\
b' &= -b - 1/4 + a/4 + 5a^3/24,\\
d' &= -d + (1/4 + 5a^2/8)b + 17/320 - a/24 - a^3/12 - 17a^5/240,
\end{aligned} \tag{145}$$

with all initial values zero. For the a^2b coefficient we used $(as + bs^3 + ds^5)^3 = a^3s^3 + 3a^2bs^5 + O(s^7)$. Solving the first equation gives $a = 3(1 - E)/4$. Substitution into the second yields

$$b' + b = \frac{13}{512} - \frac{231}{512}E + \frac{135}{512}E^2 - \frac{45}{512}E^3.$$

Differentiating the expression for b in (66) verifies this identity term by term, and its numerator is zero at $x = 0$. Uniqueness of the scalar linear ODE proves that expression.

For the final equation, substitute this a, b and expand its forcing $G = (1/4 + 5a^2/8)b + 17/320 - a/24 - a^3/12 - 17a^5/240$. With $D = 2621440$, the coefficient of E^k in DG is

$$\begin{aligned} q_0 &= -38840, & q_1 &= 838380 - 711480x, \\ q_2 &= -1467720 + 831600x, & q_3 &= 1267200 - 415800x, \\ q_4 &= -544320, & q_5 &= 84564. \end{aligned}$$

For example the term containing b is explicitly

$$\left(\frac{77}{128} - \frac{45}{64}E + \frac{45}{128}E^2\right)\frac{26 + (199 - 462x)E - 270E^2 + 45E^3}{1024},$$

and the other terms are $17/320 - (1 - E)/32 - 9(1 - E)^3/256 - 1377(1 - E)^5/81920$. Collecting these polynomials gives exactly the six displayed q_k. For $d = D^{-1}\sum_k P_k(x)E^k$,

$$d' + d = D^{-1}\sum_{k=0}^{5}\{P_k'(x) + (1 - k)P_k(x)\}E^k.$$

Each coefficient equals q_k: they are respectively

$$-38840, \quad 838380 - 711480x, \quad -831600 - (636120 - 831600x),$$

$$207900 - 2(-529650 + 207900x), \quad -3(181440), \quad -4(-21141).$$

Also $\sum_k P_k(0) = -38840 - 227929 + 636120 - 529650 + 181440 - 21141 = 0$. Thus (67) has the correct forcing and initial value, which proves it by uniqueness. Dividing the inner expansion by $\beta_c = a/2$ gives $\chi = -2b/a$, $\psi = -2d/a$.

We next justify the certification without using floating-point signs. Differentiating (95), with $x = \tau/4$, gives

$$-\chi'(\tau) = -\frac{77}{256}\frac{E(1 - x - E)}{(1 - E)^2} - \frac{75}{512}E + \frac{15}{256}E^2. \tag{146}$$

The additional function `certify_quintic_fold` in Appendix F evaluates (95) at $\tau_L = 420903/25000$ and $\tau_U = 1683613/100000$ and obtains strictly positive and negative enclosures, respectively. The existing uniqueness of the zero therefore places τ_* inside this interval. On the entire interval it evaluates (67), $\psi = -8d/[3(1 - E)]$, and (146) with rational interval arithmetic. The exponential enclosure uses the finite positive Taylor sum and its geometric tail bound already justified in Appendix F. All arithmetic is rounded outward to the rational grid. The returned bounds are (3.5); substituting the same interval enclosures into $\beta_c\kappa^2/(8\psi)$ gives (68). These operations enclose the whole time interval, including its unknown root, rather than just its midpoint.

Finally, the vector field is real analytic near the compact zero solution segment with $0 \le \tau \le 17$. The standard local analytic dependence of ODE solutions on parameters, continued over finitely many overlapping time neighborhoods of that segment, makes $\nu_\tau(s)$ jointly analytic for small s and τ near τ_*. One may equivalently obtain its successive parameter derivatives from the triangular variational ODEs; their coefficients are bounded on this compact neighborhood. Reflection makes ν odd. Dividing by s, grouping the even powers as powers of $z = s^2$, and dividing by the positive analytic $\beta_c(\tau)$ gives a jointly analytic R in (69). Choose a closed rectangle strictly inside this analytic neighborhood. The finitely many continuous derivatives $\partial_z^i\partial_\tau^j R$ with $i + j \le 2$ have a finite maximum there; taking its maximum with one defines C. This proves the parameter-uniform remainder and all claimed bounds. □

D.23 Coexistence for finite inner time and exact minimization

Proof of Theorem B.10. For finite $\tau > \tau_*$, (95) is negative. Its representation in the preceding proof implies $B(x) > 32/45$. The density $e^v/(e^x - 1)$ on $[0, x]$ integrates to one, and $(1 - e^{-v})^3 < (1 - e^{-x})^3 = d_\tau^3$ except at its endpoint. Therefore $d_\tau^3 > B(x) > 32/45 > (2/3)^3$, giving $\beta_{c,\tau} > 1/4 > \delta$. At $\tau = \infty$ the coefficient is $-13/192 < 0$ and $\beta_{c,\infty} = 3/8 > \delta$.

For either case, Lemma A.10 verifies all scalar hypotheses of Lemma A.14. It gives a nonempty interval $\mathcal{B}_\tau$ above $\beta_{c,\tau}$ and an attracting longitudinal cycle at an amplitude $a_\tau(\beta) > 0$. The transverse bound $1 - \kappa_\tau(a)/\beta \in (0, 1)$ verifies the additional hypothesis for full-dimensional attraction. Thus the same common lemma applies to finite GD, flow, and exact inner solves; the distinct work is verifying its hypotheses for each solver.

If $b_\theta \neq 0$, the initial inner gradient is nonzero. For finite $\tau > 0$, continuity makes $\int_0^\tau \|\nabla L_{\mathrm{D}}(w_{\mathrm{D}}(u); \theta)\|^2 du > 0$, so $L_{\mathrm{D}}(w_{\mathrm{D}}(\tau); \theta) < L_{\mathrm{D}}(0; \theta)$ and $w_{\mathrm{D}}(\tau) \neq 0$. For exact minimization, $w_{\mathrm{D}}^* = 0$ is equivalent to $b_\theta = 0$. Thus zero is the unique fixed point in both cases. Its Jacobian is

$$D\mathsf{F}_{\mathrm{D},\tau}(0) = \mathrm{diag}(1 - 2\beta_{c,\tau}/\beta, 1 - \kappa_\tau(0)/\beta),$$

whose entries have modulus less than one. It is locally stable and, by (8), uniquely reward-optimal.

We verify the global IPO comparison. The exact solution of its linear ODE is

$$w_{\mathrm{Q}}(\tau; \theta) = (\mathrm{Id} - e^{-\tau M_\theta/4}) w_{\mathrm{Q}}^*(\theta).$$

Hence (140)–(141) hold with $a_x(s) = 1 - e^{-\tau M_{xx}(s)/4}$ and $a_z(t) = 1 - e^{-\tau M_{zz}(t)/4}$. They satisfy $0 < a_x \leq d_\tau$, $0 < a_z \leq e_\tau = 1 - e^{-\tau/3}$ and

$$\frac{e_\tau}{d_\tau} = \frac{4}{3} \frac{\int_0^\tau e^{-u/3} du}{\int_0^\tau e^{-u/4} du} \leq \frac{4}{3}.$$

For exact IPO use $a_x = a_z = d_\infty = e_\infty = 1$. The proof of Lemma A.9 uses precisely these positive bounds, this ratio bound, and $\beta > c d_\tau/2$. In particular its transverse relative displacement is below $8/45$, its first coordinate is bounded by $\max(|s_0|, 2/\beta)$, and it first obtains $t_n \to 0$ and then $s_n \to 0$. These steps apply unchanged and prove global convergence. Second-order agreement is Lemma A.10.

The reward formula gives the stated positive gap. In a frozen round starting at $(a, 0)$, each intermediate candidate has first coordinate $a - \nu_u(a)/\beta$. For finite τ, strict scalar increase gives $0 < \nu_u(a) < \nu_\tau(a) = 2\beta a$ for $0 < u < \tau$. For exact minimization, $0 < \nu_u(a) < \nu_\infty(a) = 2\beta a$ for every finite $u > 0$. Thus every specified intermediate coordinate lies in $(-a, a)$ and has strictly larger reward than the returned point $(-a, 0)$. Reflection proves the assertion at the other point. The inner gradient is nonzero at every finite time on this trajectory, so its loss decreases strictly there; convergence to the minimizer was proved above. □

D.24 The local fold, reward scaling, and the axial basin

Proof of Theorem 3.3. Use the analytic function $\mathcal{H}$ from Lemma A.13. Write $\beta_* := \beta_c(\tau_*) > 0$ and $\varepsilon = \tau - \tau_*$. Its local time expansions are

$$\chi(\tau_* + \varepsilon) = -\kappa\varepsilon + O(\varepsilon^2), \quad \psi(\tau_* + \varepsilon) = \psi_* + O(\varepsilon), \quad \beta_c(\tau_* + \varepsilon) = \beta_* + O(\varepsilon).$$

All remainders and their needed derivatives are uniform by the lemma. At $(z, \tau) = (0, \tau_*)$,

$$\mathcal{H}_z = 0, \qquad \mathcal{H}_{zz} = -\beta_* \psi_* < 0, \qquad \mathcal{H}_{z\tau} = -\frac{\beta_*}{2} \chi'(\tau_*) = \frac{\beta_* \kappa}{2} > 0.$$

The implicit function theorem applied to $\mathcal{H}_z = 0$ gives an analytic stationary branch $z_{\mathrm{f}}(\varepsilon)$ with

$$z_{\mathrm{f}}(0) = 0, \qquad z_{\mathrm{f}}'(0) = -\frac{\mathcal{H}_{z\tau}}{\mathcal{H}_{zz}} = \frac{\kappa}{2\psi_*} > 0.$$

After reducing the rectangle, $\mathcal{H}_{zz} < 0$ throughout it. For small $\varepsilon > 0$, this branch is a strict local maximum at positive z_{f}, and $a_{\mathrm{f}}^2 = z_{\mathrm{f}} = \kappa\varepsilon/(2\psi_*) + O(\varepsilon^2)$. Define $\beta_{\mathrm{f}}(\tau) = \mathcal{H}(z_{\mathrm{f}}, \tau)$. Substitution

into (69) gives

$$\begin{aligned}\beta_{\mathrm{f}} - \beta_c &= \frac{\beta_c}{2}\{\kappa\varepsilon z_{\mathrm{f}} - \psi_* z_{\mathrm{f}}^2\} + O(\varepsilon^3) \\ &= \frac{\beta_*}{2}\left\{\frac{\kappa^2}{2\psi_*} - \frac{\kappa^2}{4\psi_*}\right\}\varepsilon^2 + O(\varepsilon^3) = \frac{\beta_*\kappa^2}{8\psi_*}\varepsilon^2 + O(\varepsilon^3).\end{aligned}$$

The error estimate uses $z_{\mathrm{f}} = O(\varepsilon)$, so the $z^3 R$ term and the products of the displayed coefficient errors are all $O(\varepsilon^3)$. The leading constant is positive. Thus the local scale interval $(\beta_c, \beta_{\mathrm{f}})$ is nonempty.

We verify both branches throughout that interval before expanding them. For small positive ε, $\mathcal{H}_z(0,\tau) > 0$ and $\mathcal{H}_z$ strictly decreases in z on the chosen rectangle. Choose $z_R = 2\kappa\varepsilon/\psi_*$, which is inside it for small ε. Expansion gives

$$\mathcal{H}(z_R, \tau) - \beta_c = -\frac{\beta_*\kappa^2}{\psi_*}\varepsilon^2 + O(\varepsilon^3) < 0.$$

Hence every scale strictly between β_c and β_{f} has one root of $\mathcal{H}(z,\tau) = \beta$ in $(0, z_{\mathrm{f}})$ and one in (z_{f}, z_R); strict concavity makes both unique in this neighborhood. Write them z_-, z_+ and let $a_\pm = \sqrt{z_\pm}$. Oddness of the outer map turns each positive root into a distinct two-cycle on the invariant axis.

Now fix $\rho \in (0,1)$ and set $\beta = \beta_c + \rho(\beta_{\mathrm{f}} - \beta_c)$. Substitute $z = \varepsilon y$ and divide the branch equation by ε^2. The uniform analytic remainder yields

$$\frac{\beta_*}{2}(\kappa y - \psi_* y^2) - \rho\frac{\beta_*\kappa^2}{8\psi_*} + O(\varepsilon) = 0,$$

with the error uniform together with its derivative in y on bounded sets. Its limiting roots are

$$y_\pm^0 = \frac{\kappa}{2\psi_*}(1 \pm \sqrt{1-\rho}).$$

The derivatives at these roots are

$$\frac{\beta_*}{2}(\kappa - 2\psi_* y_\pm^0) = \mp\frac{\beta_*\kappa}{2}\sqrt{1-\rho} \neq 0.$$

The implicit function theorem therefore gives $y_\pm = y_\pm^0 + O(\varepsilon)$, proving (27). For ρ in any fixed compact subset of $(0,1)$, the two limiting roots stay positive, distinct, and in a bounded set, and the displayed derivatives are bounded away from zero. Compactness gives uniform neighborhoods and error bounds in ρ.

At any cycle amplitude a, differentiating $\nu_\tau(a) = 2a\mathcal{A}_\tau(a)$ and using $\mathcal{A}_\tau(a) = \beta$ gives the longitudinal multiplier

$$\lambda_x(a) = -1 - \frac{2a\mathcal{A}_\tau'(a)}{\beta} = -1 - \frac{4z\mathcal{H}_z(z,\tau)}{\beta}.$$

At z_+, $\mathcal{H}_z < 0$, so this is greater than -1; it is less than 1 for small ε by continuity. At z_-, $\mathcal{H}_z > 0$, so it is less than -1. More precisely, putting $q = \sqrt{1-\rho}$ in this calculation gives

$$\lambda_x(a_\pm) = -1 \pm \frac{\kappa^2}{\psi_*}(1 \pm q)q\,\varepsilon^2 + O(\varepsilon^3). \tag{147}$$

Indeed $\mathcal{H}_z = (\beta_c/2)\{\kappa\varepsilon - 2\psi_* z + O(\varepsilon^2)\}$ and $\beta_c/\beta = 1 + O(\varepsilon^2)$ on these branches. Substituting $z = y_\pm^0\varepsilon + O(\varepsilon^2)$ gives the formula.

The transverse multiplier is $1 - \kappa_\tau(a)/\beta$, where $0 < \kappa_\tau(a) < \delta$ by Lemma A.10. The symbol $\kappa_\tau(a)$ here is the transverse response defined there, not the constant $\kappa = -\chi'(\tau_*)$. Since $\tau_* > 16$, $\beta_c(\tau_*) > \delta$; shrinking the neighborhood keeps $\beta > \delta$. Both transverse multipliers lie in $(0,1)$. Coordinate reflections make both Jacobians diagonal and equal at reflected points. Thus the larger cycle has both second-iterate multipliers in $(0,1)$ and attracts a full-dimensional open set, while the smaller has one multiplier above one and one in $(0,1)$. At zero the longitudinal multiplier is $1 - 2\beta_c/\beta \in (-1,1)$; the transverse bound gives local stability there as well. The global IPO flow argument in the proof of Theorem B.10 uses only $\beta > \beta_c$, $\tau > 0$, and the product-coordinate

bounds. It therefore applies at every scale in this local interval, regardless of how that interval was constructed.

The meeting point is a nondegenerate fold of two-cycles for each small fixed $\varepsilon > 0$. At $a = a_{\mathrm{f}} > 0$, the amplitude derivatives satisfy $\mathcal{A}'_\tau(a) = 0$ and $\mathcal{A}''_\tau(a) = 4a^2\mathcal{H}_{zz}(a^2, \tau) < 0$. For $g = \mathsf{F}_{\mathrm{D},\tau,x}(\cdot, 0)$ at the fold, $g(a) = -a$ and $g'(\pm a) = -1$. Direct differentiation gives

$$\partial_\beta\{g(g(a)) - a\} = -4a/\beta \neq 0, \qquad \partial_s^2\{g(g(s)) - s\}|_{s=a} = 4a\mathcal{A}''_\tau(a)/\beta \neq 0.$$

For the second equality, oddness gives $g''(-a) = -g''(a)$ and $g''(a) = -\nu''_\tau(a)/\beta = -2a\mathcal{A}''_\tau(a)/\beta$. These identities verify the fold terminology for the second iterate.

The reward formula (8) gives

$$J_* - J(a, 0) = \frac{\gamma}{2}\tanh^2(a/2) = \frac{\gamma}{8}a^2 + O(a^4).$$

Inserting (27) yields (28), uniformly in the stated range of ρ. Finally, for $0 < s < a_-$, strict increase of $\mathcal{H}$ up to its maximum gives $0 < \mathcal{A}_\tau(s) < \beta$. Hence

$$|g(s)| = s\left|1 - \frac{2\mathcal{A}_\tau(s)}{\beta}\right| < s.$$

The same holds for negative s by oddness. From any initial $|s_0| < a_-$, the absolute iterates decrease to a limit. A positive limit would lie in a compact annulus inside $(-a_-, a_-)$ where the continuous absolute multiplier is uniformly below one, a contradiction. Thus every such point converges to zero. The boundary points $\pm a_-$ are a two-cycle and cannot be in its basin. A connected subset of the real line containing zero cannot pass either missing boundary point; therefore the connected component of this axial basin is exactly $(-a_-, a_-)$. This does not classify other, disconnected basin components or the full two-dimensional basin. □

D.25 Differentiable consistency of the GD discretization

Proof of Lemma A.11. Enlarge Θ to a compact rectangle and put $T = \tau + 1$. Write $G(w, \theta) = \nabla_w L_{\mathrm{D}}(w; \theta)$, $H = D_w G$, and $C = D_\theta G$, with the latter derivative taken at fixed w. Let $D = 2\sqrt{2}$. Since $\|d\| \le D$ and $|\sigma - q| \le 1$, $\|G\| \le D$, while $\|H\| \le L = 1/3$. Both the flow and its GD iterates up to time T obey $\|w\| \le DT$. On this compact state and parameter set choose a finite bound C_0 for $\|C\|$. The sensitivity $Z = D_\theta w$ solves

$$\dot{w} = -G(w, \theta), \qquad \dot{Z} = -H(w, \theta)Z - C(w, \theta), \qquad (w(0), Z(0)) = (0, 0).$$

Differentiating GD gives exactly forward Euler for this augmented system:

$$w_{j+1} = w_j - \eta G(w_j, \theta), \qquad Z_{j+1} = (\mathrm{Id} - \eta H(w_j, \theta))Z_j - \eta C(w_j, \theta).$$

The integral inequality for the flow gives $\|Z(u)\| \le C_0 u e^{Lu}$; induction in $\|Z_{j+1}\| \le (1 + \eta L)\|Z_j\| + \eta C_0$ gives $\|Z_j\| \le C_0 j\eta e^{Lj\eta}$. Thus all these states lie in a common compact convex set.

Let $Y = (w, Z)$ and let $\mathcal{E}_\theta(Y)$ denote its augmented vector field. Smoothness gives uniform bounds $\|\mathcal{E}_\theta\| \le B_E$ and a Lipschitz constant $L_E \ge 1$ on this set. For one exact flow step,

$$\begin{aligned}
&\|Y(u + \eta) - Y(u) - \eta\mathcal{E}_\theta(Y(u))\| \\
&\quad \le \int_0^\eta \|\mathcal{E}_\theta(Y(u + v)) - \mathcal{E}_\theta(Y(u))\|\, dv \\
&\quad \le \int_0^\eta L_E B_E v\, dv = \tfrac{1}{2}L_E B_E \eta^2.
\end{aligned}$$

For $e_j = \|Y_j - Y(j\eta)\|$, comparison with Euler consequently gives

$$e_{j+1} \le (1 + \eta L_E)e_j + \tfrac{1}{2}L_E B_E \eta^2, \qquad e_0 = 0.$$

Summing the geometric series,

$$e_K \le \tfrac{1}{2}B_E\eta\{(1 + \eta L_E)^K - 1\} \le \tfrac{1}{2}B_E\eta(e^{L_E T} - 1).$$

Also $\|Y(K\eta) - Y(\tau)\| \le B_E|K\eta - \tau|$. These estimates hold uniformly over $\theta \in \Theta$. Since the outer map and its derivative are $\theta + w/\beta$ and $\mathrm{Id} + Z/\beta$, division by β proves (63). □

D.26 Resolving the disappearing stability margin with finite GD

Proof of Proposition B.12. Throughout this proof, $\tau = K\eta = \tau_* + \varepsilon$ lies in a fixed compact interval around τ_*. On the invariant axis the linear inner coefficient is exactly

$$v_1^{\eta,K} = \frac{3}{4}\{1 - (1 - \eta/4)^K\} = 2\beta_c^{\eta,K}.$$

For η sufficiently small, Taylor's formula for the logarithm, with a remainder uniform on $0 \le \eta \le \eta_0 < 1$, gives

$$\begin{aligned} K\log(1-\eta/4) &= \frac{\tau}{\eta}\{-\eta/4 - \eta^2/32 + O(\eta^3)\} \\ &= -\tau/4 - \tau\eta/32 + O(\eta^2), \\ (1-\eta/4)^K &= e^{-\tau/4}\{1 - \tau\eta/32 + O(\eta^2)\}. \end{aligned}$$

Consequently

$$\beta_c^{\eta,K} - \beta_c(\tau) = \frac{3\tau e^{-\tau/4}}{256}\eta + O(\eta^2) = B_*\eta + O(\eta\varepsilon + \eta^2). \tag{148}$$

The shift is positive to first order. It comes from the exact linear recursion and does not rely on a perturbative cycle argument.

Write the finite-GD inner expansion as $\nu_{\eta,K}(s) = v_1^{\eta,K}s + v_3^{\eta,K}s^3 + v_5^{\eta,K}s^5 + O(s^7)$. The first part of Lemma A.15 gives errors $O(\eta)$ in each of these coefficients relative to the flow at τ. Since the linear coefficient stays bounded away from zero,

$$\begin{aligned} \chi_{\eta,K} &= -2v_3^{\eta,K}/v_1^{\eta,K} = \chi(\tau) + O(\eta) = -\kappa\varepsilon + O(\varepsilon^2 + \eta), \\ \psi_{\eta,K} &= -2v_5^{\eta,K}/v_1^{\eta,K} = \psi_* + O(\varepsilon + \eta). \end{aligned}$$

Let $\mathcal{H}_{\eta,K}(z) = \mathcal{A}_{\eta,K}(\sqrt{z})$. Its uniform expansion, including two differentiated remainders, is

$$\mathcal{H}_{\eta,K}(z) = \beta_c^{\eta,K}\{1 - \chi_{\eta,K}z/2 - \psi_{\eta,K}z^2/2\} + O(z^3). \tag{149}$$

If $\eta = o(\varepsilon)$, rescaling $z = \varepsilon y$ gives

$$\frac{\mathcal{H}_{\eta,K}(\varepsilon y) - \beta_c^{\eta,K}}{\varepsilon^2} \longrightarrow \frac{\beta_c(\tau_*)}{2}(\kappa y - \psi_* y^2) \quad \text{in } C^2 \tag{150}$$

on every bounded interval of $y \ge 0$. The coefficient error from χ is $O(\varepsilon + \eta/\varepsilon)$ after rescaling; that from ψ is $O(\varepsilon + \eta)$, and the differentiated remainder is $O(\varepsilon)$. Each tends to zero under the stated condition. This is the gain from using the exact discrete critical scale: there is no constant error of size η/ε^2.

The limiting quadratic is strictly concave and has its positive maximum at $y_{\mathrm{f}} = \kappa/(2\psi_*)$, with value C_*. Opposite derivative signs on either side of this point, followed by strict concavity, give a unique nearby local maximum of the rescaled finite-GD curve. Its location and value converge to those of the quadratic. This proves (103). At level ρC_*, the two simple limiting roots are

$$y_\pm = \frac{\kappa}{2\psi_*}(1 \pm \sqrt{1-\rho}).$$

Uniform C^1 convergence gives finite-GD roots $a_\pm^2 = \varepsilon y_\pm + o(\varepsilon)$ and makes the estimates uniform on compact subsets of $0 < \rho < 1$. With $q_\rho = \sqrt{1-\rho}$, differentiation at these roots yields

$$\begin{aligned} \lambda_x^{\eta,K}(a_\pm) &= -1 - \frac{4a_\pm^2\mathcal{H}'_{\eta,K}(a_\pm^2)}{\beta} \\ &= -1 \pm \frac{\kappa^2}{\psi_*}(1 \pm q_\rho)q_\rho\varepsilon^2 + o(\varepsilon^2). \end{aligned}$$

Thus the larger branch attracts longitudinally and its second-iterate multiplier is

$$1 - \frac{2\kappa^2}{\psi_*}(1 + q_\rho)q_\rho\varepsilon^2 + o(\varepsilon^2),$$

whereas the smaller branch is longitudinally unstable. Lemma A.8 gives the other entry as $1 - \kappa_{\eta,K}(a)/\beta \in (0,1)$ because $\beta \to \beta_c(\tau_*) > \delta$. The uniform two-dimensional C^1 convergence in Lemma A.15 gives its strict limit $1 - \kappa_{\tau_*}(0)/\beta_c(\tau_*) \in (0,1)$. The diagonal Jacobians and reflection symmetry therefore prove full attraction and a saddle, respectively.

At zero the longitudinal entry is $1 - 2\beta_c^{\eta,K}/\beta \in (-1,1)$; the transverse bound gives local stability. The uniqueness argument in Theorem B.2 uses only strict initial frozen loss decrease and applies to every $0 < \eta \le 3$, $K \ge 1$. The reward optimum remains zero by (8), and Lemma A.9 gives global convergence of matched IPO because $\beta > \beta_c^{\eta,K}$. Expanding reward as $\gamma a^2/8 + O(a^4)$ proves the claimed reward constants. Between zero and the smaller amplitude, $0 < \mathcal{A}_{\eta,K}(s) < \beta$, since the rescaled amplitude increases strictly up to its local maximum. Hence $|g(s)| < |s|$ there. A positive limiting absolute coordinate would contradict the strict continuous contraction on a compact annulus. The endpoints are a two-cycle and are outside the basin, so its connected component is exactly $(-a_-, a_-)$. This proves the first part.

For the unmodified flow scale, its excess above the discrete critical scale is, by (26) and (148),

$$\beta - \beta_c^{\eta,K} = \rho C_* \varepsilon^2 - B_* \eta + O(\varepsilon^3 + \eta\varepsilon + \eta^2). \tag{151}$$

If $\eta/\varepsilon^2 \to h < \infty$, then $\eta = o(\varepsilon)$ and the rescaled excess converges to $\rho C_* - B_* h$. For $h < \rho C_*/B_*$, divide by C_* to obtain the interior relative position ρ_{eff}. Applying (150) at this limiting level proves all the asserted branch, stability, reward, and basin results. If $h > \rho C_*/B_*$, the excess in (151) is negative for sufficiently small parameters. Then $1 - 2\beta_c^{\eta,K}/\beta < -1$, so zero is unstable. If the ratio tends to infinity, dividing the same identity by η gives a limit $-B_* < 0$ because $\varepsilon^2/\eta \to 0$, $\varepsilon \to 0$, and $\eta \to 0$. This proves instability in that case as well. At equality the leading terms cancel, so no conclusion follows without higher terms.

Finally, the interval $16.83612 < \tau_* < 16.83613$ and the outward exponential bounds enclose $3\tau e^{-\tau/4}/256$ on the entire interval. Combining that enclosure with the previously certified expression for C_* gives (105); the added certificate evaluates these expressions using exact interval arithmetic. For the rounding claim, $|K\eta - (\tau_* + e)| < \eta$ and bounded time derivatives give an additional $O(\eta)$ error. If $\eta = o(e^2)$, this error divided by e^2 vanishes; if the discrete critical scale is used, only the cubic coefficient error divided by e is needed. This proves the two stated rounding conclusions. □

D.27 Persistence of the cycle for small GD steps

Proof of Corollary B.11. Write $a = a_\tau(\beta) > 0$ and $E_n = \eta_n + |K_n\eta_n - \tau| \to 0$. By construction $\mathcal{A}'_\tau(a) < 0$. Choose $r \in (0, a/2)$ and $m > 0$ such that $\mathcal{A}'_\tau(s) \le -m$ on $[a - r, a + r]$ and $\mathcal{A}_\tau(a - r) > \beta > \mathcal{A}_\tau(a + r)$. Lemma A.11, restricted to this interval, implies convergence of the discrete amplitude curves $\mathcal{A}_n$ and their derivatives with error at most CE_n; division by $2s$ is harmless since $s \ge a/2$. For large n their endpoint inequalities persist and $\mathcal{A}'_n \le -m/2$. A unique intersection $a_n \in (a - r, a + r)$ exists. The mean value theorem gives

$$m|a_n - a| \le |\mathcal{A}_\tau(a_n) - \mathcal{A}_\tau(a)| = |\mathcal{A}_\tau(a_n) - \mathcal{A}_n(a_n)| \le CE_n.$$

Odd symmetry makes $\{(a_n, 0), (-a_n, 0)\}$ a two-cycle. Its longitudinal derivative tends to $1 - \nu'_\tau(a)/\beta \in (-1,1)$, and its transverse derivative lies in $(0,1)$ by Lemma A.8, since $\beta > \beta_{c,\tau} > \delta$ and eventually $\eta_n \le 3$. The disjoint-ball argument therefore gives an open attracting basin.

The elementary expansion $\log(1 - \eta_n/4) = -\eta_n/4 + O(\eta_n^2)$ and $K_n\eta_n^2 \to 0$ imply

$$\beta_{c,K_n}^{(\eta_n)} = \frac{c}{2}\{1 - (1 - \eta_n/4)^{K_n}\} \longrightarrow \frac{c}{2}(1 - e^{-\tau/4}) = \beta_{c,\tau} < \beta.$$

Hence the origin is locally stable for large n, remains the unique fixed point by strict initial loss descent, and equal-depth IPO converges globally by Lemma A.9. Finally $\gamma \tanh^2(s/2)/2$ has bounded derivative near a. Its mean value bound and $|a_n - a| \le CE_n$ give $|\Delta_n - \Delta_\tau(\beta)| \le C' E_n$, proving (100). □

D.28 Robustness under unrestricted finite-model perturbations

Proof of Theorem B.13. We first ensure that frozen GD remains a stable solver at every outer parameter, including parameters outside compact sets. Restrict $e(\xi) \le 1/8$. The two feature differences

from $(0,0)$ to $(1,0)$ and $(0,1)$ form a matrix $\mathrm{Id}+E$ with $\|E\| \le 2\sqrt{2}e(\xi) < 1$. They remain linearly independent. All weights e^{ℓ_i} are positive and all BT probabilities are strictly between zero and one. Consequently the finite family is minimal and both frozen losses are strictly convex and coercive, by Lemma A.1. Moreover, for every response, $\|\phi_i\| \le \sqrt{2} + 1/8 < 13/8$. For any outer law,

$$0 \prec \nabla^2 L_{\mathrm{D}} \preceq M_\theta^\xi/4 = \tfrac{1}{2}\,\mathrm{Cov}_\theta^\xi(\phi) \preceq \tfrac{1}{2}\mathbb{E}_\theta^\xi[\phi\phi^\top] \preceq \frac{169}{128}\mathrm{Id} \prec \frac{3}{2}\mathrm{Id}.$$

The IPO Hessian is $M_\theta^\xi/4$ and satisfies the same upper bound. The descent lemma at step one gives

$$L_X(w - \nabla L_X(w);\theta) \le L_X(w;\theta) - \tfrac{1}{4}\|\nabla L_X(w;\theta)\|^2. \tag{152}$$

Summation shows that gradient norms tend to zero. Coercivity bounds the iterates in a compact sublevel set. Every cluster point is stationary by continuity and hence is the unique minimizer. The full sequence therefore converges. This argument applies to every finite outer parameter.

All finite sums defining b_θ^ξ, $J_\xi(\theta)$, and the two inner gradients are real analytic in (ξ,θ,w). Finite composition gives joint analyticity of the three outer maps. On any fixed compact outer set, their values and first derivatives differ from those at ξ_0 by at most $Ce(\xi)$, using the mean value theorem in the finite-dimensional data vector. The same holds for any fixed number of outer compositions. Constants may depend on the chosen compact set.

At the base point, (118) gives

$$D_\theta b^{\xi_0}(0) = \mathrm{diag}(-c/4, -2c_z(0)/9),$$

which is invertible. The analytic implicit function theorem gives a root θ_ξ of $b_\theta^\xi = 0$, unique in a fixed neighborhood, with $\|\theta_\xi\| \le Ce(\xi)$. At this root every inner iterate is zero, so it is fixed for all three maps. Conversely, if $b_\theta^\xi \neq 0$, the first step in (152) strictly lowers the loss and later steps cannot increase it. Thus no positive finite-depth inner iterate can return to zero. Fixed points of all three maps are therefore exactly roots of b_θ^ξ.

At ξ_0 the three origin Jacobians have norm below one by (138). Choose a sufficiently small closed ball B_0 about zero on which all three base maps have derivative norm below one. By joint continuity, for sufficiently small $e(\xi)$ the perturbed derivative norms on B_0 are bounded by a common $q_0 < 1$. Also the images of zero differ from zero by $O(e(\xi))$. Shrinking the data neighborhood makes this error less than $(1-q_0)$ times half the radius of B_0, so each perturbed map takes B_0 into itself and is a contraction there. Its fixed point is θ_ξ, proving local asymptotic stability. The general Theorem B.4 applies to the perturbed minimal family at $b_{\theta_\xi}^\xi = 0$, proving second-order agreement for DPO-32 and IPO-32.

We next construct the cycle uniformly over perturbations. Write $p_\pm^0 = (\pm\log 3, 0)$ and $F_\xi = F_{\mathrm{D},32}^\xi$. The base cycle Jacobians have norm below one. Choose disjoint closed balls $B_\pm$ of radius r centered at $p_\pm^0$, on which the base derivative norms are below one. For sufficiently small perturbations,

$$\sup_{B_+\cup B_-} \|DF_\xi\| \le q < 1, \qquad \|F_\xi(p_\pm^0) - p_\mp^0\| \le Ce(\xi) < (1-q)r/2.$$

Hence for $\theta \in B_\pm$,

$$\|F_\xi(\theta) - p_\mp^0\| \le q\|\theta - p_\pm^0\| + Ce(\xi) < r.$$

Thus $F_\xi(B_\pm) \subset B_\mp$, and F_ξ^2 is a contraction on B_+ with constant at most q^2. It has a unique fixed point $p_{+,\xi}$ there. Put $p_{-,\xi} = F_\xi(p_{+,\xi}) \in B_-$. Then $F_\xi(p_{-,\xi}) = p_{+,\xi}$, and disjointness makes the period exactly two. Even and odd iterates from B_+ converge to these respective points. If $d_\pm = \|p_{\pm,\xi} - p_\pm^0\|$, the same estimates give $d_- \le qd_+ + Ce(\xi)$ and $d_+ \le qd_- + Ce(\xi)$. Consequently

$$\max(d_+, d_-) \le \frac{C}{1-q}e(\xi).$$

It remains to make the initialization set common to the comparison procedures. At ξ_0, both good procedures converge globally to zero. Choose a finite N such that their N-th iterates from p_+^0 lie in the ball concentric with B_0 of one quarter its radius. Continuity gives an open neighborhood U

of p_+^0, with compact closure contained in the interior of B_+, whose images under both base N-fold compositions lie in the half-radius ball. Joint continuity of these finite compositions makes their perturbed images lie in the three-quarter-radius ball, uniformly over $\overline{U}$, after decreasing ϵ_0. Subsequent iterates contract to θ_ξ in B_0. The bad procedure starts in B_+ and converges alternately to $p_{+,\xi}$ and $p_{-,\xi}$. Thus U is independent of the perturbation.

Joint smoothness of reward on the compact neighborhoods and the location bounds imply

$$|J_\xi(\theta_\xi) - J_{\xi_0}(0)| + \max_{\pm} |J_\xi(p_{\pm,\xi}) - J_{\xi_0}(p_\pm^0)| \le Ce(\xi).$$

The base difference is $\gamma/8$. Choosing $C\epsilon_0 \le \gamma/16$ proves (107). Because the reward values at the two perturbed points may differ, this is a statement about both phase limits, not necessarily one limiting reward value.

Finally, differentiating (8) gives

$$\nabla J_{\xi_0}(0) = 0, \qquad \nabla^2 J_{\xi_0}(0) = \operatorname{diag}(-\gamma/4, -2\delta/9) \prec 0.$$

The implicit function theorem applied to ∇J_ξ gives a stationary point $\theta_\xi^r = O(e(\xi))$ near zero. On a common small convex neighborhood, the perturbed Hessians remain negative definite and have norm at most L_J. Thus θ_ξ^r is a strict local maximum. Let $h = \theta_\xi - \theta_\xi^r = O(e(\xi))$. Taylor's integral formula and the zero gradient at θ_ξ^r give

$$0 \le J_\xi(\theta_\xi^r) - J_\xi(\theta_\xi) = -\int_0^1 (1-t) h^\top \nabla^2 J_\xi(\theta_\xi^r + th) h \, dt \le \tfrac{1}{2} L_J \|h\|^2 \le Ce(\xi)^2.$$

Taking the minimum of the finitely many positive neighborhood radii used above completes the uniform construction. □

D.29 Global guarantees within a structured perturbation family

Proof of Corollary C.2. The data in (112) depend continuously on the six parameters. A sufficiently small open neighborhood therefore lies within Theorem B.13, with all six parameters positive. The reward is

$$J(s,t) = -\gamma' \frac{2\cosh(us)}{\alpha + 2\cosh(us)} - \delta' \frac{2\cosh(vt)}{\zeta + 2\cosh(vt)}. \tag{153}$$

For positive a, the derivative of $2C/(a+2C)$ is $2a/(a+2C)^2 > 0$. Since $u, v > 0$, both hyperbolic cosines have their unique minimum at zero. Thus zero is uniquely globally reward-optimal. Reflection in either coordinate preserves the data and the reward, so $b_0 = 0$. The nearby common fixed point is consequently exactly zero. Reflection in z maps the ball B_+ of the preceding proof to itself and commutes with the second iterate. Uniqueness of its fixed point forces the cycle point in B_+ to have $t = 0$. Reflection in x sends it to the unique second-iterate fixed point in B_-, namely its partner. Thus the cycle is $\{(a_\xi, 0), (-a_\xi, 0)\}$ with $a_\xi > 0$, and its gap is at least $\gamma/16$ by the preceding theorem.

We prove the stronger global convergence of the comparison procedures. Define $c_x(t)$ and $c_z(s)$ as in (36)–(37), with the new marginal laws and rewards γ', δ'. The symmetry of the difference of iid squared coordinates, and the hyperbolic-tangent identity in the proof of Lemma A.2, give

$$0 < c_x(t) \le \tanh(\gamma'/2), \qquad 0 < c_z(s) \le \tanh(\delta'/2).$$

For $C = \cosh(us)$, $S = \sinh(us)$ and $Z = \alpha + 2C$, direct calculation gives

$$\operatorname{Var}(x) = \frac{2C}{Z} - \frac{4S^2}{Z^2} = \frac{2(\alpha C + 2)}{Z^2}, \qquad b_x = -\frac{2u\alpha c_x(t) S}{Z^2}, \qquad M_{xx} = \frac{4u^2(\alpha C + 2)}{Z^2}.$$

The same calculation applies to z with α, u replaced by ζ, v. The covariance matrix is diagonal. Therefore

$$w_{\mathrm{Q}}^* = \left(-\frac{2\alpha c_x(t)}{u} \frac{\sinh(us)}{\alpha\cosh(us) + 2}, -\frac{2\zeta c_z(s)}{v} \frac{\sinh(vt)}{\zeta\cosh(vt) + 2} \right). \tag{154}$$

Set $R(\alpha) = \max\{1, 2\alpha/(\alpha+2)\}$. Shrink the parameter neighborhood so that

$$u^2, v^2 < 2, \qquad \beta_{32} > \tfrac{1}{2} R(\alpha) \tanh(\gamma'/2), \qquad \beta_{32} > \tanh(\delta'/2). \tag{155}$$

These are strict open conditions and hold at the base point: use (39) and $\beta_{32} > 151/400 > 3/8 + 1/7200$ for the first scale inequality, and $\tanh(\delta/2) < 1/40 < \beta_{32}$ for the second.

For $C \geq 1$,

$$\frac{\alpha(C+1)}{\alpha C + 2} \leq R(\alpha).$$

Indeed its derivative is $\alpha(2-\alpha)/(\alpha C+2)^2$, and its value at one and limit at infinity are $2\alpha/(\alpha+2)$ and one. Thus for $q > 0$,

$$\frac{\alpha \sinh q}{q(\alpha \cosh q + 2)} = \frac{\alpha(\cosh q + 1)}{\alpha \cosh q + 2} \frac{\tanh(q/2)}{q} \leq \frac{R(\alpha)}{2}.$$

Equation (154) and (155) imply, for $s \neq 0$,

$$0 < \frac{|w^*_{\mathrm{Q},x}|}{\beta_{32}|s|} \leq \frac{R(\alpha)c_x(t)}{\beta_{32}} < 2.$$

For the other coordinate, $\sinh q \leq q \cosh q$ gives

$$\frac{\zeta \sinh q}{q(\zeta \cosh q + 2)} < 1, \qquad 0 < \frac{|w^*_{\mathrm{Q},z}|}{\beta_{32}|t|} < \frac{2c_z(s)}{\beta_{32}} < 2 \quad (t \neq 0).$$

Both displacements point toward zero, so exact IPO strictly decreases each nonzero coordinate in absolute value and preserves zero coordinates. For a finite depth L, the inner solution multiplies the two exact displacements by the diagonal entries of $\mathrm{Id} - (\mathrm{Id} - M/4)^L$. Since $x, z \in [-1, 1]$, their variances are at most one, giving $0 < M_{xx}/4 \leq u^2/2 < 1$ and $0 < M_{zz}/4 \leq v^2/2 < 1$. The two multipliers therefore belong to $(0, 1)$, preserving the strict absolute-coordinate decrease. All trajectories stay in their initial compact rectangle. If $V_n = |s_n| + |t_n|$ decreases to ℓ, any subsequential limit $\bar{\theta}$ has $V(\bar{\theta}) = \ell = V(F(\bar{\theta}))$ by continuity. Strict decrease outside zero forces $\bar{\theta} = 0$, proving full convergence. Formula (153) gives strict reward increase outside zero. The identity DPO-1=IPO-1 completes the comparison. All conditions were open inequalities; none reinstates reward calibration. □

D.30 Why global optimality needs structural restrictions

Proof of Proposition C.3. This perturbation has $e(\xi) = \epsilon$. For fixed $\epsilon > 0$, along $\theta = (0, t)$ the probability of response $(0, 1)$ is

$$p^{\epsilon}_{0,t}(0, 1) = \frac{2e^{(1+\epsilon)t}}{2e^{(1+\epsilon)t} + 2e^t + 4 + 4e^{-t}} \longrightarrow 1 \qquad (t \to \infty).$$

All other probabilities tend to zero. As rewards are fixed and bounded, $J_\epsilon(0, t) \to -\delta$, whence $J^*_\epsilon \geq -\delta$. Theorem B.13 gives $\theta_\epsilon \to 0$; continuity at this finite parameter gives $J_\epsilon(\theta_\epsilon) \to -\gamma/2 - 2\delta/3$. Consequently

$$\liminf_{\epsilon \downarrow 0} \{J^*_\epsilon - J_\epsilon(\theta_\epsilon)\} \geq -\delta + \gamma/2 + 2\delta/3 = \gamma/2 - \delta/3.$$

Calibration implies $3/4 \leq \tanh((\gamma + \delta)/2) < (\gamma + \delta)/2$, so $\gamma > 3/2 - \delta$ and $\gamma/2 - \delta/3 > 3/4 - 5\delta/6 > 0$ for $\delta = 1/20$. This proves the claim. The improvement occurs at parameters escaping to infinity as the perturbation vanishes, so it does not contradict the local smoothness and local reward bound proved above. □

D.31 Safeguarding without a bound on inner depth

Proof of Theorem 3.4. Write $d = (d_1, d_2)$. Under the product sampling law, d_1 and d_2 are independent, and each has a symmetric distribution: exchanging two iid draws changes the sign of their difference. The identity $\log(1 + e^v) = v/2 + \log(2 \cosh(v/2))$ gives

$$L_{\mathrm{D}}(w; \theta) = G_\theta(w) - b_\theta^\top w, \qquad G_\theta(w) = \mathbb{E}_\theta \log(2 \cosh(d^\top w/2)).$$

Independently changing the sign of either d_i leaves its joint law unchanged. Consequently G_θ is even in each coordinate separately. In particular, $\partial_i G_\theta(w_1, \ldots, 0, \ldots) = 0$. The fundamental theorem of calculus, varying only coordinate i, yields

$$\partial_i G_\theta(w) = H_i(\theta, w) w_i, \qquad H_i(\theta, w) = \int_0^1 \mathbb{E}_\theta[d_i^2 \sigma'(d^\top w^{(i,t)})], dt, \tag{156}$$

where $w^{(i,t)}$ equals w except that its i-th coordinate is tw_i. For finite θ, w, minimality and positive weights give $0 < H_i(\theta, w) \le M_{\theta,ii}/4 \le 1/3$. Thus unit-step GD obeys the componentwise identity

$$w_{j+1,i} = (1 - H_i(\theta, w_j))w_{j,i} + b_{\theta,i}. \tag{157}$$

Starting from zero, this implies that $w_{j,i}$ has the sign of $b_{\theta,i}$ and $|w_{j,i}| \ge |b_{\theta,i}|$ for every $j \ge 1$, unless $b_{\theta,i} = 0$, in which case every such coordinate is zero. Taking limits proves the same assertions for the exact minimizer. By (118), $b_{\theta,i}$ has the opposite sign to θ_i when $\theta_i \ne 0$, and vanishes exactly when $\theta_i = 0$. Since M_θ is diagonal, $z_{\theta,i} = 4b_{\theta,i}/M_{\theta,ii}$ has these same properties. Therefore the ratios in (30) are positive, $0 < a \le 1$, and the accepted displacement points toward zero in every nonzero coordinate.

We first prove invariance of each initial rectangle. Put

$$L_* = \max\left\{\frac{\tanh(\gamma/2)}{\beta}, \frac{2\tanh(\delta/2)}{\beta}\right\} < 2, \qquad \varepsilon_* = 2 - L_* > 0.$$

The strict inequality follows from (39), the assumed bound on β, and $\tanh(\delta/2) < 1/40$. For $s \ne 0$, the exact IPO formula and $|\tanh(s/2)| \le |s|/2$ give

$$\frac{|z_{\theta,1}|}{\beta|s|} \le \frac{c_x(t)}{\beta} \le \frac{\tanh(\gamma/2)}{\beta} \le L_*.$$

For $t \ne 0$, $\sinh|t| \le |t|\cosh t$ gives

$$\frac{|z_{\theta,2}|}{\beta|t|} \le \frac{2c_z(s)}{\beta}\frac{\cosh t}{\cosh t + 2} \le L_*.$$

If $D_i = a|w_i|/\beta$ denotes the accepted coordinate motion, then $0 < D_i \le |z_{\theta,i}|/\beta \le (2 - \varepsilon_*)|\theta_i|$ for every nonzero coordinate. A displacement toward zero of this size strictly decreases its absolute value. Zero coordinates remain zero. Thus all outer iterates stay in their initial rectangle, and (8) proves strict reward increase outside zero.

We now exclude stagnation uniformly over unbounded depth choices, including exact minimization. Fix $C = [-S, S] \times [-T, T]$, $S, T \ge 0$, and set

$$\begin{aligned} p_* &= \frac{e^{-2(S+T)}}{12}, \qquad q_* = \sigma(-\gamma - \delta), \qquad R = \frac{\log 2}{p_*^2 q_*}, \qquad h_* = 2p_*^2\sigma'(4R), \\ c_{x,*} &= \frac{\sinh\gamma}{\cosh\gamma + \cosh\delta}, \qquad c_{z,*} = \frac{\sinh\delta}{\cosh\delta + \cosh\gamma}, \\ m_* &= \min\left\{\frac{c_{x,*}}{(1+\cosh S)^2}, \frac{2c_{z,*}}{(1+2\cosh T)^2}\right\}, \qquad \rho_C = \min\{1/2,\, 3h_*m_*/\beta,\, \varepsilon_*\}. \end{aligned} \tag{158}$$

All these constants are strictly positive. For $\theta \in C$, each response probability is at least p_*: its unnormalized weight is at least $e^{-(S+T)}$, and the normalizer is at most $12e^{S+T}$. Every ordered comparison probability is consequently at least p_*^2. Also $q_{IJ}, 1 - q_{IJ} \ge q_*$ because the reward range is $\gamma + \delta$. For any real v and $q \in [q_*, 1 - q_*]$,

$$\log(1 + e^v) - qv \ge q_*|v|.$$

Indeed for $v \ge 0$ use $\log(1 + e^v) \ge v$, and for $v < 0$ use $\log(1 + e^v) \ge 0$. Selecting an ordered pair with feature difference $(1, 0)$ or $(0, 1)$ proves $L_{\mathrm{D}}(w; \theta) \ge p_*^2 q_* |w_i|$ for either coordinate. Every inner GD iterate and the exact minimizer satisfy $L_{\mathrm{D}}(w; \theta) \le L_{\mathrm{D}}(0; \theta) = \log 2$ by frozen descent. Thus $|w_i| \le R$, uniformly over all depths and all $\theta \in C$.

For any such w, every segment point $w^{(i,t)}$ remains in this box, so $|d^\top w^{(i,t)}| \le 4R$. The logistic derivative is even and decreasing in absolute argument. The ordered unit-difference pair and its reverse give $M_{\theta,ii} \ge 2p_*^2$. Equation (156) therefore gives $H_i(\theta, w) \ge h_*$. Together with (157), induction yields

$$|b_{\theta,i}| \le |w_{\mathrm{D},K,i}(\theta)| \le \frac{|b_{\theta,i}|}{h_*} \qquad (K \ge 1).$$

For the upper bound, if it holds at j, then $(1 - H_i)|w_{j,i}| + |b_{\theta,i}| \le (1 - h_*)|b_{\theta,i}|/h_* + |b_{\theta,i}| = |b_{\theta,i}|/h_*$. It holds initially at zero and passes to the exact minimizer by convergence. Also $|z_{\theta,i}| =$

$4|b_{\theta,i}|/M_{\theta,ii} \geq 3|b_{\theta,i}|$. Consequently each nonzero ratio in (30) is at least $3h_*$, and $a \geq 3h_*$, since $h_* \leq 1/3$.

For completeness, the positive constants in the bounds on b can be chosen independently of $\theta \in C$. The identity

$$\frac{1}{2}\left\{\tanh\frac{u+v}{2}+\tanh\frac{u-v}{2}\right\}=\frac{\sinh u}{\cosh u+\cosh v}$$

and symmetry of differences of iid squared coordinates show $c_x(t) \geq c_{x,*}$ and $c_z(s) \geq c_{z,*}$. Since $|\sinh s| \geq |s|$ and $|\sinh t| \geq |t|$, (118) then implies $|b_{\theta,i}| \geq m_*|\theta_i|$. For each nonzero coordinate we have proved

$$\frac{3h_*m_*}{\beta}|\theta_i| \leq D_i \leq (2-\varepsilon_*)|\theta_i|.$$

Because the displacement points toward zero,

$$\begin{aligned}|\theta_i| - |\theta_i^+| &= |\theta_i| - \big||\theta_i| - D_i\big| \\ &= \min\{D_i, 2|\theta_i| - D_i\} \geq \rho_C|\theta_i|.\end{aligned}$$

The same inequality is trivial for a zero coordinate. Summing and iterating proves (31). No estimate depends on the chosen depth, so arbitrary history-dependent depth choices are allowed. Finally, convexity and $0 < a \leq 1$ give

$$L_{\mathrm{D}}(aw;\theta) \leq (1-a)L_{\mathrm{D}}(0;\theta) + aL_{\mathrm{D}}(w;\theta) < L_{\mathrm{D}}(0;\theta)$$

when $\theta \neq 0$: the first inner step strictly decreases the loss, and later steps, or exact minimization, preserve that strict inequality. The constants in (158) are conservative certificates, not practical convergence-rate estimates. □

D.32 A COMMON RATIONAL COEXISTENCE INTERVAL

Proof of Corollary B.3. The additional certificate in Appendix F proves, simultaneously for $K = 31, 32, 33, 34$,

$$\mathcal{A}_K(1) > \frac{189}{500}, \qquad \mathcal{A}_K(6/5) < \frac{377}{1000}, \qquad \frac{2}{3} < \nu_K'(s) < \frac{3}{4} \quad (1 \leq s \leq 6/5). \tag{159}$$

For clarity, we justify both the use and the uniformity of these bounds. Partition $[1, 6/5]$ into the 100 rational intervals $[1 + j/500, 1 + (j+1)/500]$. The program encloses the complete scalar orbits at the two endpoints. On this strip $\cosh s < 2$, certified by the exponential enclosure. Equations (119)–(120) give $1 + f_v \geq 3/4 > 0$ and $f_s > 0$ for $v \geq 0$. Thus $s \mapsto \nu_j(s)$ is increasing by induction, and its endpoint enclosures contain the entire range on each subinterval. Starting from $u_0 = 0$, the interval recursion

$$u_{j+1} = (1 + f_v(s, \nu_j(s)))u_j + f_s(s, \nu_j(s))$$

therefore encloses ν_{j+1}' throughout that subinterval. It uses intervals for every occurrence of s and $\nu_j(s)$, retaining outward rounding. The strict derivative assertions at steps 31–34 on all 100 subintervals prove the third inequality of (159); the endpoint assertions prove the first two. Only rational arithmetic and proved exponential bounds enter these assertions.

Fix β in (81). The continuous function $H(s) = \nu_K(s) - 2\beta s$ is positive at 1 and negative at $6/5$. Moreover

$$H'(s) < \frac{3}{4} - \frac{754}{1000} = -\frac{1}{250} < 0 \qquad (1 \leq s \leq 6/5).$$

It consequently has a unique root a in $(1, 6/5)$, giving the symmetric two-cycle. At this root,

$$\lambda_x = 1 - \frac{\nu_K'(a)}{\beta} > 1 - \frac{3/4}{377/1000} = -\frac{373}{377} > -1, \qquad \lambda_x < 1 - \frac{2/3}{189/500} = -\frac{433}{567} < -\frac{3}{4}.$$

The uniform transverse lemma gives $\lambda_z = 1 - \kappa_K(a)/\beta \in (0, 1)$ because $0 < \kappa_K(a) < \delta < \beta$. Both coordinate reflections force a diagonal Jacobian at $(a, 0)$, and the partner has the same multipliers. Hence the two-cycle attracts an open neighborhood in two dimensions. The only fixed point is zero, since strict inner descent prohibits a return to the starting point unless $b_\theta = 0$. Its longitudinal multiplier is $1 - c(1 - (3/4)^K)/\beta \in (-1, 1)$ because $\beta > 377/1000 > 3/8$, and its transverse

multiplier belongs to $(0, 1)$ by the same transverse estimate at zero. The reward optimum and global comparison convergence follow from Lemma A.2, since $377/1000 > 3/8 + 1/7200$.

Finally, $e > 1 + 1 + 1/2 + 1/6 = 8/3$ implies $\tanh(1/2) = (e-1)/(e+1) > 5/11$. Therefore

$$J_* - J(a, 0) = \frac{\gamma}{2} \tanh^2(a/2) > \frac{25\gamma}{242} > \frac{\gamma}{10}.$$

The established bounds $151/400 < \beta_{32} < 236/625$ place β_{32} strictly inside (81). □

For the illustrated scale $\beta = 151/400$, the function `certify_illustrated_saddle` additionally encloses an intersection in $(491/1000, 123/250)$ for $K = 32$. Its endpoint signs have opposite order to those at the large cycle, and the derivative enclosure satisfies $\nu'_{32}(s) > 2\beta$ on that entire small interval. Hence $\nu_{32}(s) - 2\beta s$ has a unique zero there, with $g'_{32,\beta} < -1$. The transverse multiplier lies in $(0, 1)$ by Lemma A.8. This proves the unstable-cycle label in Figure 1; Corollary B.3 proves the attracting-cycle and stable-origin labels.

D.33 Constant damping and the IPO envelope

Proof. For any outer parameter in C, the inner sublevel argument in Theorem 3.4 gives $|w_{\mathrm{D},K,i}| \le |b_{\theta,i}|/h_*$, uniformly over depth, whereas $|z_{\theta,i}| \ge 3|b_{\theta,i}|$. Thus $a_0 \le 3h_*$ implies $|a_0 w_i| \le |z_{\theta,i}|$. The sign-preservation argument puts the next parameter in C, so induction justifies use of the same bounds at every round. Since $|w_{\mathrm{D},K,i}| \ge |b_{\theta,i}| \ge m_*|\theta_i|$, the accepted motion D_i satisfies

$$\frac{a_0 m_*}{\beta}|\theta_i| \le D_i \le (2 - \varepsilon_*)|\theta_i|.$$

The identity $x - |x - D| = \min\{D, 2x - D\}$ gives the asserted rate, and also strict reward increase and frozen-loss decrease as in the safeguard proof. This proves part 1 without a bound on the depth sequence.

For part 2, oddness permits restricting to $s > 0$. Since $\nu_K(s) > 0$,

$$\begin{aligned} |s - a_0\nu_K(s)/\beta| < s \quad &\iff \quad 0 < a_0\nu_K(s)/\beta < 2s \\ &\iff \quad a_0 \mathcal{A}_K(s) < \beta. \end{aligned}$$

Under the stated negative-cubic hypothesis, the amplitude maximum is attained at a positive point by Theorem B.2. The inequality at every $s > 0$ is therefore equivalent to $a_0 A_{\max} < \beta$. Equality maps a maximizer to its negative; oddness maps that negative point back. A larger a_0 violates the displayed strict decrease at this point. No assertion about all other orbits is needed.

For part 3, each active coordinate imposes exactly $a \le |z_{\theta,i}|/|w_i|$. Inactive coordinates impose no restriction. Intersecting these intervals with $[0, 1]$ gives exactly the minimum in (30). At depth one $w = b_\theta$, so every active ratio is $4/M_{\theta,ii} \ge 3$, and the minimum is one. □

D.34 Empirical perturbation bounds under batch reuse

Proof of Lemma A.5. For a triple (i, j, a), define the DPO gradient integrand $g_{ij,a}(w) = d_{ij}\{\sigma(d_{ij}^\top w) - a\}$. Since $\|d_{ij}\| \le D = 2\sqrt{2}$ and $0 < \sigma < 1$, $\|g_{ij,a}(w)\| \le D$ for every w. Writing the difference of expectations as a finite sum gives

$$\sup_w \left\| \nabla \widehat{L}_{\mathrm{D}}(w) - \nabla L_{\mathrm{D}}(w; \theta) \right\| \le \sum_{i,j,a} |\widehat{P}_{ij,a} - P_{\theta,ij,a}| D \le D\epsilon. \tag{160}$$

This is an inequality for every w on a fixed batch event; no independence between an iterate and that batch is invoked.

Let $T_{\mathrm{D}}(w) = w - \nabla L_{\mathrm{D}}(w; \theta)$. For any u, v,

$$T_{\mathrm{D}}(u) - T_{\mathrm{D}}(v) = \left(\mathrm{Id} - \int_0^1 \nabla^2 L_{\mathrm{D}}(v + a(u - v); \theta)\, da \right)(u - v).$$

The symmetric matrix in parentheses has eigenvalues in $[2/3, 1]$ by Lemma A.1. Hence T_{D} is nonexpansive. Splitting the empirical recursion into this population map and the gradient error gives

$$\|\widehat{w}_{\mathrm{D},j+1} - w_{\mathrm{D},j+1}\| \le \|\widehat{w}_{\mathrm{D},j} - w_{\mathrm{D},j}\| + D\epsilon.$$

Both sequences start at zero; induction gives the DPO bound in (108) after dividing by β.

For IPO, finite sums similarly imply

$$\left\|\widehat{M} - M_\theta\right\| \le D^2\epsilon, \qquad \left\|\widehat{b} - b_\theta\right\| \le (D/2)\epsilon.$$

For every empirical distribution, irrespective of its rank, $0 \preceq \widehat{M} \preceq D^2\mathrm{Id} = 8\mathrm{Id}$. Therefore $\widehat{R} = \mathrm{Id} - \widehat{M}/4$ has eigenvalues in $[-1, 1]$ and $\left\|\widehat{R}\right\| \le 1$. Also each coordinate of w^*_{Q} in (40) has absolute value less than two, because $c_x, c_z < 1$, $|\tanh(s/2)| < 1$, and $|\sinh t|/(\cosh t + 2) < 1$. Thus $\left\|w^*_{\mathrm{Q}}\right\| \le D$ and (41) gives $\|w_{\mathrm{Q},j}\| \le D$ for all j. Subtracting the two affine updates gives, with $e_j = \widehat{w}_{\mathrm{Q},j} - w_{\mathrm{Q},j}$,

$$e_{j+1} = (\mathrm{Id} - \widehat{M}/4)e_j - \tfrac{1}{4}(\widehat{M} - M_\theta)w_{\mathrm{Q},j} + (\widehat{b} - b_\theta),$$
$$\|e_{j+1}\| \le \|e_j\| + \left(\frac{D^3}{4} + \frac{D}{2}\right)\epsilon = \|e_j\| + \frac{5D}{2}\epsilon.$$

Starting from zero proves the IPO part of (108).

If $\epsilon \le 1/12$, then

$$\nabla^2 \widehat{L}_{\mathrm{D}}(w) \preceq \widehat{M}/4 \preceq M_\theta/4 + (D^2\epsilon/4)\mathrm{Id} \preceq (1/3 + 2\epsilon)\mathrm{Id} \preceq \mathrm{Id}/2.$$

The empirical IPO Hessian is exactly $\widehat{M}/4$ and obeys the same bound. The integral Taylor argument of Lemma A.1, with smoothness $1/2$ and step one, gives (50). No assertion of empirical strict convexity or a finite empirical logistic minimizer is needed for these finite-depth statements.

We finally prove the concentration bound. For a vector $v \in \mathbb{R}^S$, $\|v\|_1 = \max_{a\in\{-1,1\}^S} a^\top v$. For a fixed sign vector a, the random variable $Z = a_X$ under $X \sim P$ lies in $[-1, 1]$. To see its exponential bound directly, let $k(\lambda) = \log \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}$. Its second derivative is the variance of Z under the exponentially tilted law; this variance is at most one because Z remains in $[-1, 1]$. As $k(0) = k'(0) = 0$, integrating twice gives $k(\lambda) \le \lambda^2/2$ for $\lambda \ge 0$. For iid $Z_1, \dots, Z_N$, Markov's inequality and independence yield

$$\mathbb{P}\left\{N^{-1}\sum_{i=1}^N Z_i - \mathbb{E}Z > \epsilon\right\} \le \exp\{-N\lambda\epsilon + N\lambda^2/2\}.$$

Setting $\lambda = \epsilon$ gives $e^{-N\epsilon^2/2}$. A union bound over the 2^S sign vectors proves (51). Conditioning on any past history fixes the current law; the identical calculation then applies to the conditionally iid batch. □

D.35 A COMPLETE DETERMINISTIC TRACKING ARGUMENT

Proof of Lemma A.6. Choose $r_D > 0$ so that the closed balls $B_\pm = \overline{B}(a_\pm, r_D)$ are disjoint and $\sup_{B_+\cup B_-} \|DG\| \le q_D < 1$. Then for $\theta \in B_\pm$, $\|G(\theta) - a_\mp\| \le q_D \|\theta - a_\pm\|$. If a trajectory starts in B_+ and its additive errors have norm at most $(1 - q_D)r_D$, induction keeps it in the alternating balls and gives

$$\left\|\widehat{\theta}_{G,n} - a_{(-1)^n}\right\| \le q_D^n \|\theta_0 - a_+\| + \xi\sum_{j=0}^{n-1} q_D^j \le q_D^n \|\theta_0 - a_+\| + \frac{\xi}{1 - q_D}. \tag{161}$$

For both H_i, choose a common radius $r_0 > 0$ and $q_H < 1$ such that $\sup_{\overline{B}(0,r_0)} \|DH_i\| \le q_H$. Both H_i contract distance to zero on this ball. Because $H_i^n(a_+) \to 0$, choose a common integer $n_0 \ge 1$ with $\|p_{i,n_0}\| \le r_0/4$, where $p_{i,n} = H_i^n(a_+)$. Take a closed Euclidean ball $\mathcal{B}$ containing the closed unit neighborhoods of all the finitely many $p_{i,n}$ for $0 \le n \le n_0$. Enlarge it if necessary. On this convex compact ball, choose

$$L = \max\left(1, \sup_{\mathcal{B}} \|DH_1\|, \sup_{\mathcal{B}} \|DH_2\|\right), \qquad S_0 = \sum_{j=0}^{n_0-1} L^j > 0.$$

Set

$$u_0 = \min\left(\frac{r_D}{2}, \frac{r_0}{8L^{n_0}}, \frac{1}{4L^{n_0}}\right) > 0,$$
$$\xi_0 = \min\left(\frac{(1-q_D)r_D}{2}, \frac{(1-q_H)r_0}{2}, \frac{r_0}{8S_0}, \frac{1}{4S_0}\right) > 0, \qquad U = B(a_+, u_0).$$

For $\theta_0 \in U$, let $e_{i,n} = \left\|\widehat{\theta}_{H_i,n} - p_{i,n}\right\|$. While the perturbed states lie in $\mathcal{B}$, the segment formula and convexity of $\mathcal{B}$ give $e_{i,n+1} \le Le_{i,n} + \xi$. Starting with $e_{i,0} < u_0$ yields

$$e_{i,n} \le L^n u_0 + \xi \sum_{j=0}^{n-1} L^j \le L^{n_0} u_0 + \xi_0 S_0 \le \min(r_0/4, 1/2), \qquad 0 \le n \le n_0.$$

This is also an induction verifying the premise: each perturbed state is within $1/2$ of the reference state and hence inside $\mathcal{B}$. At n_0 it gives $\left\|\widehat{\theta}_{H_i,n_0}\right\| \le r_0/4 + r_0/4 = r_0/2$. For later rounds, errors at most $(1-q_H)r_0/2$ preserve $\overline{B}(0, r_0)$, since $q_H r_0 + \xi_0 \le (1+q_H)r_0/2 < r_0$. Iterating gives

$$\left\|\widehat{\theta}_{H_i,n}\right\| \le \frac{r_0}{2} q_H^{n-n_0} + \frac{\xi}{1-q_H}, \qquad n \ge n_0.$$

Take $q = \max(q_D, q_H) < 1$ and $C = r_0/2$. This proves both desired bounds, uniformly over U and over all admissible error sequences. □

D.36 Finite-sample reward separation and its limit

Proof of Theorem B.14. Apply Lemma A.6 with $G = F_{\mathrm{D},32}$, $H_1 = F_{\mathrm{D},1}$ and $H_2 = F_{\mathrm{Q},32}$. Its assumptions follow from Theorem 3.1 and (138). Let its constants be denoted U, ξ_0, q, n_0, r_0, using the construction in its proof. For the three procedures, (108) bounds all map errors by $C_F\epsilon$, with the explicit common choice

$$C_F = \frac{80D}{\beta_{32}}, \qquad \epsilon_0 = \min\left(\frac{1}{12}, \frac{\xi_0}{C_F}\right) > 0.$$

For each of the $3T$ adaptive rounds, the conditional failure probability of the batch event at tolerance ϵ is at most $2^{162}e^{-N\epsilon^2/2}$. Taking the expectation over its own history and then a union bound, with no independence assumption between these round events, gives total failure probability at most $3T\,2^{162}e^{-N\epsilon^2/2}$. Substitution of (109) makes this upper bound α. On the simultaneous event, Lemma A.6 applies with $\xi = C_F\epsilon$.

We supply the conversion from parameter to reward error explicitly. Put $a = \tanh^2(s/2)$ and $b = \tanh^2(t/2)$. Using $\cosh s = (1+a)/(1-a)$ and the corresponding identity for t,

$$J_* - J(s,t) = \frac{\gamma}{2}a + \frac{4\delta b}{3(3+b)}. \tag{162}$$

Since $|\tanh u| \le |u|$, this gives the global bound

$$0 \le J_* - J(s,t) \le \frac{\gamma}{8}s^2 + \frac{\delta}{9}t^2 \le C_J \left\|(s,t)\right\|^2, \qquad C_J = \max(\gamma/8, \delta/9). \tag{163}$$

For either H_i, the tracking bound and $(a+b)^2 \le 2a^2 + 2b^2$ therefore give

$$J_* - J(\widehat{\theta}_{H_i,n}) \le \frac{C_J r_0^2}{2} q^{2(n-n_0)} + \frac{2C_J C_F^2}{(1-q)^2}\epsilon^2.$$

On the compact union $B_+ \cup B_-$, let $L_J = \max_{B_+\cup B_-} \|\nabla J\| < \infty$. The straight segment from a state to its cycle center stays in the same ball. Because $J(a_+) = J(a_-)$,

$$\left|J_* - J(\widehat{\theta}_{G,n}) - \gamma/8\right| \le L_J q^n \left\|\theta_0 - a_+\right\| + \frac{L_J C_F}{1-q}\epsilon.$$

Choosing C as the maximum of the four displayed coefficients gives (110)–(111). Every chosen constant is independent of N, T, α.

For the triangular-array assertion take $\alpha_T = 1/T$ for $T \geq 2$. Then

$$\epsilon^2_{N_T,T,1/T} = \frac{2}{N_T}(162 \log 2 + \log 3 + 2 \log T) \longrightarrow 0.$$

For sufficiently large T the small-error condition holds; the failure probability tends to zero and $q^{T-n_0} \to 0$. This proves convergence in probability of all three gaps to the claimed constants. All reward gaps lie between zero and the finite reward range $\gamma + \delta$. If bounded random variables Z_T converge in probability to z, then for every $a > 0$,

$$\mathbb{E}|Z_T - z| \leq a + B\mathbb{P}(|Z_T - z| > a),$$

where B is a common bound on $|Z_T - z|$. First let $T \to \infty$ and then $a \downarrow 0$ to obtain convergence in expectation. This proves the last part without asserting any infinite-horizon result at fixed N. □

D.37 Strict opposition at the cycle

Proof. Lemma A.3 gives $v_j < v_\infty$ for every finite j, so $f(a, v_j) > 0$ and the inner gradient is nonzero. The strict loss inequality follows from (10). Let $\Delta_j = v_{j+1} - v_j = f(a, v_j)$. Since $f_v(a, v) < 0$ and v_j strictly increases, Δ_j strictly decreases while remaining positive. In particular $\Delta_{31} < \Delta_0 = v_1 \leq v_{31}$, whence

$$v_{31} < v_{32} = v_{31} + \Delta_{31} < 2v_{31}.$$

As $v_{32} = 2\beta_{32}a$ with $a = \log 3$, the two inequalities imply

$$-a < a - v_{31}/\beta_{32} < 0, \qquad a - v_{32}/\beta_{32} = -a.$$

The second coordinate remains zero. Formula (8) is strictly decreasing in $|s|$ at $t = 0$ for $|s| > 0$, giving the reward inequality.

Each inner iterate is a finite composition of continuous functions of the outer parameter, distribution probabilities, and additive errors. Evaluations of the loss and reward are also continuous. At the stated point both signed differences have a strictly positive margin. Therefore there is an open neighborhood of all these inputs on which both margins remain positive. For an empirical batch sufficiently close to the population distribution, this also proves the strict inequality for its own empirical DPO loss, by applying continuity to that empirical loss functional. This is a finite-step statement and does not assert global convergence of DPO-31. □

D.38 Persistence of the explicit cycle in the scale

Proof. Write $G_\beta = F_{\mathrm{D},32}$ with the dependence on β explicit, and define $\mathcal{H}(\theta, \beta) = G_\beta(G_\beta(\theta)) - \theta$. At (a_+, β_{32}) this function is zero. Its derivative in θ is

$$D_\theta \mathcal{H} = DG_{\beta_{32}}(a_-)DG_{\beta_{32}}(a_+) - \mathrm{Id}.$$

The product is diagonal with entries $\lambda^2_{\mathrm{cyc},x}, \lambda^2_{\mathrm{cyc},z} \in (0, 1)$, so this derivative is invertible. The implicit function theorem provides an analytic branch $a_+(\beta)$ of solutions and $a_-(\beta) = G_\beta(a_+(\beta))$ satisfies $G_\beta(a_-(\beta)) = a_+(\beta)$. The points remain distinct because they start a positive distance apart. By continuity the one-step Jacobian norms at the two points remain below one, and the same local contraction argument used in Theorem 3.1 proves attraction. The rewards are continuous and their gaps start at $\gamma/8$, so after shrinking the interval they both exceed $\gamma/16$. Choose the interval also inside $(3/8 + 1/7200, \infty)$. Lemma A.2 proves the stated global convergence there. The fixed-point argument using strict inner descent is independent of $\beta > 0$, and the inequalities in (138) stay strict. This proves all claims. □

D.39 Frozen-Fisher loss decrease and approximation order

Proof. Integrating the Hessian bound along an arbitrary increment v gives

$$L_{\mathrm{D}}(w + v; \theta) \leq L_{\mathrm{D}}(w; \theta) + g^\top v + \tfrac{1}{8} v^\top M_\theta v.$$

Substituting $v = -4M_\theta^{-1}g$ makes the last two terms $-4g^\top M_\theta^{-1}g + 2g^\top M_\theta^{-1}g$, proving (115). At zero the gradient is $-b_\theta$, hence $\bar{w}_1 = 4M_\theta^{-1}b_\theta = w_{\mathrm{Q}}^*$.

At $(w, \theta) = (0, \theta_*)$, the derivative in w of the equation $\nabla L_{\mathrm{D}}(w;\theta) = 0$ is $M_*/4 \succ 0$. The implicit function theorem gives a smooth minimizer branch $w_{\mathrm{D}}^*(\theta_* + h) = O(\|h\|)$; uniqueness identifies it with the exact minimizer. For fixed K, all $\bar{w}_j(\theta_* + h) = O(\|h\|)$ as well. Define $T_\theta(w) = w - 4M_\theta^{-1}\nabla L_{\mathrm{D}}(w;\theta)$. Because $\sigma'(u) = 1/4 + O(u^2)$ near zero,

$$\nabla^2 L_{\mathrm{D}}(w;\theta) = M_\theta/4 + O(\|w\|^2), \qquad D_w T_\theta(w) = O(\|w\|^2),$$

uniformly for θ near θ_*; M_θ^{-1} is uniformly bounded there. Along the segment between $\bar{w}_j$ and w_{D}^*, the derivative is $O(\|h\|^2)$. Since $T_\theta(w_{\mathrm{D}}^*) = w_{\mathrm{D}}^*$,

$$\bar{w}_{j+1} - w_{\mathrm{D}}^* = \left\{ \int_0^1 D_w T_\theta(w_{\mathrm{D}}^* + u(\bar{w}_j - w_{\mathrm{D}}^*))\, du \right\} (\bar{w}_j - w_{\mathrm{D}}^*).$$

The initial error $\bar{w}_0 - w_{\mathrm{D}}^*$ is $O(\|h\|)$; induction gives (116). This is a finite-order local statement, not a uniform-in-K asymptotic bound. □

The first correction has the explicit form

$$\begin{aligned} \bar{w}_2 - \bar{w}_1 &= 4M_\theta^{-1}\mathbb{E}_\theta \left[d \left\{ \frac{d^\top \bar{w}_1}{4} - \sigma(d^\top \bar{w}_1) + \frac{1}{2} \right\} \right] \\ &= \frac{1}{12} M_\theta^{-1}\mathbb{E}_\theta[d(d^\top \bar{w}_1)^3] + O(\|\bar{w}_1\|^5), \end{aligned} \tag{164}$$

locally where M_θ^{-1} is bounded. The first equality follows by subtracting $\nabla L_{\mathrm{Q}}(\bar{w}_1) = 0$ from the logistic gradient; the second follows from (122).

D.40 The one-step versus two-step frozen-Fisher separation

Proof. Since $4/M_{xx}(s, 0) = 2(1 + \cosh s)$, the scalar version of (114), obtained from (43), gives

$$\begin{aligned} \bar{\nu}_1(s) &= \frac{3}{2}\tanh(s/2), \\ \bar{\nu}_2(s) &= 2\bar{\nu}_1(s) - \frac{2\cosh s \tanh(\bar{\nu}_1(s)/2) + \tanh \bar{\nu}_1(s)}{1 + \cosh s}, \end{aligned} \tag{165}$$

where $\bar{\nu}_j = -\bar{w}_{j,x}$. At $a = \log 2$, $\cosh a = 5/4$, $\sinh a = 3/4$ and $\bar{\nu}_1(a) = 1/2$. Thus (117) equals $\bar{\nu}_2(a)/(2a)$ and defines the claimed cycle exactly. The certificate in Appendix F proves

$$\frac{471}{1250} < \bar{\beta} < \frac{377}{1000}, \qquad -\frac{197}{200} < 1 - \frac{\bar{\nu}_2'(a)}{\bar{\beta}} < -\frac{123}{125}. \tag{166}$$

For verification of the derivative, write $A = \tanh(1/4)$, $B = \tanh(1/2)$ and note $\bar{\nu}_1'(a) = 2/3$. If $N(s) = 2\cosh s \tanh(\bar{\nu}_1(s)/2) + \tanh \bar{\nu}_1(s)$, then

$$N'(a) = \frac{3}{2}A + \frac{2}{3}\left\{ \frac{5}{4}(1 - A^2) + (1 - B^2) \right\}.$$

The quotient rule in (165) gives

$$\begin{aligned} \bar{\nu}_2'(a) &= \frac{4}{3} - \frac{4}{9}N'(a) + \frac{4}{27}\left(\frac{5}{2}A + B \right) \\ &= \frac{2}{3} + \frac{10A^2 + 8B^2 - 8A + 4B}{27}, \end{aligned}$$

which is the expression used by the certificate.

For the transverse calculation, the zz preconditioner is three at $t = 0$, and its t derivative is zero there because M_{zz} is even. The mixed Hessian vanishes as in Lemma A.4. Let

$$H = \frac{4}{3}\mathbb{E}_{x,x'}[\sigma'(\bar{\nu}_1(s)(x - x'))] \in (0, 1/3].$$

The distribution derivative of the z gradient is still $\mathcal{N}(s)$, independent of the inner x coordinate. The two derivatives are therefore

$$\bar{U}_1 = -3\mathcal{N}(s), \qquad \bar{U}_2 = (1-3H)\bar{U}_1 - 3\mathcal{N}(s) = -3\mathcal{N}(s)(2-3H).$$

Since $0 < H \le 1/3$, $0 < 3\mathcal{N}(s)(2-3H) < 6\mathcal{N}(s) \le 2\delta/3$. With $\bar{\beta} > 3/8$ this gives

$$\frac{41}{45} < 1 + \bar{U}_2/\bar{\beta} < 1.$$

Both off-diagonal derivatives vanish by reflection; (166) then gives a contracting neighborhood at both cycle points. The lower bound $471/1250$ also exceeds $3/8 + 1/7200$. Thus global IPO convergence follows from Lemma A.2 and $\bar{F}_1$ is exactly that map. Strict loss decrease in (115) proves the fixed-point claim as in Theorem 3.1. At zero, all positive MM depths have Jacobian

$$\operatorname{diag}(1 - c/\bar{\beta},\ 1 - 2c_z(0)/(3\bar{\beta})),$$

by Proposition C.6; its entries lie in $(-1, 1)$. Finally $\cosh a = 5/4$ gives $J_* - J(\pm a, 0) = \gamma(5/9 - 1/2) = \gamma/18$.

For the cubic coefficient, exact IPO has $\bar{w}_{1,x} = -cs + cs^3/12 + O(s^5)$, and its linear outer coefficient crosses -1 at $\beta = c/2 = 3/8$. Its cubic coefficient there is $(c/12)/(c/2) = 1/6$. At zero, $M_{xx} = 1$ and $\mathbb{E}d_x^4 = 5/2$. Formula (164) adds the cubic outer coefficient

$$-\frac{c^3}{12(c/2)}\frac{5}{2} = -\frac{5c^2}{12} = -\frac{15}{64}.$$

This gives $1/6 - 15/64 = -13/192$ for $\bar{F}_2$. Proposition C.6 transfers this cubic coefficient to exact DPO and every $K \ge 2$. □

E Further Calculations

The results above use exact population updates or their finite-sample counterparts. This section records the effects of finite precision and the empirical form of the complementary frozen-Fisher update.

E.1 Finite precision in ordinary GD

Suppose an additional vector of norm at most ζ is added to each inner empirical update. In the DPO proof of Lemma A.5, the recursion becomes

$$\|\widetilde{w}_{\mathrm{D},j+1} - w_{\mathrm{D},j+1}\| \le \|\widetilde{w}_{\mathrm{D},j} - w_{\mathrm{D},j}\| + D\epsilon + \zeta.$$

For IPO use $\left\|\mathrm{Id} - \widehat{M}/4\right\| \le 1$ in the same way. Thus

$$\left\|\widetilde{F}_{\mathrm{D},K} - F_{\mathrm{D},K}\right\| \le \frac{K}{\beta}(D\epsilon + \zeta),$$
$$\left\|\widetilde{F}_{\mathrm{Q},K} - F_{\mathrm{Q},K}\right\| \le \frac{K}{\beta}(5D\epsilon/2 + \zeta).$$

Substitution into Lemma A.6 replaces the error scales in Theorem B.14 by a constant times $\epsilon + \zeta$. A positive reward separation therefore survives sufficiently small fixed ζ. The exact limiting gaps require $\zeta = \zeta_T \to 0$. The strict inequalities of Proposition C.4 survive small errors locally, but monotone empirical loss decrease is not asserted for arbitrary approximate updates at all states.

E.2 Empirical frozen-Fisher updates, including singular batches

Let $\widehat{M} = N^{-1}\sum_n d_n d_n^\top$ and $\widehat{b} = N^{-1}\sum_n d_n(A_n - 1/2)$ for a fixed batch, and use $\widehat{M}^\dagger$, its Moore–Penrose inverse, in (114). Let $\mathcal{S} = \operatorname{span}\{d_1, \dots, d_N\}$. All empirical gradients and $\widehat{b}$ lie in $\mathcal{S}$, which is the range of $\widehat{M}$; on $\mathcal{S}$ the matrix is positive definite. In particular,

$$\widehat{\bar{w}}_1 = 4\widehat{M}^\dagger\widehat{b}$$

is the minimum-norm minimizer of the empirical IPO loss. Indeed its normal equation is $\widehat{M}w = 4\widehat{b}$, and all solutions are this vector plus an element of $\ker \widehat{M}$. The empirical Hessian is bounded by $\widehat{M}/4$. Integrating that bound and substituting $v = -4\widehat{M}^\dagger \widehat{g}$ yields

$$\widehat{L}_{\mathrm{D}}(w+v) \le \widehat{L}_{\mathrm{D}}(w) - 2\widehat{g}^\top \widehat{M}^\dagger \widehat{g},$$

because $\widehat{g} \in \mathcal{S}$ and $\widehat{M}^\dagger \widehat{M} \widehat{M}^\dagger = \widehat{M}^\dagger$. Thus the one-step empirical identity and monotone inner descent hold for all batches, including rank-deficient ones.

For quantitative perturbation, suppose a compact set of outer states has $M_\theta \succeq m\mathrm{Id}$ with $m > 0$. If $\left\|\widehat{P} - P_\theta\right\|_1 \le \epsilon \le m/(2D^2)$, then

$$\widehat{M} \succeq (m - D^2\epsilon)\mathrm{Id} \succeq (m/2)\mathrm{Id},$$
$$\left\|\widehat{M}^{-1}\right\| \le 2/m, \qquad \left\|\widehat{M}^{-1} - M_\theta^{-1}\right\| = \left\|\widehat{M}^{-1}(M_\theta - \widehat{M})M_\theta^{-1}\right\| \le 2D^2\epsilon/m^2.$$

Since $\|b\| \le D/2$ and $\left\|\widehat{b} - b\right\| \le D\epsilon/2$,

$$\left\|\widehat{\bar{w}}_1 - \bar{w}_1\right\| \le B_1\epsilon, \qquad B_1 = \frac{4D}{m}\left(1 + \frac{D^2}{m}\right). \tag{167}$$

For the second step use $\|\nabla L_{\mathrm{D}}\| \le D$, the uniform sample gradient bound $D\epsilon$, and the population gradient Lipschitz bound $D^2/4$. Subtracting the two updates gives

$$\begin{aligned}\left\|\widehat{\bar{w}}_2 - \bar{w}_2\right\| &\le B_1\epsilon + \frac{8}{m}\left(D\epsilon + \frac{D^2}{4}B_1\epsilon\right) + \frac{8D^3}{m^2}\epsilon \\ &= B_2\epsilon, \qquad B_2 = \left(3 + \frac{2D^2}{m}\right)B_1.\end{aligned}$$

Dividing these bounds by $\bar{\beta}$ gives the outer map error. The same batch is used at both steps; the argument is uniform in w.

These bounds imply the analogue of Theorem B.14 for $\widehat{\bar{F}}_1$ and $\widehat{\bar{F}}_2$, with cycle center $(\log 2, 0)$, gap $\gamma/18$, and

$$\bar{\epsilon}_{N,T,\alpha} = \sqrt{\frac{2}{N}\left(162\log 2 + \log\frac{2T}{\alpha}\right)}.$$

Explicitly, for suitable U, C, q, n_0, ϵ_0 and all $n_0 \le n \le T$, with probability at least $1 - \alpha$,

$$0 \le J_* - J(\widehat{\bar{\theta}}_{1,n}) \le C\{q^{2(n-n_0)} + \bar{\epsilon}^2_{N,T,\alpha}\},$$
$$\left|J_* - J(\widehat{\bar{\theta}}_{2,n}) - \gamma/18\right| \le C\{q^n \left\|\theta_0 - (\log 2, 0)\right\| + \bar{\epsilon}_{N,T,\alpha}\}.$$

To verify that the compact-set assumption is not circular, take $\mathcal{K} = [-1,1]^2$. The elementary bound (121) gives $\cosh 1 < (11/4 + 3/8)/2 = 25/16$. For $|s|, |t| \le 1$,

$$M_{xx} > 32/41 > 3/4, \qquad M_{zz} > \frac{912}{1089} > 3/4,$$

where the second estimate uses that $4(C+2)/(1+2C)^2$ decreases for $C \ge 1$. Thus $m = 3/4$ is valid. The unperturbed IPO trajectory from $(\log 2, 0)$ has coordinates bounded by $\log 2 < 1$; all its finite burn-in points and both cycle points lie in the interior of $\mathcal{K}$. In the proof of Lemma A.6, shrink the initial neighborhood and error bound so that the finite burn-in deviations are less than $(1 - \log 2)/2$, and choose the origin and cycle balls inside $\mathcal{K}$. Then the one- and two-step error bounds apply at the initial state and inductively at every subsequent state. Conditional concentration and the reward calculation (163) finish the proof exactly as in Appendix D.36, with $C_F = B_2/\bar{\beta}$ and $\epsilon_0 \le m/(2D^2)$. The corresponding triangular-array limits are zero and $\gamma/18$ if $N_T/\log T \to \infty$; boundedness again gives convergence in expectation.

F EXACT RATIONAL CERTIFICATES

The finite numerical inequalities used to establish the specified cycle multipliers, common scale interval, and criticality thresholds are certified below using Python's standard-library `fractions.Fraction`. The program uses no floating-point numbers. All other parts of the proofs, including transverse attraction and stochastic tracking, are analytic.

F.1 WHY THE ENCLOSURE OPERATIONS ARE VALID

Let $G = 2^{100}$. For a rational a, define

$$\lfloor a \rfloor_G = G^{-1}\lfloor Ga \rfloor, \qquad \lceil a \rceil_G = -\lfloor -a \rfloor_G.$$

Every interval operation first computes its rational endpoint candidates exactly and then rounds its lower endpoint down and upper endpoint up. For multiplication it takes the minimum and maximum of the four endpoint products; these enclose every product by bilinearity on a rectangle. For division it first verifies that the divisor interval excludes zero and uses the reciprocal interval $[1/b, 1/a]$ for a divisor $[a, b]$. Addition and negation have their usual endpoint formulas. Therefore, inductively, every stored interval encloses the exact expression, even when dependent quantities appear multiple times. Such dependence may widen an interval but cannot invalidate its enclosure.

For $z \geq 0$ and $n \geq 0$ with $z < n + 2$, let $S_n(z) = \sum_{j=0}^{n} z^j/j!$. The first omitted term is $z^{n+1}/(n+1)!$, and each subsequent ratio is at most $z/(n+2) < 1$. All terms are nonnegative. Thus

$$S_n(z) \leq e^z \leq S_n(z) + \frac{z^{n+1}}{(n+1)!}\frac{1}{1 - z/(n+2)}. \tag{168}$$

The program uses $n = 48$ and verifies the domain condition at each call. For a nonnegative interval argument, monotonicity of $\tanh u = (e^{2u} - 1)/(e^{2u} + 1)$ turns the two endpoint bounds in (168) into a valid hyperbolic-tangent interval.

For $0 < z < 1$,

$$\log \frac{1+z}{1-z} = 2\sum_{j=0}^{\infty} \frac{z^{2j+1}}{2j+1}.$$

This follows by integrating the geometric series for $2/(1 - z^2)$ from zero to z. After the first n terms, its remainder obeys

$$0 < 2\sum_{j=n}^{\infty} \frac{z^{2j+1}}{2j+1} \leq \frac{2z^{2n+1}}{(2n+1)(1-z^2)}. \tag{169}$$

Use $z = 1/2$ for $\log 3$ and $z = 1/3$ for $\log 2$, again with $n = 48$. All inputs are rational.

The program evaluates the exact recurrences for v_j, u_j through 32 steps. An induction using the enclosure rules proves that the resulting intervals contain their mathematically defined values. Assertions that an entire interval lies strictly between the reported rational bounds therefore prove those strict inequalities. The depth-15 and depth-16 coefficients are evaluated directly as rational numbers. The continuous coefficient is enclosed at $\tau = 16, 17$; strict monotonicity, already proved analytically, turns these signs into a unique-root enclosure. The MM calculation and calibration-root enclosure work in the same way.

Table 3: Exact certified open intervals. Each bound is rational; the root intervals additionally use the monotonicity proofs in the text.

Quantity	Lower bound	Upper bound
β_{32}	$151/400$	$236/625$
$1-u_{32}/\beta_{32}$	$-179/200$	$-893/1000$
$\chi_{\mathrm{D},15},\ \eta=1$	$3/1000$	$1/250$
$\chi_{\mathrm{D},16},\ \eta=1$	$-11/1000$	$-1/100$
τ_*	16	17
$\bar{\beta}$	$471/1250$	$377/1000$
$1-\bar{\nu}_2'(\log 2)/\bar{\beta}$	$-197/200$	$-123/125$
γ	$19463/10000$	$2433/1250$

F.2 The refined threshold and fifth-order certificate

The function `certify_quintic_fold` first brackets the unique cubic zero by two rational times. It then encloses the entire time interval, evaluates the explicit fifth coefficient (67) and the analytic derivative (146), and certifies their signs and the fold-width coefficient. The new assertions use only rational arithmetic and the same outward-rounded exponential bounds as the other certificates.

The additional functions `certify_near_ipo_depths` and `certify_joint_gd_margin` check the adjacent cubic signs for the displayed mixture depths, the endpoint IPO fifth coefficient, and the discrete critical-scale shift and margin ratio. These checks support the finite numerical claims; the joint asymptotic rates follow from the analytic proofs.

F.3 Complete executable certificate

Run `python3 certificates.py` without Python's `-O` option (which would disable assertions). The complete source is embedded below. It prints the certified intervals as JSON and terminates normally only after every assertion has passed.

```
"""Exact rational interval certificates; no floating point is used in
    assertions."""
from fractions import Fraction as F

GRID = 1 << 100

def floor_grid(x):
    return F((x.numerator * GRID) // x.denominator, GRID)

def ceil_grid(x):
    return -floor_grid(-x)

class IV:
    def __init__(self, lo, hi=None):
        self.lo = floor_grid(F(lo))
        self.hi = ceil_grid(F(lo if hi is None else hi))
        assert self.lo <= self.hi

    @staticmethod
    def cast(x):
        return x if isinstance(x, IV) else IV(x)

    def __add__(self, other):
        y = self.cast(other)
        return IV(self.lo + y.lo, self.hi + y.hi)

    __radd__ = __add__

    def __neg__(self):
        return IV(-self.hi, -self.lo)
```

```
    def __sub__(self, other):
        return self + (-self.cast(other))

    def __rsub__(self, other):
        return self.cast(other) - self

    def __mul__(self, other):
        y = self.cast(other)
        corners = [self.lo*y.lo, self.lo*y.hi,
                   self.hi*y.lo, self.hi*y.hi]
        return IV(min(corners), max(corners))

    __rmul__ = __mul__

    def __truediv__(self, other):
        y = self.cast(other)
        assert y.lo > 0 or y.hi < 0
        return self * IV(1/y.hi, 1/y.lo)

    def __rtruediv__(self, other):
        return self.cast(other)/self

def exp_bounds(x, terms=48):
    assert x >= 0 and x < terms + 2
    term = F(1)
    total = term
    for j in range(1, terms + 1):
        term *= x / j
        total += term
    next_term = term * x / (terms + 1)
    return total, total + next_term / (1 - x/(terms + 2))

def tanh_iv(x):
    x = IV.cast(x)
    assert x.lo >= 0
    lower, _ = exp_bounds(2*x.lo)
    _, upper = exp_bounds(2*x.hi)
    return IV((lower-1)/(lower+1), (upper-1)/(upper+1))

def log3_iv():
    n = 48
    z = F(1, 2)
    lower = 2*sum((z**(2*j+1)/F(2*j+1) for j in range(n)), F(0))
    remainder = 2*z**(2*n+1)/(F(2*n+1)*(1-z*z))
    return IV(lower, lower+remainder)

def interval_inside(name, x, lower, upper):
    assert F(lower) < x.lo <= x.hi < F(upper), name
    return {"quantity": name, "strict_lower": str(F(lower)),
            "strict_upper": str(F(upper))}

def certify_cycle():
    v = IV(0)
    dv = IV(0)
    for _ in range(32):
        a = tanh_iv(v/2)
        b = tanh_iv(v)
        dv_next = (
            (F(13,16) + F(15,128)*a*a + F(9,128)*b*b)*dv
            + F(9,256) + F(3,64)*a + F(9,128)*b
        )
        v_next = v + F(9,64) - F(15,64)*a - F(9,128)*b
        v, dv = v_next, dv_next
        assert v.lo > 0 and v.hi < 1
```

```
    beta = v/(2*log3_iv())
    slope = 1 - dv/beta
    return [
        interval_inside("beta_32", beta, F(151,400), F(472,1250)),
        interval_inside("outer derivative at log(3)", slope,
                        F(-179,200), F(-893,1000)),
    ]

def cubic(k):
    r = F(3,4)
    u = k*(1-r)*r**(k-1)/(1-r**k)
    weighted = (
        (1-r)*sum(
            (r**(k-1-j)*(1-r**j)**3 for j in range(k)), F(0)
        )/(1-r**k)
    )
    return F(1,6) + u/2 - F(15,64)*weighted

def certify_cubic_sign():
    c15, c16 = cubic(15), cubic(16)
    assert F(3,1000) < c15 < F(1,250)
    assert F(-11,1000) < c16 < F(-1,100)
    return [
        {"quantity": "DPO cubic, K=15", "strict_lower": "3/1000",
         "strict_upper": "1/250"},
        {"quantity": "DPO cubic, K=16", "strict_lower": "-11/1000",
         "strict_upper": "-1/100"},
    ]

def continuous_cubic(tau):
    x = F(tau, 4)
    ex = IV(*exp_bounds(x))
    em = IV(1)/ex
    return (-F(13,192) + F(77,64)*x/(ex-1)
            - F(75,128)*em + F(15,128)*em*em)

def certify_continuous_threshold():
    left, right = continuous_cubic(16), continuous_cubic(17)
    assert left.lo > 0 and right.hi < 0
    return {"quantity": "unique continuous inner-time cubic threshold",
            "strict_lower": "16", "strict_upper": "17"}

def log2_iv():
    n = 48
    z = F(1, 3)
    lower = 2*sum((z**(2*j+1)/F(2*j+1) for j in range(n)), F(0))
    remainder = 2*z**(2*n+1)/(F(2*n+1)*(1-z*z))
    return IV(lower, lower+remainder)

def certify_mm_cycle():
    a, b = tanh_iv(F(1,4)), tanh_iv(F(1,2))
    nu2 = 1 - F(10,9)*a - F(4,9)*b
    beta = nu2/(2*log2_iv())
    derivative = F(2,3) + (10*a*a + 8*b*b - 8*a + 4*b)/27
    slope = 1 - derivative/beta
    assert nu2.lo > F(1,2) and nu2.hi < 1
    return [
        interval_inside("MM beta", beta, F(471,1250), F(377,1000)),
        interval_inside("MM outer derivative at log(2)", slope,
                        F(-197,200), F(-123,125)),
    ]

def certify_gamma():
    delta = F(1,20)
    def calibration(g):
```

```
        return (F(5,9)*tanh_iv(g/2)
                + F(2,9)*tanh_iv((g-delta)/2)
                + F(2,9)*tanh_iv((g+delta)/2))
    lo, hi = F(19463,10000), F(2433,1250)
    assert calibration(lo).hi < F(3,4)
    assert calibration(hi).lo > F(3,4)
    return {"quantity": "unique gamma", "strict_lower": str(lo),
            "strict_upper": str(hi)}

def sinh_cosh_iv(s):
    s = IV.cast(s)
    assert s.lo >= 0
    el = IV(*exp_bounds(s.lo))
    er = IV(*exp_bounds(s.hi))
    cl, cr = (el+1/el)/2, (er+1/er)/2
    sl, sr = (el-1/el)/2, (er-1/er)/2
    return IV(sl.lo, sr.hi), IV(cl.lo, cr.hi)

def scalar_step(s, v):
    S, C = sinh_cosh_iv(s)
    a, b = tanh_iv(v/2), tanh_iv(v)
    return v + (F(3,2)*S-2*C*a-b)/(2*(1+C)*(1+C))

def endpoint_orbit(s):
    S, C = sinh_cosh_iv(s)
    v = IV(0)
    result = [v]
    for _ in range(34):
        a, b = tanh_iv(v/2), tanh_iv(v)
        v = v + (F(3,2)*S-2*C*a-b)/(2*(1+C)*(1+C))
        assert v.lo > 0
        result.append(v)
    return result

def certify_common_interval(parts=100):
    # Uniform derivative enclosure on [1,6/5], not a pointwise grid check
    .
    beta_lo, beta_hi = F(377,1000), F(189,500)
    left = endpoint_orbit(F(1))
    for k in range(31,35):
        assert (left[k]/2).lo > beta_hi
    for j in range(parts):
        slo = 1+F(j,5*parts)
        shi = 1+F(j+1,5*parts)
        right = endpoint_orbit(shi)
        S, C = sinh_cosh_iv(IV(slo,shi))
        assert C.hi < 2
        u = IV(0)
        for k in range(34):
            # nu_k is increasing on this strip by f_s>0 and 1+f_v>0.
            v = IV(left[k].lo,right[k].hi)
            a, b = tanh_iv(v/2), tanh_iv(v)
            fv = -(C*(1-a*a)+1-b*b)/(2*(1+C)*(1+C))
            fs = (F(3,2)*(2-C)*(1+C)+2*S*((C-1)*a+b)) / (2*(1+C)*(1+C)
    *(1+C))
            u = (1+fv)*u+fs
            if k+1 >= 31:
                assert F(2,3) < u.lo <= u.hi < F(3,4), (j,k+1,str(u.lo),
    str(u.hi))
        left = right
    for k in range(31,35):
        assert (left[k]/F(12,5)).hi < beta_lo
    return {"quantity":"common coexistence interval, K=31,32,33,34",
            "strict_lower":str(beta_lo),"strict_upper":str(beta_hi),
            "amplitude_lower":"1","amplitude_upper":"6/5",
```

```
            "nu_derivative_lower":"2/3","nu_derivative_upper":"3/4"}

def certify_illustrated_saddle():
    slo, shi, beta = F(491,1000), F(123,250), F(151,400)
    left, right = endpoint_orbit(slo), endpoint_orbit(shi)
    assert (left[32]/(2*slo)).hi < beta
    assert (right[32]/(2*shi)).lo > beta
    S, C = sinh_cosh_iv(IV(slo,shi))
    assert C.hi < 2
    u = IV(0)
    for k in range(32):
        v = IV(left[k].lo,right[k].hi)
        a, b = tanh_iv(v/2), tanh_iv(v)
        fv = -(C*(1-a*a)+1-b*b)/(2*(1+C)*(1+C))
        fs = (F(3,2)*(2-C)*(1+C)+2*S*((C-1)*a+b))/(2*(1+C)*(1+C)*(1+C))
        u = (1+fv)*u+fs
    assert u.lo > 2*beta
    return {"quantity":"illustrated unstable cycle amplitude, K=32, beta
    =151/400",
            "strict_lower":str(slo),"strict_upper":str(shi)}


def continuous_quintic_data(tau_lo, tau_hi):
    x = IV(tau_lo, tau_hi)/4
    ex = IV(exp_bounds(x.lo)[0], exp_bounds(x.hi)[1])
    e = 1/ex
    p1 = (-355740*x+838380)*x-227929
    p2 = 636120-831600*x
    p3 = -529650+207900*x
    numerator = ((((-21141*e+181440)*e+p3)*e+p2)*e+p1)*e-38840
    fifth_inner = numerator/2621440
    psi = -8*fifth_inner/(3*(1-e))
    kappa = (-F(77,256)*e*(1-x-e)/((1-e)*(1-e))
             -F(75,512)*e+F(15,256)*e*e)
    beta = F(3,8)*(1-e)
    fold_width_coefficient = beta*kappa*kappa/(8*psi)
    return psi, kappa, fold_width_coefficient

def certify_quintic_fold():
    lo, hi = F(420903,25000), F(1683613,100000)
    assert continuous_cubic(lo).lo > 0
    assert continuous_cubic(hi).hi < 0
    psi, kappa, width = continuous_quintic_data(lo,hi)
    return [
        {"quantity":"unique continuous cubic zero, refined",
         "strict_lower":str(lo),"strict_upper":str(hi)},
        interval_inside("positive fifth outer coefficient at cubic zero",
     psi,
                        F(8681,100000), F(8683,100000)),
        interval_inside("negative time derivative of cubic at zero",
    kappa,
                        F(1268,100000), F(1269,100000)),
        interval_inside("quadratic coefficient of local fold width",
    width,
                        F(855,10000000), F(857,10000000))
    ]


def mixture_cubic(lam, k, c=F(2,3), vartheta=F(1,12)):
    alpha = 1+vartheta*lam
    r = 1-1/(alpha+2)
    u = k*(1-r)*r**(k-1)/(1-r**k)
    w = (1-r)*sum((r**(k-1-j)*(1-r**j)**3
                   for j in range(k)), F(0))/(1-r**k)
    big_lam = 2*alpha*alpha*(alpha+8)*c*c/(3*(alpha+2)**3)
```

```
    gap = lam*big_lam-2*(alpha-1)/(3*(alpha+2))
    linear = (4-alpha)/(alpha+2)
    assert gap > 0 and linear > 0
    assert u <= 1-w <= 3*u
    value = -gap+linear*u+lam*big_lam*(1-w)
    assert -gap+(linear+lam*big_lam)*u <= value
    assert value <= -gap+(linear+3*lam*big_lam)*u
    return value

def certify_near_ipo_depths():
    result = []
    for power, k in [(1,18),(2,24),(3,30),(4,37),(5,43)]:
        lam = F(1,10**power)
        assert mixture_cubic(lam,k-1) > 0
        assert mixture_cubic(lam,k) < 0
        result.append({"quantity":"first negative mixture cubic depth",
                       "lambda":str(lam),"depth":k,
                       "c":"2/3","vartheta":"1/12"})
    # Exact Taylor division for the limiting IPO fifth coefficient.
    q1 = F(1,3)
    q3 = (F(1,6)-q1/2)/3
    q5 = (F(1,120)-q3/2-q1/24)/3
    assert q3 == 0 and -6*q5 == F(1,90)
    return result

def certify_joint_gd_margin():
    lo, hi = F(420903,25000), F(1683613,100000)
    tau = IV(lo,hi)
    x = tau/4
    ex = IV(exp_bounds(x.lo)[0],exp_bounds(x.hi)[1])
    shift = 3*tau/(256*ex)
    _, _, width = continuous_quintic_data(lo,hi)
    return [
        interval_inside("linear discrete critical-scale shift", shift,
                        F(2931,1000000),F(2933,1000000)),
        interval_inside("flow-scale GD margin coefficient C_star/B_star",
                        width/shift,F(291,10000),F(293,10000))
    ]


if __name__ == "__main__":
    import json
    result = (certify_cycle() + certify_cubic_sign()
              + certify_near_ipo_depths() + certify_joint_gd_margin()
              + [certify_continuous_threshold()] + certify_quintic_fold()
              + certify_mm_cycle()
              + [certify_gamma(), certify_common_interval(),
                 certify_illustrated_saddle()])
    print(json.dumps(result, indent=2))
```