SPECTRAL STARVATION IN ON-POLICY DISTILLATION

Anonymous authors
Paper under double-blind review

ABSTRACT

On-policy distillation can improve from a small selected query set while absorbing teacher supervision progressively more slowly. We study this phenomenon through the spectrum induced by student-generated prefixes. Specifically, for every smooth f-divergence over tokens, the local OPD operator is an occupancy-weighted pullback of categorical Fisher geometry. The exact cross-entropy Hessian separates this Gauss-Newton operator from a teacher-gap-weighted network-curvature remainder. Under this local geometry, the linearized dynamics show that the step-to-step decline in absorption equals a normalized variance of active contraction factors, while the continuous log-rate derivative equals minus four times the active eigenvalue variance. Fast modes disappear first and leave the residual concentrated on weak directions. Kernel components form an invariant floor, and the smallest active eigenvalue gives the finite-time envelope and limiting rate. Consequently, equal-budget query choice is horizon dependent, while changing-operator bounds characterize the validity window of a frozen spectral forecast. Experimentally, in 160 unclipped Transformer runs, the initial student-Fisher predictor explains short-horizon gap reduction with $R^2 = 0.906$ and rank correlation 0.925. Across 64 matched replays, tangent drift tracks frozen-forecast error with rank correlation 0.504 and seed-cluster interval $[0.392, 0.623]$. The broader 1,088-run study spans GSM8K and ARC-Challenge; all eight ARC selector and training-mode conditions exhibit positive starvation strength with intervals bounded away from zero. Consistent with this local-validity picture, direct measurements find order-one occupancy-induced operator replacement by update 64. Finally, after a common 32-candidate screen and at equal 256 online teacher calls, four selected Residual E prompts absorb more maximum teacher gap than sixteen selected random prompts. Explicit accounting specifies screening, online-trajectory, and token-position cost.

1 INTRODUCTION

On-policy distillation (OPD) trains a student on its own generated prefixes and supplies dense teacher supervision at the states the student actually visits (Agarwal et al., 2024; Gu et al., 2024). This learner-generated distribution addresses a central weakness of static sequence distillation: a fixed corpus can describe teacher behavior on expert trajectories while leaving the student's mistakes and off-trajectory states unseen. OPD instead couples data collection to the current student. Specifically, a teacher query is therefore more than one labeled example; and it induces a rollout, a prefix occupancy, and a structured collection of token-level gradients.

However, this coupling creates a puzzle: A small query set can generate many supervised states, yet hundreds of updates may still be required to close the teacher-student gap (Fu et al., 2026; Zhang et al., 2026; Zhao et al., 2026). Query count and supervision volume do not determine how much of the remaining discrepancy an update can absorb. Repeated trajectories can excite nearly parallel parameter directions, whereas a few complementary trajectories can span the directions carrying the residual teacher gap. Fig. 1(a) depicts this mismatch between abundant supervision and diminishing absorption.

To resolve this puzzle, we develop a spectral account of OPD dynamics. Specifically, at a fixed student checkpoint, each query distribution induces a prefix occupancy measure. The occupancy-weighted pullback of categorical Fisher curvature through the student logit Jacobian defines a positive semidefinite local OPD operator in the trainable coordinates. Its spectrum records which teacher-discrepancy directions receive strong updates, which receive weak updates, and which are absent

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.

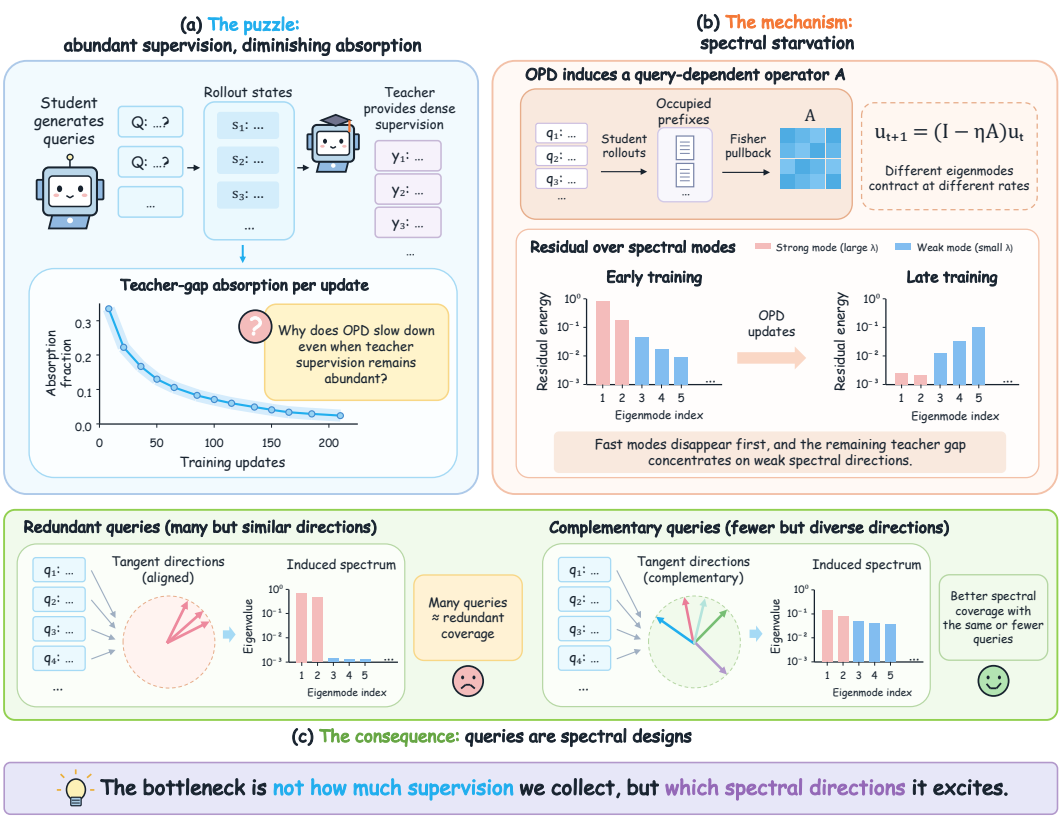


Fig. 1. Overview of spectral starvation in on-policy distillation. Student rollouts may provide abundant teacher-supervised prefixes while repeatedly exciting similar tangent directions. The induced OPD operator contracts strong modes first, concentrating the residual teacher gap on weak modes and reducing absorption per update. Query sets with complementary tangent directions improve spectral coverage at the same or smaller selected-query budget.

from the current tangent span. Fig. 1(b) summarizes the mechanism: strong modes contract first, so the residual becomes progressively concentrated in weak modes.

Under this fixed-operator view, we call the resulting residual-composition effect *spectral starvation*. In exact linearized dynamics, the fraction of squared teacher gap absorbed per update is nonincreasing. Its discrete decline is exactly a normalized variance of the active contraction factors, and the derivative of the continuous log-absorption rate is minus four times the active eigenvalue variance. Heterogeneous spectra therefore generate strict slowdown, a single active eigenvalue gives constant fractional absorption, and a kernel component forms an invariant residual floor. These results jointly characterize the initial slope, the complete absorption curve, and the terminal rate.

As a consequence, the same operator turns query selection into spectral experimental design. Gradient energy measures immediate progress on the teacher gap currently present. The smallest active eigenvalue protects the weakest relevant direction and controls the limiting rate. These objectives can rank equal-budget query sets differently, so the best design depends on the optimization horizon. Fig. 1(c) illustrates the consequence: complementary trajectories can cover weak tangent directions more effectively than a larger set of redundant queries. Gradient norm, E optimality, and residual-weighted spectral selection emerge as distinct horizon-dependent criteria.

Beyond this frozen-operator picture, fully on-policy training adds a second timescale because each update changes both the student and its future occupancy. The relevant object is the operator path $A_0, A_1, \ldots$. We derive a perturbation bound separating contraction supplied by the instantaneous spectrum from occupancy, tangent, and nonlinear drift. This yields a falsifiable local claim: A_0 should predict the short neural future while the tangent geometry remains near its reference, and forecast

error should grow with measured operator replacement. For complete notation and definitions of the measured objects, please refer to Appendix A.

Guided by this decomposition, our experiments proceed from exact finite systems to neural OPD. An exact finite system verifies the modal identities, residual floor, and equal-budget horizon reversal. The neural study uses Qwen3.5 students and teachers, GSM8K and ARC-Challenge prompts, ten query selectors, fixed-prefix SGD, fully on-policy AdamW, and sixteen training seeds. The initial student-Fisher operator predicts short-horizon teacher-gap reduction across 160 Transformer runs with $R^2 = 0.906$ and rank correlation 0.925. Across 64 matched replays, tangent drift tracks frozen-forecast error with rank correlation 0.504 and seed-cluster interval $[0.392, 0.623]$. The temporal starvation signature persists across mathematics and science, fixed and regenerated prefixes, and a larger student. At equal online teacher calls, four Residual E queries absorb 0.078 more maximum teacher gap than sixteen selected random queries, with paired interval $[0.009, 0.152]$.

This paper makes three contributions:

- We derive the local OPD absorption operator and prove exact modal, variance, kernel, and perturbation results that characterize spectral starvation.
- We separate query count from tangent coverage, prove equal-budget horizon reversal, and formulate horizon-aware query choice through gradient energy and E optimality.
- We calibrate the frozen spectral forecast in 160 neural runs, measure its drift-controlled breakdown, and test the starvation signature across 1,088 math and science runs.

2 Related Work

Knowledge and sequence distillation. Classical knowledge distillation transfers a teacher predictive distribution into a smaller student (Hinton et al., 2015), while sequence-level distillation replaces or augments token targets with teacher-generated sequences (Kim & Rush, 2016). Policy distillation extends the same principle to sequential decisions (Rusu et al., 2015; Czarnecki et al., 2019). These methods differ in target construction, but their training distributions are commonly fixed before an update. Our analysis concerns the additional geometry created when the current student generates the states on which token-level distillation is performed.

On-policy and interactive distillation. MiniLLM, GKD, and related approaches train on student-generated sequences to reduce exposure mismatch (Gu et al., 2024; Agarwal et al., 2024). Recent work studies prompt efficiency, prefix reuse, and prefix-guided supervision (Fu et al., 2026; Zhang et al., 2026; Zhao et al., 2026). We complement these algorithmic results by identifying the occupancy-weighted operator induced by a query set and by characterizing how its active spectral measure changes during optimization.

Learning geometry and data design. Neural tangent and Fisher methods describe local optimization through Jacobian pullbacks and predictive curvature (Jacot et al., 2018; Amari, 2016). Optimal experimental design uses trace, determinant, and minimum-eigenvalue criteria to select informative measurements (Pukelsheim, 2006). Gradient matching brings related objectives to data-efficient subset selection (Killamsetty et al., 2021). Our setting couples these ideas: each query first induces a student occupancy, and that occupancy determines the operator being designed. This endogenous data distribution distinguishes spectral query coverage from semantic diversity, prefix count, and static coreset selection.

Endogenous training distributions. Interactive imitation learning and decision-dependent optimization also study learners whose actions determine later observations (Ross et al., 2011; Rajaraman et al., 2021; Drusvyatskiy & Xiao, 2023; Li & Wai, 2022). Their central object is usually distribution shift or equilibrium under feedback. Our object is the local spectrum created by that feedback: student-generated prefixes change the Fisher-weighted tangent directions on which teacher discrepancy can be removed. The changing-operator bound then connects the frozen local geometry to the evolving on-policy distribution. This directional account resolves which modes each occupancy activates, how residual mass migrates among those modes, and how the resulting geometry determines finite-horizon query value. It also makes training horizon an explicit property of data value.

3 SPECTRAL THEORY OF OPD ABSORPTION

3.1 A QUERY INDUCES AN OPTIMIZATION OPERATOR

Let $\mathcal{Q}$ be a finite query pool and ρ a distribution over it. Sampling $q \sim \rho$ and rolling out the current student induces a prefix occupancy $d_{\theta,\rho}$. During one stop-gradient OPD update this occupancy is fixed while an optimization copy ϑ is differentiated:

$$\mathcal{L}_\rho(\vartheta;\theta) = \mathbb{E}_{s\sim d_{\theta,\rho}}\left[D_f(\pi_{\mathrm{T}}(\cdot \mid s), \pi_\vartheta(\cdot \mid s))\right]. \tag{1}$$

Let $z_\vartheta(s)$ be centered logits, $J_s = \nabla_\vartheta z_\vartheta(s)$ their Jacobian in the trainable coordinates, and

$$C(p) = \mathrm{Diag}(p) - pp^\top \tag{2}$$

the categorical Fisher curvature. The coordinates may be all model weights or an adapter subspace; our neural experiments use the flattened LoRA parameters actually updated by the optimizer.

Theorem 3.1 (Local OPD geometry)**.** *Assume that f is twice differentiable at one with $f''(1) > 0$, occupied teacher probabilities are positive, and $\theta_\star$ matches the teacher on those prefixes. The Hessian of Equation 1 at $\vartheta = \theta_\star$ is*

$$H_\rho = f''(1)\,\mathbb{E}_{s\sim d_{\theta_\star,\rho}}\left[J_s^\top C(\pi_{\mathrm{T}}(\cdot \mid s))\, J_s\right]. \tag{3}$$

Thus every smooth f divergence has the same local eigendirections at agreement up to the scalar $f''(1)$, while changing the query distribution changes the occupancy-weighted operator itself.

For teacher cross entropy away from agreement, the exact Hessian at one prefix decomposes as

$$\nabla^2_\vartheta \ell_s = J_s^\top C(\pi_\vartheta(\cdot \mid s))\, J_s + \sum_a \left(\pi_{\vartheta,a}(s) - \pi_{\mathrm{T},a}(s)\right) \nabla^2_\vartheta z_{\vartheta,a}(s). \tag{4}$$

The first term is the current student-Fisher Gauss Newton operator. The second vanishes at agreement and records nonlinear network curvature. This distinction motivates both teacher-Fisher designs, which estimate the equality-limit geometry, and student-Fisher designs, which estimate the current Gauss Newton term.

The linear operator has a quantitative local error scale. If $\nabla^2\mathcal{L}_\rho$ is M-Lipschitz on the segment from an agreement point $\theta_\star$ to $\theta_\star + \delta$, Taylor's theorem gives

$$\left\|\nabla\mathcal{L}_\rho(\theta_\star + \delta;\theta) - H_\rho\delta\right\|_2 \le \frac{M}{2}\|\delta\|_2^2. \tag{5}$$

If the weakest active linear update has scale $\mu\|\delta\|_2$, the relative error is at most $M\|\delta\|_2/(2\mu)$. Weak modes are therefore also the first directions in which nonlinear curvature can compete with the linear term. This produces a finite validity window governed by the local approximation scale.

Whiten and stack the teacher-student logit residuals over occupied prefixes into u. If J stacks the corresponding Jacobians and W_ρ contains occupancy and token curvature, the gap-space operator is

$$A_\rho = W_\rho^{1/2} J J^\top W_\rho^{1/2} \succeq 0. \tag{6}$$

Its nonzero eigenvalues equal those of $H_\rho = J^\top W_\rho J$. The gap-space form evolves the teacher discrepancy directly; the pullback form is estimated in the trainable parameter coordinates. Complete definitions of occupancy, teacher gap, empirical absorption, and the two operator forms are given in Appendix A.

3.2 SPECTRAL STARVATION

Under a fixed linearized operator, one gradient step gives

$$u_{t+1} = (I - \eta A)u_t, \qquad a_t = 1 - \frac{\|u_{t+1}\|_2^2}{\|u_t\|_2^2}, \tag{7}$$

where a_t is the fraction of squared teacher gap absorbed by update t.

Theorem 3.2 (Spectral starvation). *Let A have eigenpairs (λ_i, v_i), write $u_0 = \sum_i c_i v_i$, and assume $0 < \eta < 1/\lambda_{\max}(A)$. Then*

$$\|u_t\|_2^2 = \sum_i c_i^2 (1 - \eta\lambda_i)^{2t}. \tag{8}$$

The absorption sequence a_t is nonincreasing and decreases strictly whenever the active residual occupies two distinct contraction factors. Let $r_i = (1 - \eta\lambda_i)^2$ and let ν_t assign mass proportional to $c_i^2 r_i^t$ to r_i. Its one-step decline is exactly

$$a_t - a_{t+1} = \frac{\mathrm{Var}_{\nu_t}(r)}{\mathbb{E}_{\nu_t}[r]}. \tag{9}$$

If μ is the smallest positive eigenvalue carrying residual, then

$$\lim_{t\to\infty} a_t = 1 - (1 - \eta\mu)^2, \qquad \lim_{t\to\infty} \|u_t\|_2^2 = \|P_{\ker A} u_0\|_2^2. \tag{10}$$

The mechanism is a change in residual composition. Large-eigenvalue components disappear first, so the normalized residual measure shifts toward weak modes. Log convexity of the residual energy makes fractional absorption decline even though every active mode contracts. In continuous time, $u(t) = \exp(-tA)u_0$ yields the sharper identity

$$-\frac{\mathrm{d}}{\mathrm{d}t} \log\|u(t)\|_2^2 = 2\mathbb{E}_{\nu_t}[\lambda], \qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2} \log\|u(t)\|_2^2 = 4\,\mathrm{Var}_{\nu_t}(\lambda). \tag{11}$$

Equivalently, the derivative of the log-absorption rate is $-4\,\mathrm{Var}_{\nu_t}(\lambda)$. Spectral heterogeneity therefore controls how quickly the rate deteriorates, while the kernel and the smallest active eigenvalue determine the residual floor and late contraction.

The smallest active eigenvalue also controls a finite-time envelope. Let P_0 project onto $\ker A$, let μ be the smallest positive eigenvalue carrying residual, and let P_μ project onto its eigenspace. Then

$$\|P_\mu u_0\|_2^2 (1 - \eta\mu)^{2t} \le \|u_t\|_2^2 - \|P_0 u_0\|_2^2 \le \|(I - P_0)u_0\|_2^2 (1 - \eta\mu)^{2t}. \tag{12}$$

For a target removable residual ϵ, the upper bound gives the sufficient update count

$$t \ge \frac{\log\big(\|(I - P_0)u_0\|_2^2/\epsilon\big)}{-2\log(1 - \eta\mu)}. \tag{13}$$

This separates how much residual is removable from how poorly its weakest direction is conditioned. More queries help only when they reduce the kernel component, increase μ, or reweight the realized residual toward better covered modes. Changing one scalar learning rate cannot remove the underlying condition number. With active spectrum in $[\mu, L]$, the best constant scalar step has worst-case contraction $(L - \mu)/(L + \mu)$, so spectral design changes a bottleneck that step-size tuning can only balance.

3.3 Queries Are Horizon-Dependent Spectral Designs

Let A_q be the average operator induced by query q. A normalized design w produces

$$A(w) = \sum_{q=1}^{n} w_q A_q, \qquad w \in \Delta_n. \tag{14}$$

Query count records support size, while tangent coverage depends on the ranges and eigenvalues of the selected operators. Since

$$\ker A(w) = \bigcap_{q : w_q > 0} \ker A_q, \tag{15}$$

a query improves span only when it acts on a direction left in the current kernel. Replicating an existing operator changes record count but leaves the normalized operator, spectrum, and linearized residual curve unchanged.

For an equal-cardinality subset S, four natural criteria expose different notions of value:

$$\begin{aligned} V_{\mathrm{trace}}(S) &= \mathrm{tr}(A_S), & V_{\mathrm{D}}(S) &= \log\det(A_S + \alpha I), \\ V_{\mathrm{E}}(S) &= \lambda_{\min}(A_{S|U}), & V_{\mathrm{short}}(S; u_0) &= u_0^\top A_S u_0. \end{aligned} \tag{16}$$

Trace measures total curvature, the regularized determinant measures covered volume, E optimality protects the weakest direction, and residual energy measures immediate progress on the present teacher gap. Equal trace, support size, or prefix count need not equalize the other criteria. Direction-normalized gradient factors emphasize coverage; residual-weighted factors retain both amplitude and interaction among selected directions.

For a teacher-relevant subspace U, E optimality maximizes $\lambda_{\min}(A(w)|_U)$ and therefore minimizes the worst-case residual

$$\sup_{u_0 \in U \setminus \{0\}} \frac{\|\exp(-tA(w))u_0\|_2^2}{\|u_0\|_2^2} = \exp[-2t\lambda_{\min}(A(w)|_U)]\,. \tag{17}$$

The realized teacher gap instead defines the horizon-specific objective

$$\Phi_\tau(w; u_0) = u_0^\top \exp(-2\tau A(w))u_0. \tag{18}$$

Theorem 3.3 (Horizon-dependent query value)**.** *For every fixed design w,*

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\Phi_\tau(w; u_0)\bigg|_{\tau=0} = -2u_0^\top A(w)u_0, \tag{19}$$

whereas the long-horizon exponent of the removable residual is the smallest eigenvalue of $A(w)$ carrying nonzero teacher gap. Hence short-horizon design maximizes realized gradient energy, while long-horizon design protects weak active directions. The two objectives can reverse the ranking of equal-budget query sets.

The reversal already occurs in two dimensions. With $u_0 = (\sqrt{0.9}, \sqrt{0.1})^\top$, $A_1 = \mathrm{Diag}(1.9, 0.1)$, and $A_2 = I$, both designs have trace two, but $u_0^\top A_1 u_0 = 1.72 > 1$ and $\lambda_{\min}(A_1) = 0.1 < 1 = \lambda_{\min}(A_2)$. The first design wins initially; the second wins at a sufficiently long horizon. Gradient norm, direction-normalized E design, and residual-weighted E design operationalize immediate energy, robust coverage, and their combination.

3.4 Changing Occupancy Defines the Validity Window

Fully on-policy training replaces the fixed operator by A_t because updated students generate new prefixes and new Jacobians. Let

$$u_{t+1} = (I - \eta A_t)u_t + r_t, \tag{20}$$

where r_t collects nonlinear and approximation error.

Theorem 3.4 (Drifting operator bound)**.** *Suppose $A_t \succeq \mu I$ and $A_t \preceq LI$ on a teacher-relevant invariant subspace, $\|r_t\|_2 \le \varepsilon_t$, and $0 < \eta \le 1/L$. Then*

$$\|u_T\|_2 \le (1 - \eta\mu)^T \|u_0\|_2 + \sum_{t=0}^{T-1} (1 - \eta\mu)^{T-1-t} \varepsilon_t. \tag{21}$$

Relative to a frozen reference A_0, the accumulated deviation is bounded by the sum of $\eta\|A_t - A_0\|_{op}\|u_t\|_2$ and $\|r_t\|_2$, with subsequent frozen contractions applied to earlier perturbations.

When the operators commute and $r_t = 0$, changing occupancy remains exactly modewise. If $\lambda_{i,t}$ denotes the eigenvalue of A_t in their common basis, then

$$\|u_T\|_2^2 = \sum_i c_i^2 \prod_{t=0}^{T-1} (1 - \eta\lambda_{i,t})^2. \tag{22}$$

This bound converts a local approximation into a measurable prediction. Freeze query gradients and student-Fisher operators at the common checkpoint, replay the realized schedule, and predict the evaluation gap with the resulting quadratic recurrence. The forecast should be accurate while A_t remains close to A_0 and should degrade as tangent or occupancy replacement grows. Fixed-prefix training suppresses occupancy drift; matched regeneration isolates its additional effect.

Table 1: Main empirical tests of the spectral account. Intervals resample independent training seeds.

Prediction	**Measurement**	**Result**
Exact modal dynamics	Deterministic and randomized operators	Maximum variance-identity error 9.42×10^{-16}
Frozen neural forecast	160 runs through update 16	$R^2 = 0.906$, rank correlation 0.925
Forecast validity window	64 matched tangent replays	Drift-error rank correlation 0.504 $[0.392, 0.623]$
Fixed-operator starvation	160 fixed-prefix $q = 4$ runs	158 negative temporal associations, mean -0.649
Cross-domain starvation	128 ARC-Challenge runs	All eight selector and mode intervals have positive lower endpoints
Selected-set efficiency	Four Residual E versus sixteen random queries	Gap advantage 0.078 $[0.009, 0.152]$
Operator replacement	Current-occupancy student Fisher at endpoint	Relative drift 0.984 to 0.989, angle approximately 89 degrees

Concretely, let g_i and H_i be the gradient and student-Fisher operator of scheduled query i at the common checkpoint. The frozen parameter displacement and evaluation-gap forecast are

$$\widehat{\delta}_{t+1} = \widehat{\delta}_t - \eta(g_{i_t} + H_{i_t}\widehat{\delta}_t), \qquad \widehat{D}_t = D_0 + g_{\mathrm{eval}}^\top \widehat{\delta}_t + \frac{1}{2}\widehat{\delta}_t^\top H_{\mathrm{eval}}\widehat{\delta}_t. \tag{23}$$

This recurrence is exact for a fixed quadratic objective. In a neural model it simultaneously defines a quantitative short-horizon prediction and a validity test: prediction error can be paired with direct measurements of $\|A_t - A_0\|$ along the same trajectory.

The complete theorem family also treats nonlinear remainders, finite-time envelopes, stable overshoot, coverage by kernel intersection, exact E optimality, task decision margins, commuting operator paths, and deviation from a frozen reference. For these statements, please refer to Appendix B; for every proof, please refer to Appendix G.

4 Experiments and Results

4.1 Experimental Design

Table 1 summarizes three tests: heterogeneous spectra should produce declining fractional absorption, the frozen operator A_0 should predict the short neural future, and its prediction should weaken as the realized operator drifts. A deterministic finite system first verifies the modal formulas, variance identity, kernel floor, and equal-trace horizon reversal. The neural study then uses official Qwen3.5 0.8B and 4B checkpoints as student and teacher, with rank-eight LoRA training on GSM8K (Cobbe et al., 2021). A scale panel uses a 4B student and 9B teacher, and an ARC-Challenge panel tests the temporal mechanism on science reasoning (Clark et al., 2018; Qwen Team, 2026).

Each of sixteen seeds supplies a common warm-start adapter, 32 candidate queries, and disjoint evaluation prompts. Ten selectors span random choice, semantic diversity, gradient energy, log determinant, direction-normalized E design, residual-weighted E design, and teacher- or student-Fisher variants. Query counts $q \in \{1, 4, 16\}$ separate selected-set size from geometry. Fixed-prefix SGD reuses the initial student trajectories and isolates the frozen operator. Fully on-policy AdamW regenerates a trajectory before every update. A matched AdamW panel holds the initial adapter, optimizer, schedule, query subset, clipping, and update count fixed while changing only prefix regeneration. The complete study contains 1,088 training runs. For model revisions, prompts, sketches, subset solvers, schedules, and statistical procedures, please refer to Appendix C.

4.2 Empirical Findings

The observations below unpack the empirical tests summarized in Table 1.

Observation 1: Spectral starvation is governed by conditioning rather than query count or rank, and the predicted slowdown appears in neural OPD. In the exact equal-budget certificate,

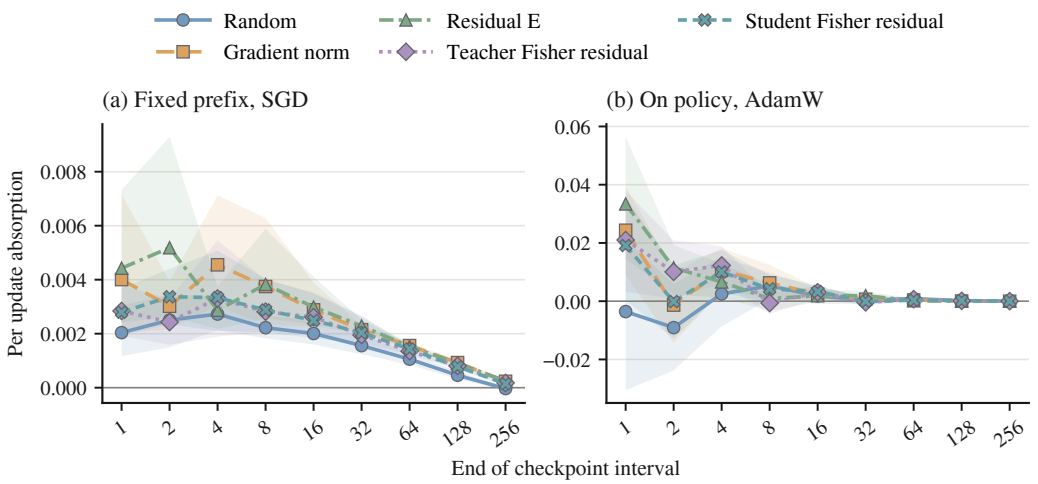


Fig. 2. Per-update teacher-gap absorption in the $q = 4$ neural matrix. Lines are seed means and bands are paired bootstrap intervals. Fixed-prefix SGD exposes progressive depletion of strong modes. Fully on-policy AdamW adds occupancy drift while retaining the overall decline toward weak absorption.

redundant, uniform, and spectrum-balanced operators all have full algebraic rank but active condition numbers 1.30×10^5, 3.10×10^4, and 1.08, respectively. After 200 updates, their squared residuals are 0.943, 0.450, and 3.85×10^{-14}. Redundant and uniform designs rapidly expose a weak tail, whereas the balanced design maintains nearly constant absorption. Randomized certificates recover the discrete variance identity to machine precision and locate the equal-trace horizon reversal at time 1.221. The same temporal pattern appears in neural OPD. Among 160 fixed-prefix $q = 4$ runs, 158 have negative absorption-time association, with mean -0.649; under fully on-policy AdamW, 128 of 160 remain negative despite regenerated trajectories. The fixed-prefix condition exposes the modal depletion predicted by Theorem 3.2, while regeneration introduces a changing operator. Fig. 2 shows this temporal signature. Complete finite certificates are reported in Appendix D.

Observation 2: The initial spectrum quantitatively predicts the short neural future, while tangent drift marks the boundary of that prediction. Across 160 unclipped runs and 800 seed, selector, and checkpoint predictions through update 16, the frozen student-Fisher recurrence attains $R^2 = 0.906$, rank correlation 0.925, and RMSE 0.0039 for design-pool teacher-gap reduction. Matched replay identifies where this forecast degrades. Mean relative tangent drift grows from 0.344 after one update to 0.505 after sixteen and 0.649 after 64. Across 448 replay records, tangent drift and forecast error have rank correlation 0.504, with seed-cluster interval $[0.392, 0.623]$. The replayed gap remains within 4.9×10^{-4} of the archived trajectory, separating forecast breakdown from replay error. Fig. 3 summarizes the forecast and its validity window, consistent with Theorem 3.4.

Observation 3: Query value is horizon dependent, and different selectors optimize distinct properties of the same local operator. Consistent with Theorem 3.3, Gradient norm and Residual E improve maximum-gap reduction over random selection by 0.200 and 0.205 under fixed prefixes, with paired intervals $[0.127, 0.276]$ and $[0.133, 0.280]$. Under changing occupancy, the corresponding improvements are 0.108 and 0.110, with intervals $[0.041, 0.180]$ and $[0.055, 0.168]$. Pure Fisher E instead raises the realized smallest eigenvalue, matching its weakest-direction objective. This reflects the distinction between immediate residual energy and persistent spectral coverage.

Observation 4: Query number and query geometry interact; a small geometrically selected set can outperform a substantially larger random set. Increasing the selected set from one to four queries especially benefits Gradient norm and Residual E by combining complementary high-energy directions. At $q = 4$, on-policy Residual E exceeds the $q = 16$ random endpoint by 0.078, with paired interval $[0.009, 0.152]$. This shows that a smaller set can outperform a larger one when its trajectories provide better coverage of the remaining teacher gap. Complete query-count curves and selector comparisons are reported in Appendix D.

Observation 5: The selected-set advantage holds at matched online teacher cost, while geometry-aware selection incurs an explicit screening cost. The $q = 4$ Residual E and $q = 16$ random

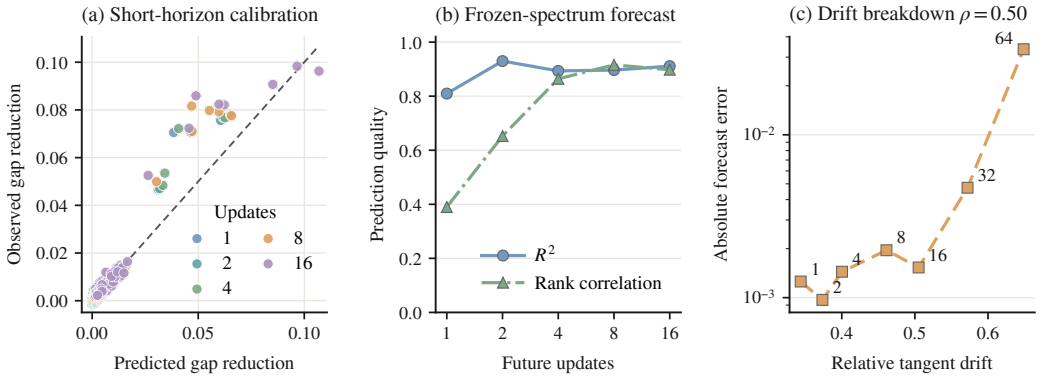


Fig. 3. Local spectral prediction and its measured validity window. Panel (a) compares the frozen A_0 forecast with observed design-pool gap reduction. Panel (b) reports horizon-specific calibration and ordering. Panel (c) follows matched tangent replay through update 64. Growing student-Fisher drift tracks growing absolute forecast error with pooled rank correlation 0.504.

schedules both use 256 online teacher trajectories. Residual E additionally requires a 32-candidate teacher-supervised screen, giving minimal method-specific costs of 288 versus 256 trajectories. The comparison therefore establishes selected-set and online-update efficiency, not four total teacher calls versus sixteen. Token and compute accounting are reported in Appendix F.

Observation 6: The starvation signature and horizon-aware query gains transfer across model scale and reasoning domain. With a 4B student and 9B teacher, Gradient norm improves maximum-gap reduction over random by 0.038 under fixed prefixes and 0.043 on-policy, with positive lower interval endpoints. On ARC-Challenge, starvation strength is positive for Random, Gradient norm, Gradient E, and Residual E under both fixed and regenerated prefixes; all eight seed-bootstrap intervals have positive lower endpoints. Complete scale-transfer and ARC results are reported in Appendix D.

Observation 7: Fully on-policy training rapidly replaces the frozen operator, making the spectral account predictive locally rather than globally. Matched AdamW regeneration changes 97.9% to 99.8% of repeated trajectories and makes absorption-time association less negative relative to prefix reuse. Current-occupancy student-Fisher drift reaches 0.976 by update 64 and 0.984–0.989 at update 256. At the archived endpoint, fixed-prefix tangent drift is 0.950–0.965, the isolated occupancy increment is of order one, and maximum principal angles are approximately $89°$. These measurements define the scope of the frozen-spectrum model: the initial operator explains the early gap trajectory, whereas parameter evolution and changing occupancy progressively replace its tangent geometry. Fully on-policy OPD is therefore naturally described by an evolving operator path rather than one global spectrum. Broader consequences are discussed in Appendix E.

5 Conclusion

On-policy distillation turns every query into a local spectral design over student-generated trajectories. Its occupancy-weighted Fisher operator yields spectral starvation: strong components disappear first, fractional absorption declines by an exact variance identity, weak directions determine the limiting rate, and kernel components form a residual floor. The same geometry separates immediate gradient energy from persistent coverage and yields equal-budget horizon reversals. Across 1,088 Qwen3.5 runs, the initial student-Fisher operator predicts sixteen updates with $R^2 = 0.906$ and rank correlation 0.925; matched replay links forecast error to tangent drift, and direct measurements expose order-one occupancy replacement. The signature transfers across tasks, occupancy rules, clipping regimes, and model scale. Four Residual E queries absorb more teacher gap than sixteen random queries at matched online teacher cost. Together, these results establish a predictive theory of OPD dynamics and query design. The operator path connects query composition, training horizon, and finite-budget distillation progress under the endogenous training distribution created by the evolving student. It is the natural unit of analysis for OPD.

AI USE STATEMENT

We use generative AI tools in supporting roles in certain aspects of research execution and manuscript preparation, including language editing, manuscript restructuring, as well as code generation and debugging. All mathematical formulations and equations are developed and drafted manually by the authors, and all associated analyses and scientific interpretations are likewise conducted manually by the authors. GPT-5.6 Sol is used only as a secondary check of the mathematical expressions for correctness and internal consistency; the authors independently review these checks and determine the final formulations. Generative AI tools are not used to generate experimental data or to make final scientific decisions. All AI-assisted code is reviewed, tested, and validated against the intended computations and experimental outputs. Similarly, all AI-assisted manuscript text is reviewed and revised by the authors, and all reported numerical results are checked against the underlying experimental outputs. The authors take full responsibility for the final content of this work, including its text, claims, mathematical formulations, code, analyses, and any other artifacts produced with the assistance of generative AI.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, 2024.

Shun-ichi Amari. *Information Geometry and Its Applications*. Springer, 2016.

Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. In *Automata, Languages and Programming*, volume 2380 of *Lecture Notes in Computer Science*, pp. 693–703. Springer, 2002. doi: 10.1007/3-540-45465-9_59.

Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. Think you have solved question answering? try ARC, the AI2 reasoning challenge. *arXiv preprint arXiv:1803.05457*, 2018. doi: 10.48550/arXiv.1803.05457.

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

Wojciech M. Czarnecki, Razvan Pascanu, Simon Osindero, Siddhant Jayakumar, Grzegorz Swirszcz, and Max Jaderberg. Distilling policy distillation. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1331–1340. PMLR, 2019.

Dmitriy Drusvyatskiy and Lin Xiao. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research*, 48(2):954–998, 2023. doi: 10.1287/moor.2022.1287.

Zixuan Fu, Bingxiang He, Yuxin Zuo, Haohuan Huang, Jinqian Zhang, Ruhang Xiao, Cheng Qian, Qinyu Luo, Huan-ang Gao, Yudong Wang, Zhiyuan Liu, Ning Ding, and Chaojun Xiao. Rethinking on-policy distillation of large language models II: One training example. *arXiv preprint arXiv:2609.04172*, 2026.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. MiniLLM: Knowledge distillation of large language models. In *International Conference on Learning Representations*, 2024.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems*, 2018.

Krishnateja Killamsetty, Sivasubramanian Durga, Ganesh Ramakrishnan, Abir De, and Rishabh Iyer. Grad-Match: Gradient matching based data subset selection for efficient deep model training. In *International Conference on Machine Learning*, pp. 5464–5474, 2021.

Yoon Kim and Alexander M. Rush. Sequence-level knowledge distillation. *arXiv preprint arXiv:1606.07947*, 2016.

Qiang Li and Hoi-To Wai. State dependent performative prediction with stochastic approximation. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 3164–3186. PMLR, 2022.

Friedrich Pukelsheim. *Optimal Design of Experiments*. SIAM, 2006.

Qwen Team. Qwen3.5 model collection. Hugging Face model repository, 2026.

Nived Rajaraman, Yanjun Han, Lin Yang, Jingbo Liu, Jiantao Jiao, and Kannan Ramchandran. On the value of interaction and function approximation in imitation learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 1325–1336, 2021.

Stephane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pp. 627–635. PMLR, 2011.

Andrei A. Rusu, Sergio Gomez Colmenarejo, Caglar Gulcehre, Guillaume Desjardins, James Kirkpatrick, Razvan Pascanu, Volodymyr Mnih, Koray Kavukcuoglu, and Raia Hadsell. Policy distillation. *arXiv preprint arXiv:1511.06295*, 2015.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Dongxu Zhang, Zhichao Yang, Sepehr Janghorbani, Jun Han, Andrew Ressler, Qian Qian, Gregory D. Lyng, Sanjit Singh Batra, and Robert E. Tillman. Fast and effective on-policy distillation from reasoning prefixes. *arXiv preprint arXiv:2602.15260*, 2026.

Qingfei Zhao, Huan Song, Shuyu Tian, Jiawei Shao, and Xuelong Li. Prefix-guided on-policy distillation: Mining golden trajectories from rollouts. *arXiv preprint arXiv:2606.21994*, 2026.

A Problem Setup and Measured Objects

A.1 Queries, Rollouts, and Occupancy

Let $\mathcal{Q} = \{q_1, \ldots, q_n\}$ be a finite candidate pool of teacher queries and let $\rho \in \Delta_n$ be a distribution over that pool. A query can be a prompt, a problem statement, or an initial context. Given student parameters θ, sampling a query and rolling out the student induces a distribution over token prefixes. We write this occupancy measure as

$$d_{\theta,\rho}(s) = \sum_{q \in \mathcal{Q}} \rho(q) \sum_{j \geq 0} \Pr_{\pi_\theta}(S_j = s \mid q)\, \omega_j, \tag{24}$$

where S_j is the prefix at position j and the nonnegative weights ω_j specify the token positions used by the training rule. Normalizing the ω_j makes $d_{\theta,\rho}$ a probability distribution; leaving them unnormalized changes only the global scale of the local operator. The occupancy contains both prompt selection and student behavior. Two students presented with the same query distribution can therefore induce different training distributions, and the same student can induce different occupancies under different query weights.

The teacher provides a categorical distribution $\pi_{\mathrm{T}}(\cdot \mid s)$ at an occupied prefix. The student distribution is $\pi_\theta(\cdot \mid s)$. During a standard OPD update, prefixes are sampled from the current student and treated

as constants while the student parameters are differentiated. It is useful to show the two roles explicitly. We introduce an optimization copy ϑ and define the frozen-occupancy loss

$$\mathcal{L}_\rho(\vartheta;\theta) = \mathbb{E}_{s\sim d_{\theta,\rho}}\left[D_f(\pi_{\mathrm{T}}(\cdot \mid s), \pi_\vartheta(\cdot \mid s))\right]. \tag{25}$$

The sampling argument θ determines which prefixes are observed. The differentiated argument ϑ determines the update at those prefixes. Fixed-prefix experiments keep the sampled states and teacher targets unchanged for many updates. Fully on-policy experiments reset $\theta = \vartheta$ after every update and regenerate the next trajectory.

A.2 Teacher Gap and Absorption

For a prefix s, define the token-level teacher gap

$$g_\theta(s) = D_f(\pi_{\mathrm{T}}(\cdot \mid s), \pi_\theta(\cdot \mid s)). \tag{26}$$

For a declared evaluation distribution μ, its mean gap is

$$G_\mu(\theta) = \mathbb{E}_{s\sim\mu}[g_\theta(s)]. \tag{27}$$

The distribution μ need not equal the current training occupancy. In the experiments, the design-pool metric evaluates all candidate-query trajectories generated at the common warm start, while a disjoint heldout pool measures transfer beyond the selected candidates. We also report the maximum per-query gap over the design pool. Mean gap describes aggregate absorption, maximum gap retains the hardest candidate, and heldout gap measures whether the update generalizes beyond the selected query set.

For two checkpoints $t_0 < t_1$, define geometric per-update absorption by

$$\widehat{a}_{t_0:t_1} = 1 - \left(\frac{G_\mu(\theta_{t_1})}{G_\mu(\theta_{t_0})}\right)^{1/(t_1-t_0)}. \tag{28}$$

This quantity is the constant fractional reduction per update that would produce the observed endpoint ratio. It can be compared across checkpoint intervals of different widths. A value of zero means that the measured gap is unchanged over the interval, a positive value means contraction, and a negative value means expansion. The primary temporal statistic is the Spearman association between the interval endpoint and $\widehat{a}_{t_0:t_1}$. The exact theory predicts a nonincreasing absorption sequence under one fixed stable operator. Neural training replaces exact monotonicity with a rank-based prediction that is robust to scale differences between intervals.

The finite-horizon reduction used to compare query selectors is

$$\mathcal{D}_T^{\max}(\rho) = G_{\max}(\theta_0) - G_{\max}(\theta_T), \qquad \mathcal{D}_T^{\mathrm{held}}(\rho) = G_{\mathrm{held}}(\theta_0) - G_{\mathrm{held}}(\theta_T). \tag{29}$$

All selector contrasts hold the query count, update count, model pair, common adapter, and random seed fixed. The query design changes only which candidates are scheduled. A positive paired difference against random selection therefore measures additional gap absorbed at an equal prompt and update budget.

A.3 Local Coordinates and Gap Directions

Let $z_\theta(s) \in \mathbb{R}^{V-1}$ be centered logits at prefix s, with one redundant softmax direction removed, and let

$$J_s = \nabla_\theta z_\theta(s) \in \mathbb{R}^{(V-1)\times p} \tag{30}$$

be the Jacobian with respect to the p trainable coordinates. In a full-parameter analysis these coordinates are all model parameters. In the neural experiments they are rank-eight LoRA parameters, so the spectrum describes the optimization subspace actually available to the student. Define the categorical Fisher curvature at a distribution p_s by

$$C(p_s) = \mathrm{Diag}(p_s) - p_s p_s^\top. \tag{31}$$

For teacher curvature, $p_s = \pi_{\mathrm{T}}(\cdot \mid s)$. For student curvature, $p_s = \pi_\theta(\cdot \mid s)$. Both are positive semidefinite on the centered-logit subspace.

Table 2: Core objects in the spectral analysis of OPD. Every operator and metric is indexed by a declared query design, occupancy, trainable coordinate system, and evaluation distribution.

Object	Type	Meaning
ρ	query distribution	Weights assigned to the finite candidate pool
$d_{\theta,\rho}$	prefix occupancy	Student-generated states and token-position weights induced by ρ
$G_\mu(\theta)$	scalar	Mean teacher gap on declared evaluation distribution μ
$\widehat{a}_{t_0:t_1}$	scalar	Geometric fraction of measured gap absorbed per update
J_s	matrix	Centered-logit Jacobian in the trainable coordinates
$C(p_s)$	matrix	Categorical Fisher curvature at teacher or student probabilities
A_ρ	operator	Gap-space occupancy-weighted OPD absorption operator
H_ρ	operator	Parameter-space pullback sharing the nonzero spectrum of A_ρ
u	vector	Whitened teacher-student discrepancy on occupied prefixes
U	subspace	Declared teacher-relevant tangent subspace for robust design
$\Phi_\tau(\rho; u)$	scalar	Residual squared gap after local optimization horizon τ

The local discrepancy direction is the centered teacher-student logit residual. After whitening by categorical curvature and stacking occupied prefixes, it becomes a vector u. The query-induced operator acting on this vector is

$$A_\rho = W_\rho^{1/2} J J^\top W_\rho^{1/2}, \tag{32}$$

where J stacks the occupied Jacobians and W_ρ contains occupancy and token curvature. Its nonzero eigenvalues equal those of the parameter-space pullback

$$H_\rho = J^\top W_\rho J. \tag{33}$$

The gap-space form evolves the teacher discrepancy directly. The parameter-space form is easier to estimate with Jacobian-vector products and sketches. We move between the two forms only through their shared nonzero spectrum.

A.4 Four Notions of Coverage

The analysis distinguishes four properties of a query set. *Query coverage* is the number of selected prompts. *State coverage* describes the number or diversity of generated prefixes. *Tangent coverage* describes the span and conditioning of their curvature-weighted Jacobians. *Residual coverage* additionally asks whether those directions align with the teacher gap currently present. The first two are properties of visited inputs. The latter two are properties of how those inputs act on the optimization problem.

This distinction explains why query count alone can be misleading. Repeating one high-gradient direction can produce many tokens and a large trace while leaving an orthogonal teacher-gap component unchanged. Conversely, a small set of trajectories can cover the relevant trainable subspace if their Jacobian factors are complementary. The kernel of A_ρ records the completely unexcited component, its smallest positive eigenvalue records the weakest active direction, and $u^\top A_\rho u$ records the energy placed on the realized residual.

Table 2 makes the conditioning explicit. A spectrum is meaningful only after the query pool, occupancy rule, trainable coordinates, curvature convention, and evaluation gap have been declared. The empirical protocol fixes these objects within every paired comparison. The next section derives the operator from the distillation objective and characterizes how its active spectral measure evolves.

B Spectral Theory of OPD Absorption

The analysis proceeds in four layers. First, we derive the local operator from the frozen-occupancy distillation objective and identify the categorical Fisher geometry shared by smooth token divergences. Second, we solve the induced linear gap dynamics and characterize the changing spectral measure responsible for decreasing absorption. Third, we distinguish span, conditioning, and residual alignment in equal-budget query design. Fourth, we allow the student occupancy to change and bound the resulting departure from the frozen operator. Each layer produces a separately testable quantity in the experiments.

The local analysis is performed in the trainable coordinates declared by the algorithm. It therefore applies to full finetuning, adapters, or any smooth parameterization after replacing J_s with the corresponding Jacobian. Reparameterization changes the Euclidean operator unless it is orthogonal, just as it changes ordinary gradient descent. The categorical curvature and occupied prefixes remain intrinsic to the predictive problem, while the pullback records the update geometry actually used by the optimizer. This distinction matters empirically because our student is trained in a LoRA subspace rather than in every weight coordinate.

B.1 Frozen Occupancy Objective

Let ρ be a distribution over training queries. A student rollout induces a distribution $d_{\theta,\rho}$ over prefixes s. During one stop-gradient OPD update, this distribution is fixed while an optimization copy ϑ is differentiated. For a token divergence D_f, define

$$\mathcal{L}_\rho(\vartheta;\theta) = \mathbb{E}_{s\sim d_{\theta,\rho}}\left[D_f(\pi_{\mathrm{T}}(\cdot \mid s), \pi_\vartheta(\cdot \mid s))\right]. \tag{34}$$

The two arguments separate the sampling policy from the differentiated policy. Let $z_\vartheta(s)$ denote centered student logits and $J_s = \nabla_\vartheta z_\vartheta(s)$. The categorical Fisher matrix at the teacher is

$$C_s = \mathrm{Diag}(\pi_{\mathrm{T}}(\cdot \mid s)) - \pi_{\mathrm{T}}(\cdot \mid s)\pi_{\mathrm{T}}(\cdot \mid s)^\top. \tag{35}$$

Theorem B.1 (Local OPD geometry)**.** *Assume f is twice differentiable at one with $f''(1) > 0$, the occupied teacher probabilities are positive, and $\theta_\star$ matches the teacher on the occupied prefixes. Then the Hessian of Equation 34 at $\vartheta = \theta_\star$ is*

$$H_\rho = f''(1)\,\mathbb{E}_{s\sim d_{\theta_\star,\rho}}\left[J_s^\top C_s J_s\right]. \tag{36}$$

Consequently, smooth forward KL, reverse KL, and other smooth f divergences share the same local eigendirections up to their scalar curvature at equality.

The proof follows by taking the second order expansion of an f divergence at equality and applying the chain rule. Equation 36 identifies the two sources of conditioning: the student tangent map and the prefix occupancy induced by the query distribution.

For the teacher cross entropy used in our experiments, the geometry away from equality has an exact decomposition. Write $p_\vartheta(s) = \pi_\vartheta(\cdot \mid s)$ and $p_{\mathrm{T}}(s) = \pi_{\mathrm{T}}(\cdot \mid s)$.

Proposition B.2 (Curvature away from agreement)**.** *For a fixed occupied prefix s, the parameter Hessian of $-\sum_a p_{\mathrm{T},a}(s)\log p_{\vartheta,a}(s)$ is*

$$\nabla^2_\vartheta \ell_s = J_s^\top C_{p_\vartheta(s)} J_s + \sum_a \left(p_{\vartheta,a}(s) - p_{\mathrm{T},a}(s)\right)\nabla^2_\vartheta z_{\vartheta,a}(s). \tag{37}$$

The first term is the current student Fisher Gauss Newton operator. The second term vanishes at teacher agreement and records the nonlinear curvature remainder.

Equation 37 motivates two measured designs. Teacher curvature estimates the equality-limit operator in Theorem B.1, while student curvature estimates the exact Gauss Newton term at the common warm start. Their difference tests whether the remaining teacher gap is small enough for the equality-limit spectrum to predict optimization.

Theorem B.1 has two immediate consequences. The first is divergence invariance of local directions. Forward KL, reverse KL, Jensen-Shannon divergence, and every other admissible f-divergence have the same occupied-prefix eigenspaces at agreement; only the scalar $f''(1)$ changes. The second is occupancy sensitivity. Reweighting queries changes the expectation in Equation 36, so the same model pair can have different local conditioning under different prompt mixtures. Changing a divergence and changing a query design therefore act on different parts of the local geometry.

Away from agreement, Proposition B.2 also separates two approximation errors. Replacing the exact Hessian by student Fisher removes the network-curvature remainder while retaining the current categorical probabilities. Replacing student Fisher by teacher Fisher additionally moves the categorical curvature to the equality limit. Our four Fisher selectors cross these two choices with direction normalization or residual weighting. Their empirical ordering reveals whether equality-limit coverage, current Gauss-Newton coverage, or realized residual amplitude dominates at the measured horizon.

Proposition B.3 (Local gradient remainder)**.** *Suppose $\nabla^2 \mathcal{L}_\rho(\cdot;\theta)$ is M-Lipschitz in operator norm on the segment from an agreement point $\theta_\star$ to $\theta_\star + \delta$. Then*

$$\|\nabla \mathcal{L}_\rho(\theta_\star + \delta;\theta) - H_\rho \delta\|_2 \le \frac{M}{2}\|\delta\|_2^2. \tag{38}$$

Thus the frozen linear operator controls the leading update, while the approximation error is second order in distance from teacher agreement.

The proposition gives a scale for local validity rather than a binary regime. If the smallest active linear update is of order $\mu\|\delta\|$, then the relative gradient error is at most $M\|\delta\|/(2\mu)$. Weak spectral directions are consequently the first directions for which nonlinear curvature can compete with the linear term. This observation connects spectral starvation and approximation quality: as the residual concentrates on small eigenvalues, the same absolute nonlinear remainder becomes more important. The unclipped panel tests the linear prediction before this late weak-mode regime dominates.

B.2 A Frozen Local Forecast

The local operator yields a quantitative prediction before the modal analysis is applied. For a realized query schedule $i_0, i_1, \ldots$, freeze each query gradient g_i and student-Fisher Gauss-Newton operator H_i at the common checkpoint. The predicted parameter displacement follows

$$\widehat{\delta}_{t+1} = \widehat{\delta}_t - \eta\left(g_{i_t} + H_{i_t}\widehat{\delta}_t\right), \qquad \widehat{\delta}_0 = 0. \tag{39}$$

For an evaluation distribution with checkpoint gradient g_{eval}, operator H_{eval}, and gap D_0, the corresponding forecast is

$$\widehat{D}_t = D_0 + g_{\mathrm{eval}}^\top \widehat{\delta}_t + \frac{1}{2}\widehat{\delta}_t^\top H_{\mathrm{eval}}\widehat{\delta}_t. \tag{40}$$

This recurrence is exact for a fixed quadratic objective. In a neural model, Proposition B.3 controls the local gradient error and Proposition B.20 later controls accumulated departure as the tangent operator changes. Equations 39 and 40 therefore define both a short-horizon prediction and a measurable validity window.

B.3 Why Absorption Slows

Stack the centered logit discrepancies over a finite occupied prefix collection and whiten them by the block categorical curvature. In the exact linearized model, the whitened gap u_t obeys

$$u_{t+1} = (I - \eta A)u_t, \qquad A = W^{1/2} J J^\top W^{1/2} \succeq 0. \tag{41}$$

The nonzero eigenvalues of A equal those of the parameter space pullback $J^\top W J$. The analysis uses A because it evolves the teacher gap directly, while the experiments sketch the equivalent pullback in the trainable LoRA coordinates.

Define the fraction of squared teacher gap absorbed by one update as

$$a_t = 1 - \frac{\|u_{t+1}\|_2^2}{\|u_t\|_2^2}. \tag{42}$$

Theorem B.4 (Spectral starvation)**.** *Let A have eigenpairs (λ_i, v_i) and write $u_0 = \sum_i c_i v_i$. If $0 < \eta < 1/\lambda_{\max}(A)$, then*

$$\|u_t\|_2^2 = \sum_i c_i^2 (1 - \eta\lambda_i)^{2t}. \tag{43}$$

The sequence a_t is nonincreasing. Its decrease from time t to $t+1$ is strict whenever the time t spectral measure contains two distinct contraction factors, and its limit is

$$\lim_{t\to\infty} a_t = 1 - (1 - \eta\lambda_{\min}^{\mathrm{active}})^2. \tag{44}$$

Moreover, let $r_i = (1-\eta\lambda_i)^2$ and let ν_t assign mass proportional to $c_i^2 r_i^t$ to r_i. Whenever $\mathbb{E}_{\nu_t}[r] > 0$, the one-step decline is exactly

$$a_t - a_{t+1} = \frac{\mathrm{Var}_{\nu_t}(r)}{\mathbb{E}_{\nu_t}[r]}. \tag{45}$$

To prove monotonicity, set $r_i = (1-\eta\lambda_i)^2$ and $S_t = \sum_i c_i^2 r_i^t$. Cauchy Schwarz gives $S_{t+1}^2 \le S_t S_{t+2}$, so S_{t+1}/S_t is nondecreasing. Equation 42 is one minus this ratio. The update progressively removes large eigenvalue components and leaves its slowest active direction.

The continuous time form gives a sharper identity. Let $u(t) = \exp(-tA)u_0$, $E(t) = \|u(t)\|_2^2$, and define the instantaneous log absorption rate $r(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\log E(t)$. If ν_t assigns mass proportional to $c_i^2\exp(-2\lambda_i t)$ to eigenvalue λ_i, then

$$r(t) = 2\mathbb{E}_{\nu_t}[\lambda], \qquad r'(t) = -4\,\mathrm{Var}_{\nu_t}(\lambda) \le 0. \tag{46}$$

Consequently, the absorption rate is constant exactly when the active spectral measure is concentrated on one eigenvalue. A scalar learning rate changes t to ηt and therefore rescales the horizontal axis without changing the shape indexed by effective time.

Equation 46 describes a moving probability measure over eigenvalues. At time zero, the mass assigned to λ_i is proportional to the squared coefficient c_i^2. Optimization exponentially tilts this measure by $\exp(-2\lambda_i t)$. Large eigenvalues lose mass first, so the mean eigenvalue decreases and the residual becomes increasingly concentrated on weak modes. The variance determines the speed of this concentration. A broad active spectrum produces a rapid early decline in absorption; a nearly isotropic spectrum produces a flatter curve because all active components decay at similar rates.

This measure view distinguishes spectral starvation from ordinary loss saturation. The remaining teacher gap can be large while its active mean eigenvalue is small. Conversely, a small residual aligned with a strong direction can still have high fractional absorption. The theory therefore predicts the fractional rate from the composition of the residual, not from its magnitude alone. It also predicts why the maximum gap and mean gap can respond differently: they assign different weights to query-level residual components and hence induce different active spectral measures.

Proposition B.5 (Two-mode crossover)**.** *Let $0 \le \lambda_2 < \lambda_1$, let $c_1 c_2 \ne 0$, and set $r_i = (1-\eta\lambda_i)^2$ under the step condition of Theorem B.4. Then*

$$a_t = 1 - \frac{c_1^2 r_1^{t+1} + c_2^2 r_2^{t+1}}{c_1^2 r_1^t + c_2^2 r_2^t}. \tag{47}$$

The fast and slow contributions have equal residual energy at

$$t_\times = \frac{\log(c_1^2/c_2^2)}{\log(r_2/r_1)} \tag{48}$$

whenever the right side is nonnegative. Before this crossover the observed rate can be dominated by λ_1; afterward it approaches the rate set by λ_2.

The two-mode calculation explains delayed starvation. A weak direction with a small initial coefficient can be invisible in early checkpoints even though it controls the final convergence rate. The crossover depends logarithmically on the initial coefficient ratio but inversely on the separation between contraction factors. Increasing the training horizon can therefore change which query design appears best without any change in implementation or noise.

Corollary B.6 (Finite-time active-spectrum envelope)**.** *Let P_0 project onto $\ker A$, let μ be the smallest positive eigenvalue carrying nonzero residual, and let P_μ project onto its eigenspace. Under the step condition of Theorem B.4,*

$$\|P_\mu u_0\|_2^2(1-\eta\mu)^{2t} \le \|u_t\|_2^2 - \|P_0 u_0\|_2^2 \le \|(I-P_0)u_0\|_2^2(1-\eta\mu)^{2t}. \tag{49}$$

The smallest active eigenvalue therefore controls both the asymptotic exponent and a finite-time upper envelope on the removable residual.

For a target removable residual $\varepsilon > 0$, the upper envelope yields the sufficient update count

$$t \ge \frac{\log\big(\|(I-P_0)u_0\|_2^2/\varepsilon\big)}{-2\log(1-\eta\mu)}. \tag{50}$$

This expression separates the amount of removable gap from the conditioning of the design. Adding queries can reduce the first factor by aligning with more residual, increase μ by improving weak-direction coverage, or do neither when the added trajectories are tangent-redundant.

Corollary B.7 (Best constant scalar step)**.** *Suppose the active spectrum lies in $[\mu, L]$ with $0 < \mu \le L$. The constant scalar step minimizing the worst-case one-step norm contraction is $\eta^\star = 2/(L+\mu)$, and its contraction factor is*

$$\min_{\eta>0} \max_{\lambda\in[\mu,L]} |1-\eta\lambda| = \frac{L-\mu}{L+\mu} = \frac{\kappa-1}{\kappa+1}, \qquad \kappa = \frac{L}{\mu}. \tag{51}$$

Thus changing one global learning rate cannot remove spectral conditioning. Query design changes the operator itself.

Proposition B.8 (Stable overshoot regime)**.** *Let $L = \lambda_{\max}(A)$ and suppose $0 < \eta < 2/L$. Then every active contraction factor $r_i = (1-\eta\lambda_i)^2$ lies in $[0, 1)$, the absorption sequence remains nonincreasing, and*

$$\lim_{t\to\infty} a_t = 1 - \max_{i:c_i\neq 0} (1-\eta\lambda_i)^2. \tag{52}$$

When $\eta \le 1/L$, the maximum is attained at the smallest active eigenvalue. For $\eta > 1/L$, a large eigenvalue can become the slowest contraction through sign-changing overshoot.

The proposition separates monotone absorption from monotone parameter motion. Individual modal coefficients can change sign when $\eta\lambda_i > 1$, but their squared energies still contract. Log-convexity of the total residual continues to make fractional absorption decline. The stronger step condition in Theorem B.4 identifies the late rate specifically with the smallest active eigenvalue and matches the no-overshoot regime measured in the theory-aligned experiment.

This distinction also clarifies the minimax step in Corollary B.7. The optimal scalar step deliberately equalizes the magnitudes at μ and L, placing the largest eigenvalue in the sign-changing branch whenever $\mu < L$. Its improvement comes from balancing the two endpoint contractions. Query design changes μ and L before this scalar balancing is applied.

B.4 Spectral Coverage

For query q, let A_q be the average local operator over the prefixes generated by its rollouts. A weighted query design w induces

$$A(w) = \sum_{q=1}^{n} w_q A_q, \qquad w \in \Delta_n. \tag{53}$$

Proposition B.9 (Unexcited teacher gap)**.** *For the fixed design dynamics in Equation 41, the component $P_{\ker A(w)} u_0$ is invariant. In particular,*

$$\inf_{t\ge 0} \|u_t\|_2^2 \ge \|P_{\ker A(w)} u_0\|_2^2. \tag{54}$$

Increasing the number of prefixes without enlarging the active tangent span cannot remove this residual.

This establishes a distinction between state coverage and spectral coverage. Counting distinct prefixes measures how many states were visited. Equation 54 measures whether their Jacobians span the directions that contain the teacher gap.

Proposition B.10 (Coverage by kernel intersection)**.** *For positive semidefinite query operators A_q and a design with positive support $\mathcal{I}(w) = \{q : w_q > 0\}$,*

$$\ker A(w) = \bigcap_{q\in\mathcal{I}(w)} \ker A_q. \tag{55}$$

Hence adding a positively weighted query removes an unexcited direction exactly when its operator acts nontrivially on the current kernel.

Proposition B.10 gives a precise notion of complementary queries. A new query can have a small trace or gradient norm and still be valuable if it removes a direction from the intersection. A high-energy query can be redundant if its range already lies inside the span covered by the selected set. This is the operator counterpart of a rank-increasing measurement in experimental design.

Rank alone is still insufficient. Once the kernel has been removed on a target subspace, an arbitrarily small eigenvalue can make the update count in Equation 50 arbitrarily large. Tangent coverage therefore has a qualitative part, whether a direction is present, and a quantitative part, how strongly it is excited. Residual coverage adds a third part by weighting these directions according to the discrepancy actually carried by u_0.

Proposition B.11 (Replication invariance)**.** *Suppose one query operator A_j is replaced by m identical copies whose weights sum to the original weight w_j. Then the normalized design operator $A(w)$, its kernel, its spectrum, and every linearized residual curve remain unchanged.*

Replication invariance formalizes the gap between sample count and operator coverage. A training log can contain more query identifiers, prefixes, tokens, and teacher targets while representing the same normalized tangent design. Repeated trajectories can still reduce stochastic estimation variance, but they do not change the population operator in the frozen analysis. A claim that more visited states improve conditioning therefore requires new operator directions or new weights, not only more records.

B.5 Equal Budget Optimal Design

For a subset $S \subseteq \mathcal{Q}$ with $|S| = q$, let

$$A_S = \frac{1}{q} \sum_{j \in S} A_j. \tag{56}$$

The normalization fixes total query weight, so improvements come from composition rather than from multiplying the loss by a larger batch. Four standard summaries emphasize different aspects of A_S:

$$\begin{aligned} V_{\mathrm{trace}}(S) &= \mathrm{tr}(A_S), & V_{\mathrm{D}}(S) &= \log\det(A_S + \alpha I), \\ V_{\mathrm{E}}(S) &= \lambda_{\min}(A_S|_U), & V_{\mathrm{short}}(S; u_0) &= u_0^\top A_S u_0. \end{aligned} \tag{57}$$

Trace measures total curvature, the regularized log determinant measures covered volume, E optimality protects the weakest direction, and residual energy measures immediate progress on the present gap. Equal trace does not equalize the other three criteria. The horizon-reversal construction below uses two operators with the same trace but opposite short- and long-horizon rankings.

Let U be a declared teacher-relevant tangent subspace with projector P_U. The continuous E optimal query design solves

$$\begin{aligned} \max_{w \in \Delta_n,\, \gamma \in \mathbb{R}} \quad & \gamma \\ \text{subject to} \quad & P_U A(w) P_U \succeq \gamma P_U. \end{aligned} \tag{58}$$

Theorem B.12 (Worst direction optimality)**.** *Under the continuous-time local dynamics $\dot{u} = -A(w)u$ restricted to an invariant subspace U, the worst-case squared gap satisfies*

$$\sup_{u_0 \in U \setminus \{0\}} \frac{\|u_t\|_2^2}{\|u_0\|_2^2} = \exp[-2t\lambda_{\min}(A(w)|_U)]\,. \tag{59}$$

Therefore every optimizer of Equation 58 minimizes this worst-case residual among all designs with the same total query weight.

Proposition B.13 (Concavity of E value)**.** *On a fixed invariant subspace U, the map*

$$w \longmapsto \lambda_{\min}(A(w)|_U) \tag{60}$$

is concave on the query simplex. Consequently, Equation 58 is a semidefinite-representable continuous design problem, and every feasible value γ certifies a uniform contraction rate on U.

The contrast with the short-horizon objective is instructive. Residual energy is linear in the continuous weights:

$$u_0^\top A(w) u_0 = \sum_q w_q u_0^\top A_q u_0. \tag{61}$$

Without a support, reuse, or robustness constraint, the continuous infinitesimal-horizon optimum places all mass on a query with maximal residual energy. E optimality rewards mixtures because

the minimum eigenvalue is concave and can increase when complementary ranges are combined. Residual-weighted E selection retains query amplitude while preserving this interaction between selected directions.

The finite implementation uses randomized sketches of query gradients and categorical Fisher probes. Gradient-based designs use greedy spectral updates, while the $q = 4$ Fisher designs solve the discrete subset objective exactly in the declared low-rank subspace. Gradient normalization removes scalar teacher gap magnitude in the local quadratic approximation, leaving an inexpensive proxy for tangent direction. Semantic diversity and gradient norm provide equal-budget controls.

The continuous solution in Equation 58 provides a benchmark even when the experiment requires a discrete subset. A weight vector can distribute mass across many queries, whereas a q-query experiment is constrained to a support of size at most q. In the sketched implementation, the candidate pool has 32 queries and the main budget is four. Exhaustive enumeration of the $\binom{32}{4}$ subsets is feasible in the sixteen-dimensional restricted Fisher space, so the principal $q = 4$ Fisher selectors optimize the declared discrete objective exactly rather than through a greedy approximation.

The query-gradient selectors use a lower-cost factor. Let g_q be the sketched gradient produced by the teacher residual on query q, and let $v_q = g_q/\|g_q\|_2$ when the norm is nonzero. Direction-normalized coverage uses

$$\widetilde{A}_S^{\mathrm{dir}} = \frac{1}{q}\sum_{j\in S} v_j v_j^\top, \tag{62}$$

whereas residual-weighted coverage uses

$$\widetilde{A}_S^{\mathrm{res}} = \frac{1}{q}\sum_{j\in S} g_j g_j^\top. \tag{63}$$

The first removes scalar gap magnitude and asks for complementary directions. The second keeps magnitude, so a query must be both relevant to the current residual and complementary to the selected set. Gradient norm keeps only $\|g_q\|_2$ and omits the interaction between directions. Thus, direction normalization isolates geometric coverage, while residual weighting preserves how strongly each query couples to the current teacher discrepancy. These three constructions isolate amplitude, direction, and their combination.

The Fisher selectors replace one gradient per query with multiple curvature-probe factors. For categorical probabilities p_s, draw deterministic zero-mean probe vectors ξ with covariance $C(p_s)$ and compute $J_s^\top \xi$. Averaging the resulting outer products gives an unbiased low-rank factorization of the Gauss-Newton pullback in expectation over probes. Unlike a single residual gradient, these factors probe the local curvature along multiple categorical directions and therefore provide a more direct estimate of the query-induced operator. Teacher-Fisher designs use $p_s = \pi_{\mathrm{T}}(\cdot \mid s)$, while student-Fisher designs use $p_s = \pi_\theta(\cdot \mid s)$. The experiment uses four deterministic probes per query so that every candidate has the same curvature-sampling budget. Table 3 maps every selector to its representation and local objective.

B.6 The Optimal Design Depends on Training Horizon

E optimality protects an unknown worst direction. For the realized teacher gap, define the local residual after optimization time τ as

$$\Phi_\tau(w; u_0) = u_0^\top \exp(-2\tau A(w))u_0. \tag{64}$$

Theorem B.14 (Horizon-dependent query value)**.** *For every fixed design w,*

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\Phi_\tau(w; u_0)\bigg|_{\tau=0} = -2u_0^\top A(w)u_0. \tag{65}$$

If u_0 has a nonzero component in the positive eigenspace of $A(w)$, then

$$\lim_{\tau\to\infty} -\frac{1}{2\tau}\log\big(\Phi_\tau(w; u_0) - \|P_{\ker A(w)}u_0\|_2^2\big) = \lambda_{\min}^{\mathrm{active}}(A(w), u_0). \tag{66}$$

Thus short-horizon design maximizes realized gradient energy, while long-horizon design improves the smallest eigenvalue carrying teacher gap. These objectives can rank query sets differently.

Table 3: Equal-budget query selectors and the local quantity emphasized by each. All neural comparisons use identical query counts and update schedules.

Selector	Representation	Local target
Random	candidate identifiers	Equal-probability reference design
Semantic diversity	prompt embeddings	Dispersion in input representation space
Gradient norm	$\|g_q\|_2$	Immediate residual energy without direction interaction
Gradient log determinant	$g_q g_q^\top$	Regularized volume of selected gradient factors
Gradient E	$v_q v_q^\top$	Weakest direction after gradient normalization
Residual E	$g_q g_q^\top$	Weakest residual-weighted gradient direction
Teacher Fisher E	teacher curvature probes	Weakest equality-limit Fisher direction
Teacher Fisher residual	weighted teacher probes	Equality-limit coverage retaining gap amplitude
Student Fisher E	student curvature probes	Weakest current Gauss-Newton direction
Student Fisher residual	weighted student probes	Current Gauss-Newton coverage retaining gap amplitude

Proposition B.15 (Equal budget horizon reversal)**.** *There are two one-query positive definite designs with equal trace for which one design gives greater infinitesimal absorption but the other gives the faster asymptotic rate. In two dimensions, take*

$$u_0 = (\sqrt{0.9}, \sqrt{0.1})^\top, \qquad A_1 = \mathrm{Diag}(1.9, 0.1), \qquad A_2 = I. \tag{67}$$

Both operators have trace two, while $u_0^\top A_1 u_0 = 1.72 > 1 = u_0^\top A_2 u_0$ and $\lambda_{\min}(A_1) = 0.1 < 1 = \lambda_{\min}(A_2)$. Each operator is realizable by one query through its positive semidefinite Jacobian factor. Hence A_1 is preferred at an infinitesimal horizon and A_2 at a sufficiently long horizon.

For a quadratic token loss, $u_0^\top A_q u_0$ is the squared parameter gradient norm induced by query q. Gradient norm ordering is therefore the exact infinitesimal-horizon control. Direction-normalized E optimality targets the robust long-horizon endpoint. Residual-weighted E optimality retains gradient magnitude inside the selected Gram operator and supplies an intermediate design. The experiments report all three rather than assigning one method universal dominance.

The horizon theorem also changes how query budgets should be compared. Increasing q can improve a random design by gradually filling missing directions, but a targeted design can obtain the same span with fewer queries. At a short horizon, adding low-energy directions can dilute the average operator in Equation 56; at a longer horizon, those directions can determine whether a residual floor remains. The value of an additional query therefore depends on both its marginal operator and the time available for its weak directions to matter.

For reporting, we keep finite-horizon reduction and robust spectral coverage separate. Maximum-gap reduction evaluates progress on the realized teacher residual, while the smallest active eigenvalue measures protection of the weakest covered direction. The horizon theorem predicts when their selector orderings agree and when they reverse.

B.7 From Teacher Gap to Task Decisions

Teacher-gap absorption is a continuous distributional objective. Exact task utility is determined by a sequence of discrete decoding decisions. The link between them depends on margins and on whether teacher probability changes point toward the task reference.

Proposition B.16 (Decision-cell stability)**.** *Let $z \in \mathbb{R}^V$ have unique maximizing class $a^\star$ and margin*

$$m(z) = z_{a^\star} - \max_{a \neq a^\star} z_a. \tag{68}$$

For any logit perturbation δ satisfying

$$2\|\delta\|_\infty < m(z), \tag{69}$$

the maximizing class of $z + \delta$ remains $a^\star$. Consequently, a continuous teacher-gap reduction can leave an exact token decision unchanged until the accumulated logit motion reaches a decision boundary.

The proposition applies at every autoregressive position. Exact sequence accuracy can remain unchanged even when every token distribution moves toward the teacher, because the greedy path is piecewise constant inside the product of its visited decision cells. Once one boundary is crossed, the generated suffix and all later occupied states can change discontinuously. This creates a measurement scale different from the smooth local operator.

Proposition B.17 (Teacher alignment and gold-margin sign)**.** *Consider a binary token decision with gold-class student probability $p = \sigma(\ell)$ and teacher gold probability τ. One gradient step of size $\eta > 0$ on teacher cross entropy updates the gold log odds by*

$$\ell^+ = \ell + \eta(\tau - p). \tag{70}$$

The gold margin increases when $\tau > p$, is stationary when $\tau = p$, and decreases when $\tau < p$. Teacher-gap descent therefore translates into task-margin progress only after the sign and magnitude of teacher guidance are related to the gold decision.

Proposition B.17 identifies two requirements for a task-level consequence. First, the selected operator must act on the residual direction, which is the subject of the spectral design. Second, that residual direction must improve the task margin on the evaluated examples. A discrete decision changes after the cumulative motion crosses the stability region in Proposition B.16.

B.8 Changing On-Policy Occupancy

The operator changes when updated students generate different prefixes. The next statement isolates the cost of this drift.

Theorem B.18 (Drifting operator bound)**.** *Suppose a teacher-relevant subspace U is invariant, $u_0, r_t \in U$, $A_t \succeq \mu I$ and $A_t \preceq LI$ on U, and the nonlinear remainder obeys $\|r_t\|_2 \le \varepsilon_t$. For $0 < \eta \le 1/L$, the dynamics*

$$u_{t+1} = (I - \eta A_t)u_t + r_t \tag{71}$$

satisfy

$$\|u_T\|_2 \le (1 - \eta\mu)^T \|u_0\|_2 + \sum_{t=0}^{T-1} (1 - \eta\mu)^{T-1-t} \varepsilon_t. \tag{72}$$

Proposition B.19 (Commuting operator path)**.** *Suppose the symmetric operators $A_0, \dots, A_{T-1}$ commute, $r_t = 0$, and all use the same step size. Let $\lambda_{i,t}$ be their eigenvalues in a common orthonormal basis and write $u_0 = \sum_i c_i v_i$. Then*

$$\|u_T\|_2^2 = \sum_i c_i^2 \prod_{t=0}^{T-1} (1 - \eta\lambda_{i,t})^2. \tag{73}$$

Thus changing occupancy accumulates directional exposure through products of modal contractions even when no single frozen operator represents the path.

The commuting case gives an exact intermediate model between fixed occupancy and fully nonlinear OPD. A direction can be weak at the common checkpoint and become strong after the student visits a new prefix. Its cumulative contraction depends on the complete exposure sequence. Conversely, a direction that disappears from later occupancies retains only the contraction obtained before it vanished. Noncommuting operators additionally rotate the modal basis, which is absorbed into the remainder treatment of Theorem B.18 and Proposition B.20.

The result gives a direct empirical test. A spectrum measured at one checkpoint predicts subsequent absorption when the accumulated remainder in Equation 72 is small. Fixed-prefix training sets the occupancy drift to zero, while fully on-policy training measures the additional discrepancy.

The remainder r_t combines three departures from the frozen linear model. The first is occupancy drift: the updated student generates different prefixes and therefore changes which Jacobians are averaged. The second is tangent drift: even on a repeated prefix, the Jacobian changes with the student parameters. The third is curvature error: the nonlinear network-Hessian term in Proposition B.2 becomes important away from agreement. The bound treats their aggregate norm directly and does not require them to share a sign.

Proposition B.20 (Deviation from a frozen reference)**.** *Let* $u_{t+1} = (I - \eta A_t)u_t + r_t$ *and let* $\bar{u}_{t+1} = (I - \eta A)\bar{u}_t$ *with* $u_0 = \bar{u}_0$*. If* $\|I - \eta A\|_{\mathrm{op}} \le 1$*, then*

$$\|u_T - \bar{u}_T\|_2 \le \sum_{t=0}^{T-1} \left(\eta \|A_t - A\|_{\mathrm{op}} \|u_t\|_2 + \|r_t\|_2\right). \tag{74}$$

If the reference map is strictly contractive, each summand can be multiplied by the corresponding remaining contraction factor.

Proposition B.20 links the matched experiment to an operator-level quantity. Reusing prefixes makes the occupancy component of $A_t - A$ zero by construction, while regenerated prefixes can change nearly every revisit. The paired difference in absorption curves measures the accumulated effect after the optimizer, update count, query subset, and initial state have been matched. Trajectory hash turnover is an exposure diagnostic: it measures how often the sampled prefix changes. It does not replace the operator norm in Equation 74, because two different strings can have similar Jacobians and one small textual change can alter a high-curvature token position.

The two drift results give complementary predictions. Theorem B.18 guarantees contraction when every instantaneous operator remains uniformly positive on the relevant subspace and the accumulated remainder is controlled. Proposition B.20 compares the realized path with one frozen reference even when a uniform positive lower bound is unavailable. The experiments therefore report both the temporal absorption signature and the direct fixed-versus-regenerated paired contrast.

C Experimental Design

The experiments follow the sequence theory from exact dynamics to neural OPD. Within each training mode, every selector comparison fixes query count, rollout schedule, teacher-supervised token positions, student updates, and evaluation states. Query selection is completed before training outcomes are available.

The design separates algebraic verification, controlled local prediction, changing occupancy, and scale transfer. The exact finite system verifies identities without model approximation. Fixed-prefix neural training tests the local frozen-spectrum forecast when the occupied sequences are held constant. Fully on-policy training restores the endogenous trajectory distribution. The matched AdamW panel changes only prefix regeneration, the unclipped panel aligns the optimizer with the linear theory, and the scale panel changes both student and teacher size. These components measure successive parts of one operator path.

All primary choices were made from the common warm-start state. Selector construction never uses a later training outcome, and the statistical unit is always a training seed. The analysis reports continuous estimates and paired intervals without an empirical success threshold. Pure E optimality, residual-weighted coverage, and gradient energy retain their distinct theoretical objectives throughout the design.

C.1 Theory-to-Measurement Map

Table 4 maps each theoretical statement to its empirical object. The map separates exact identities from neural predictions. The exact finite construction can verify equality to numerical precision. Neural experiments test directions, orderings, and paired contrasts because the network is nonlinear, the operator is sketched, and stochastic trajectories change occupancy.

The primary local prediction is quantitative: the initial gradient and student-Fisher operator should forecast the next 1, 2, 4, 8, and 16 updates. The temporal prediction is that absorption-time association is negative under a fixed stable operator. Gradient energy should align with early absorption, while pure E optimality should raise the smallest eigenvalue it targets. The drift prediction links growing tangent replacement to growing frozen-forecast error, then isolates the additional effect of occupancy regeneration under an otherwise matched update rule.

Three controls guard the interpretation. The equal-budget construction fixes total operator weight. The common adapter fixes the starting parameters within seed. The matched AdamW panel fixes the

Table 4: Theory-to-measurement map. Each row names the operator quantity and the experiment that evaluates its prediction.

Theory component	Measured quantity	Experimental realization
Local OPD geometry	Fisher eigendirections and Hessian scale	Finite Hessian certificate; teacher- and student-Fisher selector factorial
Spectral starvation	Decline of per-update absorption	Exact recurrence; fixed-prefix $q = 1$, $q = 4$, and $q = 16$ matrices; unclipped panel
Frozen spectral forecast	Gap-curve calibration and rank prediction	Initial gradient and student-Fisher predictor at updates 1, 2, 4, 8, and 16
Unexcited gap	Kernel component and weak-direction residual	Exact redundant design; candidate-span and minimum-eigenvalue measurements
Worst-direction optimality	Minimum restricted eigenvalue	Gradient E, teacher-Fisher E, and student-Fisher E selectors
Horizon dependence	Early energy versus endpoint reduction	Gradient norm, Residual E, and pure E designs at $q = 1$ and $q = 4$
Selected-set efficiency	Targeted $q = 4$ versus nested random $q = 16$	Within-seed equal-update maximum-gap contrast
Changing operator	Fixed-reference deviation and turnover	Matched AdamW prefix reuse versus regeneration
Scale transfer	Direction of paired selector advantage	Qwen3.5 4B student and 9B teacher panel
Cross-domain signature	Temporal starvation strength	ARC-Challenge science reasoning under fixed and regenerated prefixes

optimizer while changing prefix reuse. These controls correspond directly to the count, state, and drift terms in the theory.

C.2 Exact Finite Certificate

We construct positive semidefinite absorption operators with controlled spectra and compare redundant, uniform, and spectrum-balanced query collections at a common budget. The implementation verifies Equation 43 to floating-point precision and records the complete loss and absorption curves. It also checks the monotonicity condition in Theorem B.4.

The system has dimension twelve and a pool of nearly axis-aligned query features. The first three directions are deliberately overrepresented by a factor of ten, creating a candidate distribution in which naive selection can spend most of its budget on already strong modes. Every feature receives independent Gaussian perturbation with standard deviation 0.01 before normalization. The initial teacher-gap vector is sampled once from an isotropic Gaussian and normalized. All designs select 24 features and therefore have the same total query weight.

The redundant design takes the first 24 pool entries, the uniform design samples 24 entries without replacement, and the spectrum-balanced design applies greedy log-determinant selection with ridge 10^{-8}. Their operators are averages of selected feature outer products. A common step size is set to 0.9 divided by the largest eigenvalue among the three operators, placing every design inside the stability condition of Theorem B.4. Two hundred updates are computed both by direct recurrence and by the eigenvalue formula. The script requires agreement to absolute tolerance 10^{-12} and relative tolerance 10^{-10}, verifies nonincreasing absorption, and stores every loss and rate value.

Separate numerical certificates exercise the theorem statements over randomized operators. One hundred eight-dimensional trials verify monotonicity and the discrete variance identity. Additional certificates test the local Fisher Hessian for forward KL, reverse KL, and Jensen-Shannon divergence, kernel invariance, the E optimal solution of a two-direction design, the horizon reversal, the scalar

minimax step, and the changing-operator recurrence. These certificates are deterministic given their recorded seeds and accompany the analytic proofs.

C.3 QWEN3.5 QUERY DESIGNS

The main model pair uses the official Qwen3.5 0.8B student and Qwen3.5 4B teacher checkpoints (Qwen Team, 2026), whose architecture follows the Qwen3 lineage (Yang et al., 2025). A scale transfer panel uses the official 4B student and 9B teacher checkpoints from the same collection. The student is trained through rank-eight LoRA adapters. Student rollouts receive the teacher's normalized top-16 token distribution at every generated position. The task suite contains GSM8K mathematical reasoning (Cobbe et al., 2021) and ARC-Challenge multiple-choice science reasoning (Clark et al., 2018), with held-out prompts used only for evaluation.

The model revisions are frozen by commit hash in the released environment receipt. Prompts are truncated to 384 tokens and student rollouts generate at most 64 new tokens during training. At each generated position, the teacher top-16 probabilities are renormalized and used as the soft target. The same target representation is used by every selector and both training modes. LoRA rank is eight with scale parameter 16. The trainable coordinate system in the spectral analysis is exactly the flattened LoRA parameter vector loaded by the update code.

Each seed deterministically partitions GSM8K into a training pool of 128 prompts, a 32-query candidate design pool, sixteen heldout evaluation prompts, and sixteen probe items for the main model pair. The scale-transfer panel uses sixteen candidate queries, eight heldout prompts, and eight probes to keep the larger-model measurement cost bounded. No heldout outcome enters query selection. Prompt embeddings used by semantic diversity are computed at the common checkpoint, while every gradient or Fisher representation is derived from teacher supervision on the corresponding student rollout.

Each training seed creates a candidate query pool. For every candidate, we generate one student rollout and project all trainable coordinates into a common randomized CountSketch (Charikar et al., 2002). The first six equal budget selectors use uniform random ordering, prompt embedding diversity, gradient norm ordering, greedy log determinant ordering, direction normalized E optimal ordering, and residual weighted E optimal ordering. Four principal selectors estimate the operator in Equation 36 more directly. They draw four deterministic categorical curvature probes per query, compute the corresponding Jacobian vector products, restrict the pooled operator to its leading sixteen dimensional subspace, and solve the $q = 4$ subset search exactly for the largest minimum eigenvalue. Two use teacher curvature, which agrees with Theorem B.1 at equality, and two use the current student curvature, which is the Gauss Newton matrix of the implemented cross entropy away from equality. Within each pair, the residual weighted version retains the initial teacher gap magnitude. The query counts are one, four, and sixteen.

Throughout the paper, query count means the cardinality of the unique selected training-prompt set. Constructing a gradient or Fisher selector requires teacher supervision on all 32 candidate rollouts before four prompts are selected. Fisher probes reuse the stored categorical targets and add student backward passes but no additional teacher forward. The actual experimental protocol performs this common screen for every method so that candidate assets are identical. Random selection can be implemented without this teacher-supervised screen, and the minimal algorithmic accounting therefore assigns zero screening calls to random. Screening, selected-set cardinality, online teacher trajectories, and teacher-supervised token positions are reported as four separate costs.

The CountSketch has dimension 128 and is shared across candidates within a seed. Every trainable coordinate is mapped to one row with a deterministic sign, so inner products are comparable across all candidate queries. The leading sixteen-dimensional Fisher subspace is computed from the pooled candidate factors rather than separately for each subset. This fixes the target coordinates before subset enumeration. Realized spectra are recomputed after selection using the actual frequency with which the deterministic schedule visits each chosen query.

The four categorical probes per query use deterministic seeds derived from the base seed, candidate identifier, curvature source, and probe index. Teacher-Fisher and student-Fisher factors therefore differ only in the categorical curvature distribution used to generate the probe. Direction-normalized variants divide away query-level residual magnitude before constructing the subset operator. Residual

variants preserve it. This factorial design distinguishes the equality-limit versus current Gauss-Newton curvature choice from the amplitude versus direction choice.

One query supplies at most four curvature probes, so every $q = 1$ Fisher design is rank deficient in the sixteen dimensional target subspace. Its E optimal value is therefore exactly zero. We break this mathematical tie lexicographically by rank, pseudo log determinant, trace, and candidate identifier after projecting numerical zero eigenvalues to zero. This rule prevents floating point noise from defining the $q = 1$ selector. The random ordering is nested across $q \in \{1, 4, 16\}$; the complete $q = 1$ and $q = 4$ design matrices measure how selector value changes with budget.

The tie rule is part of the scientific definition of the $q = 1$ Fisher selectors. Rank is compared first because a direction added to the restricted span changes the removable subspace. Pseudo log determinant is compared next on the positive spectrum, followed by trace and candidate identifier. Candidate identifiers provide a deterministic final order rather than a data-dependent threshold. For $q = 4$, the sixteen curvature probes can span the declared sixteen-dimensional space, so the exact subset objective can produce a positive minimum eigenvalue.

Every formal comparison begins from a seed specific common adapter produced by eight deterministic OPD updates using queries drawn from the training pool. The adapter is hashed and restored identically for every selector and training mode. A deterministic seeded schedule chooses among the selected queries at each update. The reported gradient and Fisher spectra use the realized exposure frequencies from that schedule. Candidate gradients and all absorption measurements are computed after the common start. A separate unadapted-start matrix measures the nonlinear entry transient and is kept outside the primary estimates.

The common adapter serves three purposes. It places the student inside a nontrivial teacher-gap regime, ensures that every selector begins from identical parameters within a seed, and supplies the checkpoint at which all query representations are measured. The eight warmup updates use AdamW with learning rate 10^{-4}. The serialized adapter is hashed, and the training runner verifies that hash before applying any selector-specific update. Warm-start outputs, candidate rollouts, teacher targets, embeddings, gradient sketches, Fisher factors, evaluation states, and query orders are stored in one immutable precompute record per seed.

The selected-query schedule is deterministic but generally not perfectly uniform over a finite number of updates. We therefore distinguish the nominal subset operator from the realized exposure-weighted operator. If one selected query appears one additional time, its operator receives the corresponding additional mass in the measured spectrum. This adjustment prevents a scheduling remainder from being interpreted as a selector difference.

C.4 Fixed-Prefix and Fully On-Policy Tests

The fixed-prefix condition reuses the initial generated sequences and teacher targets throughout training. It tests the frozen operator theory without occupancy drift. The fully on-policy condition regenerates the student trajectory before every update and queries the teacher on the resulting prefixes. Both conditions report teacher gap on the complete candidate design pool and on a disjoint held out pool.

The primary fixed-prefix matrix uses SGD with learning rate 3×10^{-3} and global gradient clipping at norm one. The primary fully on-policy matrix uses AdamW with learning rate 10^{-4} and the same clipping norm. These update rules represent practical stable settings for the two modes. Because optimizer and occupancy both differ in the primary matrix, it establishes the phenomenon under two complete training pipelines rather than attributing their difference to occupancy alone.

We retain three fixed-prefix panels. The practical SGD panel uses gradient clipping at norm one. The theory-aligned panel removes clipping and reduces the SGD learning rate from 3×10^{-3} to 3×10^{-4} so that its initial update norm remains on the same scale. This panel tests the quantitative spectral-rate prediction under the update rule analyzed in Section B. A matched-optimizer panel uses AdamW at learning rate 10^{-4} with clipping at norm one for the random, gradient norm, teacher-Fisher residual, and student-Fisher residual designs. Its paired on-policy comparison holds the base adapter, query subset, optimizer, learning rate, clipping, and update count fixed, isolating prefix regeneration from the optimization rule.

All main and auxiliary 0.8B-student panels run for 256 measured updates. Checkpoints occur after updates 1, 2, 4, 8, 16, 32, 64, 128, and 256, in addition to the common step-zero measurement. The geometric schedule resolves the fast early phase without sacrificing a long late interval. Teacher gap is reevaluated on the full candidate pool and heldout pool at each checkpoint using cached evaluation prefixes. The fully on-policy update trajectory changes during training, but its evaluation distribution remains fixed within a run, so temporal gap ratios are comparable.

C.5 Frozen-Spectrum Forecast and Its Validity Window

The quantitative forecast uses the 160 unclipped fixed-prefix runs, comprising ten selectors and sixteen seeds. At the common checkpoint, each candidate query supplies a teacher-residual gradient sketch g_i and a student-Fisher Gauss-Newton sketch H_i. We apply Equations 39 and 40 along the realized query schedule, using the complete 32-query design pool to define g_{eval} and H_{eval}. Predicted and measured design-pool gaps are compared at updates 1, 2, 4, 8, and 16 using RMSE, R^2, calibration slope, and rank correlation.

The validity-window panel replays Random, Gradient norm, Residual E, and Teacher-Fisher E under the same unclipped SGD rule for all sixteen seeds. At updates 1, 2, 4, 8, 16, 32, and 64, it recomputes student-Fisher factors on the fixed selected prefixes and measures relative spectral drift from H_0. The replayed complete-pool gap is checked against the archived run at every checkpoint. Joining the 448 seed, selector, and checkpoint records tests whether tangent drift tracks absolute frozen-forecast error. A seed-cluster bootstrap resamples all selector and checkpoint records belonging to one training seed together.

C.6 Direct Operator-Drift Panel

The direct drift panel replays the on-policy $q = 4$ runs for Random, Gradient norm, Residual E, and Teacher-Fisher residual under all sixteen training seeds. Replay uses the original common adapter, selected candidates, optimizer, query schedule, and per-update generation seeds. Updates 64 and 128 define the measured replay trajectory. At update 256 we restore the archived final adapter before constructing the terminal operators, so every terminal association is evaluated at the exact parameters of the corresponding original run. We report the replay-to-archive parameter distance separately as a reproducibility diagnostic.

At updates 64, 128, and 256, the current student is evaluated in two occupancy conditions. The fixed condition recomputes gradients and student-Fisher factors on the original cached selected prefixes. The regenerated condition produces a fresh deterministic student trajectory for each selected query, obtains its teacher target, and recomputes the same factors. The common-start operator is reconstructed from the original precompute sketches. All comparisons use the realized query-frequency weights of the original 256-update schedule.

For both the teacher-residual gradient operator and the student-Fisher operator, we report relative Frobenius drift, relative spectral-norm drift, relative eigenvalue drift, maximum principal angle, mean principal angle, and mean sine principal angle. Three contrasts are retained: fixed current Jacobians versus the common-start operator, regenerated current occupancy versus common start, and regenerated occupancy versus fixed prefixes at the same checkpoint. The last contrast isolates occupancy replacement after holding the current student parameters fixed.

Each final drift measurement is joined to the matched AdamW attenuation outcomes for the same seed and selector. The joined quantities are the on-policy minus fixed-prefix absorption-time association, the absolute late-absorption difference, and the absolute maximum-gap-reduction difference. Pooled and within-selector rank associations test whether larger measured operator drift accompanies greater attenuation of the frozen-spectrum prediction.

C.7 Scale Transfer

The scale-transfer matrix replaces the 0.8B student and 4B teacher with a 4B student and 9B teacher. It uses eight new seeds, four query selectors, two training modes, four queries, and 128 updates, producing 64 runs. The selectors are random, gradient norm, teacher-Fisher residual, and student-Fisher residual. These choices retain the random reference, the short-horizon target, and both

curvature conventions while keeping the 9B teacher cost bounded. The same prompt limits, top-16 teacher targets, sketch dimension, optimizer rules, and geometric measurement schedule are used.

C.8 Cross-Domain Temporal Test

The ARC-Challenge panel retains the 0.8B student, 4B teacher, sixteen seeds, 32-query candidate pool, four selected queries, 256 updates, and both primary training modes. Science questions are rendered with their labeled answer choices, and student trajectories receive the same top-16 token supervision as the mathematics panel. Random, Gradient norm, Gradient E, and Residual E span the baseline, short-horizon energy, direction-normalized coverage, and residual-weighted coverage objectives. The primary cross-domain statistic is starvation strength $s = -\rho_t$. The 128-run matrix estimates this quantity for every selector under fixed-prefix SGD and fully on-policy AdamW.

The primary dynamical outcome is the per step absorption inferred between geometrically spaced checkpoints. Secondary outcomes are the initial restricted spectrum, teacher-gap curve, and wall clock cost. Training seed is the independent replication unit. No checkpoint, prefix, token, or repeated measurement is treated as an independent replicate.

For every run we retain mean and maximum design-pool teacher gap, mean and maximum heldout gap, selected candidate identifiers, realized exposure weights, initial gradient and Fisher spectra, per-update losses and gradient norms, trajectory hashes, wall time, and the final adapter hash. On-policy trajectory turnover is the fraction of consecutive revisits to a candidate whose generated text hash changes. It records occupancy replacement without treating individual strings as independent samples.

The absorption-time statistic is computed from the nine checkpoint intervals using Equation 28. Its sign is summarized over the complete selector by seed matrix, while selector comparisons use paired seed-level reductions. The initial gradient-energy association with early absorption and the minimum-eigenvalue association with late absorption are computed within seed before averaging across seeds. This prevents between-seed scale differences from driving the association.

C.9 Statistical Analysis

Primary contrasts are paired within model pair, query count, training mode, and seed. We report seed means, paired bootstrap intervals, and exact sign flip inference. Continuous spectrum to absorption predictions are evaluated using seedwise rank association and calibration error. The protocol contains no empirical threshold for declaring a run successful.

Bootstrap intervals resample the independent training seeds with replacement and use 10,000 draws. Exact sign-flip inference enumerates all 2^{16} sign assignments for the main panels and all 2^8 assignments for scale transfer. Max-minus-min summaries are not used as population separation tests. Query selectors are compared through prespecified paired differences against nested random selection, and the query-count analysis keeps candidate order fixed across $q = 1$, $q = 4$, and $q = 16$.

The final compute ledger contains 1,088 training runs, 40 precompute runs, 64 early tangent measurements, and 64 direct on-policy operator measurements. Runs execute on NVIDIA A100-class 40GB accelerators. Every aggregate JSON, plotting program, theorem certificate, model revision, and environment version is included in the artifact manifest.

D Results and Analysis

The exact finite certificate passes the modal identity and monotonicity checks. In the initial twelve dimensional construction with an equal budget of twenty four queries, the redundant, uniform, and spectrum balanced designs have active condition numbers approximately 1.30×10^5, 3.10×10^4, and 1.08. After two hundred steps, their squared residuals are 0.943, 0.450, and 3.85×10^{-14}, respectively. These values are deterministic consequences of the declared operators and serve as the exact theorem illustration.

All three operators have full algebraic rank, so rank alone does not explain their behavior. The redundant design has smallest eigenvalue 7.66×10^{-6} and largest eigenvalue 0.999. Its initial absorption is 0.00288 and falls to 0.000277. Uniform selection has smallest eigenvalue 1.48×10^{-5}

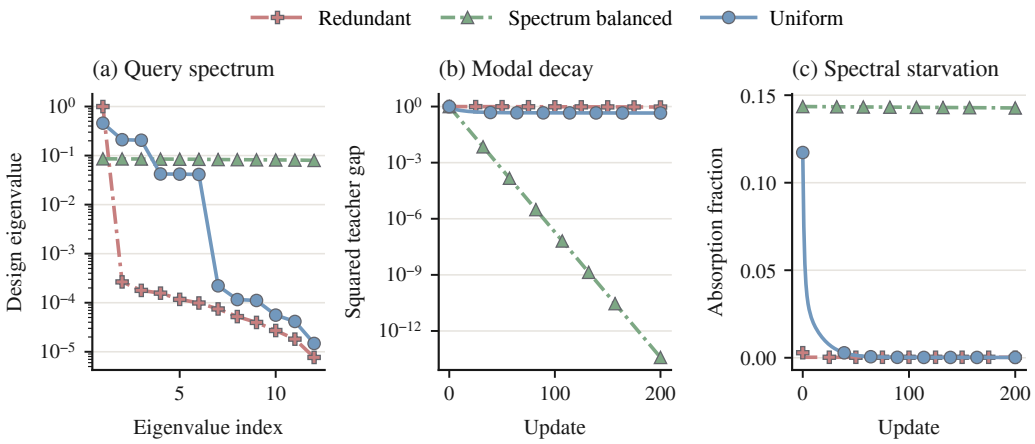


Fig. 4. Exact equal budget spectral certificate. Redundant and uniform query collections leave weak directions that dominate late optimization, while the spectrum balanced design equalizes the active modes. The right panel shows the decreasing per update absorption predicted by Theorem B.4.

Table 5: Frozen A_0 prediction of mean design-pool teacher-gap reduction. Statistics pool ten selectors and sixteen seeds at each checkpoint.

Future updates	Observed reduction	Predicted reduction	R^2	Rank correlation	RMSE
1	0.0012	0.0010	0.810	0.390	0.0029
2	0.0017	0.0016	0.930	0.652	0.0018
4	0.0043	0.0037	0.893	0.864	0.0041
8	0.0067	0.0059	0.897	0.916	0.0047
16	0.0105	0.0095	0.911	0.898	0.0051

and largest eigenvalue 0.458. It begins with absorption 0.117 but ends at 0.000183 because the early reduction is carried by a small number of strong modes. The spectrum-balanced operator places all twelve eigenvalues between 0.0797 and 0.0859. Its absorption remains near 0.143 from the first update through update 200.

This comparison isolates the two parts of spectral starvation. Redundant and uniform designs both leave weak directions despite visiting enough features to achieve full rank. Their fast components disappear early, exposing the weak tail and producing orders-of-magnitude decline in the fractional rate. The balanced design equalizes the active contraction factors, making the variance term in Equation 45 small at every step. Equal query count, rank, initial residual, and common step size therefore coexist with sharply different loss curves.

The randomized theorem certificate verifies the discrete identity over 100 operators. Its largest absorption increase is 5.55×10^{-16} and its largest rate-variance error is 9.42×10^{-16}. The local-geometry certificate recovers curvature scales 1 for forward and reverse KL and $1/4$ for Jensen-Shannon divergence, with relative Fisher errors below 9.7×10^{-8}. Kernel-component change is 1.05×10^{-14}, and the equal-trace horizon construction reverses order at optimization time 1.221. Fig. 4 visualizes the resulting separation among redundant, uniform, and spectrum-balanced designs.

D.1 The Initial Spectrum Predicts the Local Neural Future

The frozen student-Fisher predictor is quantitatively accurate over the declared short horizon. Across 160 unclipped Transformer runs and 800 seed, selector, and checkpoint predictions through update 16, it attains $R^2 = 0.906$, rank correlation 0.925, and RMSE 0.0039 for complete design-pool teacher-gap reduction. Table 5 resolves this result by forecast horizon. R^2 is 0.810 after one update, 0.930 after two, and 0.911 after sixteen. Rank correlation rises from 0.390 at the one-step scale to 0.898 at update 16 as the accumulated reduction separates the query designs.

Table 6: Formal $q = 4$ results over sixteen training seeds. Reductions are initial minus update 256 teacher gap. ρ_t is the rank association between checkpoint interval and per update absorption.

	Fixed prefix, SGD			On policy, AdamW		
Query design	**Max red.**	**Held red.**	ρ_t	**Max red.**	**Held red.**	ρ_t
Random	0.074	0.005	-0.652	0.125	0.010	-0.090
Gradient norm	0.274	0.011	-0.704	0.233	0.006	-0.408
Residual E	0.279	0.012	-0.733	0.235	0.007	-0.469
Teacher Fisher E	0.142	0.013	-0.588	0.146	0.018	-0.328
Teacher Fisher residual	0.152	0.016	-0.609	0.152	0.018	-0.343
Student Fisher residual	0.204	0.011	-0.678	0.192	0.014	-0.457

The matched replay locates the boundary of this forecast. Mean relative tangent drift grows from 0.344 after one update to 0.505 after sixteen, 0.572 after 32, and 0.649 after 64. Relative to update 16, mean absolute forecast error increases by factors of 3.1 and 21.9 at updates 32 and 64. Across all 448 replay records, tangent drift and absolute forecast error have rank correlation 0.504 with seed-cluster interval $[0.392, 0.623]$. Their seven checkpoint means have rank correlation 0.929. The replayed complete-pool gap remains within 4.9×10^{-4} of the archived trajectory at every checkpoint, separating forecast breakdown from replay error.

D.2 Neural Absorption Slows Under Fixed and Changing Occupancy

The formal $q = 4$ matrix contains 320 runs from ten query designs, two practical update rules, and sixteen seeds. The rank association between checkpoint interval and per step absorption was negative in 158 of 160 fixed-prefix SGD runs, with mean -0.649. It was negative in 128 of 160 fully on-policy AdamW runs, with mean -0.301. The frozen-occupancy rule exhibits the qualitative modal decline predicted by Theorem B.4. The fully on-policy rule has greater early variation and its absorption approaches zero over the declared intervals. The matched AdamW panel below isolates the occupancy contrast.

The fixed-prefix result is nearly universal across selectors: only two of 160 runs have positive temporal association. The mean associations of the individual methods range from -0.588 for teacher-Fisher E to -0.733 for Residual E. This consistency matters because the selectors induce very different spectra and final reductions. The shared temporal direction is a property of optimization under heterogeneous fixed operators rather than a consequence of one favorable query heuristic.

The on-policy condition retains the same aggregate direction but broadens the run-level distribution. Thirty-two of 160 runs have positive association, and the method means range from -0.231 for Gradient E to -0.469 for Residual E. Early intervals can be positive or negative because regenerated student prefixes change both the sampled teacher targets and the effective operator. By the final intervals, the mean absorption curves concentrate near zero. Fig. 2 shows this difference in scale and variability.

Table 6 separates temporal association from endpoint reduction. Random fixed-prefix training reduces maximum design gap by 0.074, while random on-policy training reduces it by 0.125. These baselines already show that the practical optimizer and occupancy rule alter finite-horizon behavior. Paired selector contrasts below therefore remain within one training mode rather than comparing raw endpoints across modes.

D.3 Realized Gap and Worst Direction Coverage Favor Different Designs

Finite-horizon reduction separates residual magnitude from spectral coverage. In fixed-prefix training, gradient norm and residual-weighted E optimal selection improved maximum-gap reduction over random selection by 0.200 and 0.205, with paired intervals $[0.127, 0.276]$ and $[0.133, 0.280]$. Their fully on-policy improvements were 0.108 and 0.110, with intervals $[0.041, 0.180]$ and $[0.055, 0.168]$. The residual design therefore retains its advantage under regenerated student trajectories.

The similarity between gradient norm and Residual E at this horizon is predicted by the short-horizon limit. Both retain teacher-gap amplitude, and 256 updates remain sensitive to directions that are

large at the common checkpoint. Their agreement does not make the spectral interaction redundant. Moving from $q = 1$ to $q = 4$ produces a larger gain for Residual E, and its four-query on-policy design exceeds the sixteen-query random endpoint. The value appears when multiple directions can be combined under a fixed budget.

Student-Fisher residual selection occupies a second favorable regime. Its fixed-prefix maximum-gap advantage over random is 0.130 with interval $[0.063, 0.202]$, and its on-policy advantage is 0.068 with interval $[0.011, 0.131]$. Replacing teacher curvature with current student curvature changes the exact subset but retains residual weighting. This result shows that both categorical conventions can support a useful residual-aware design away from complete teacher agreement.

Pure Fisher E optimality targets another quantity. Teacher-Fisher E selection increased the mean realized minimum eigenvalue from 0.025 to 1.172, and student-Fisher E selection increased its counterpart from 0.045 to 1.459. Their fixed-prefix maximum-gap reductions were 0.142 and 0.104, compared with 0.279 for residual E selection. Conversely, teacher-Fisher residual selection improved heldout reduction over random by 0.0118, with paired interval $[0.0018, 0.0217]$. Worst-direction coverage, realized training-gap reduction, and heldout transfer are empirically distinct objectives, matching the horizon separation in Theorem B.14.

D.4 Query Number Interacts with Query Geometry

At $q = 1$, 158 of 160 fixed-prefix runs had negative absorption-time association, with mean -0.799. Teacher-Fisher residual and gradient norm selection improved maximum-gap reduction over random by 0.128 and 0.126, with intervals $[0.032, 0.259]$ and $[0.029, 0.259]$. Moving from one to four queries added 0.143 for gradient norm and 0.163 for Residual E, with intervals $[0.090, 0.201]$ and $[0.106, 0.225]$.

The one-query fixed-prefix matrix establishes that temporal slowdown does not require a multi-query mixture. One rollout already contains many token positions and therefore multiple Jacobian factors. Heterogeneous modes arise inside that rollout, and their fast components can be depleted before the late checkpoints. The stronger $q = 1$ mean association is consistent with a narrower design in which the weakest query-specific modes dominate sooner.

For fully on-policy $q = 1$ training, 109 of 160 runs had negative absorption-time association, with mean -0.190. At $q = 4$, Residual E improves maximum-gap reduction over random by 0.110, with interval $[0.055, 0.168]$. Moving from one to four queries adds 0.126 for Residual E and 0.110 for Gradient norm, with intervals $[0.075, 0.178]$ and $[0.059, 0.165]$.

These direct gains isolate complementarity. Each method is constructed from the same seed-specific candidate representations at both budgets. Increasing q gives Gradient norm and Residual E access to additional high-energy directions. Four selected training queries help most when the selection rule uses their joint geometry.

Nested random fixed-prefix reductions were 0.005, 0.074, and 0.206 at $q = 1$, $q = 4$, and $q = 16$; the final increment was 0.132, with interval $[0.064, 0.206]$. All sixteen fixed-prefix $q = 16$ runs had negative temporal association, with mean -0.604. Random selection eventually fills more directions, and its benefit becomes clear between four and sixteen queries.

Table 7 and Fig. 5 summarize the cardinality interaction. Under on-policy training, the $q = 4$ Residual E mean reduction of 0.235 exceeds $q = 16$ random by 0.078, with interval $[0.009, 0.152]$. This is a selected-set efficiency result at equal update count; the implemented screening and online teacher costs are reported independently.

D.5 Selected-Query Efficiency and Teacher Cost

Table 8 separates selected-query cardinality from teacher calls. In every query-count result, q denotes the number of unique selected training prompts in the update schedule. The implemented protocol screens 32 teacher-supervised candidates for every selector, including random, and uses eight shared warmup and sixteen shared heldout-target trajectories. Fully on-policy $q = 4$ and $q = 16$ runs both make 256 online teacher calls. Their actual protocol total is therefore 312 teacher trajectories per seed in both conditions.

Table 7: Maximum teacher-gap reduction across selected-query cardinalities. Each entry is the mean over sixteen training seeds, and Δ is the $q = 4$ value minus the $q = 1$ value.

	Fixed prefix, SGD			On policy, AdamW		
Query design	$q = 1$	$q = 4$	Δ	$q = 1$	$q = 4$	Δ
Random	0.005	0.074	0.069	0.082	0.125	0.043
Gradient norm	0.131	0.274	0.143	0.123	0.233	0.110
Residual E	0.116	0.279	0.163	0.109	0.235	0.126
Teacher Fisher E	0.098	0.142	0.044	0.112	0.146	0.033
Teacher Fisher residual	0.133	0.152	0.019	0.122	0.152	0.031
Student Fisher residual	0.107	0.204	0.097	0.104	0.192	0.089

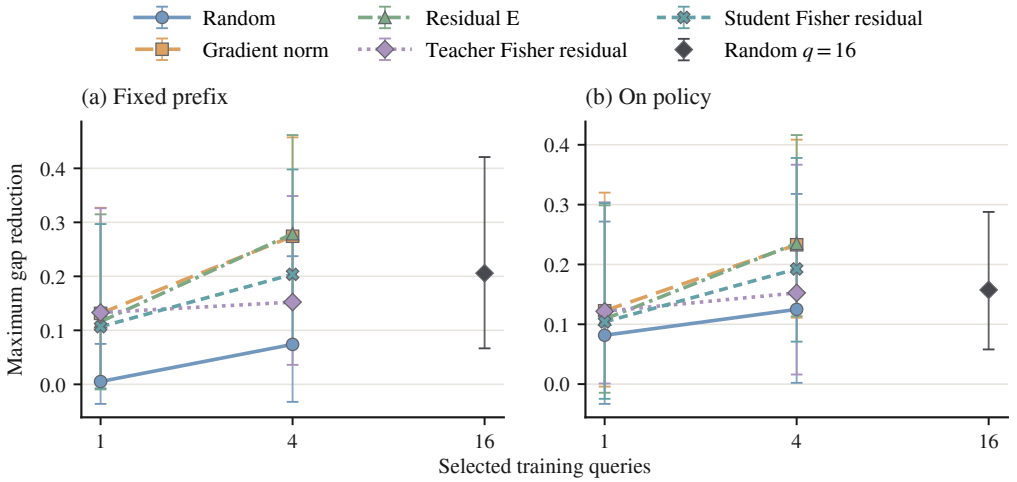


Fig. 5. Maximum teacher-gap reduction across selected-query cardinalities. Lines connect the $q = 1$ and $q = 4$ means for each principal design; diamonds show the nested $q = 16$ random condition. Error bars are training-seed bootstrap intervals. Residual E exhibits the strongest design by cardinality interaction in both update rules.

An end-to-end selector-specific accounting gives a different comparison. Residual E requires 32 screened trajectories before choosing four prompts. Random selection requires no teacher-supervised screen. Excluding shared warmup and evaluation, the minimal on-policy costs are therefore 288 teacher trajectories for q4 Residual E and 256 for q16 random, a 12.5 percent increase for the proposed selection procedure. In fixed-prefix training, the corresponding minimal costs are 32 screened trajectories and sixteen selected-target trajectories. The q4 versus q16 result is consequently a selected-set and update-count efficiency result, not a claim of four end-to-end teacher calls.

Token accounting reaches the same conclusion. The 32-candidate screen contains a mean of 2,038.6 supervised token positions. During on-policy training, q4 Residual E and q16 random use means of 16,377.3 and 16,370.4 online teacher-supervised positions. Their online token costs are effectively matched by the common 256-update schedule; the extra Residual E cost lies in candidate screening. Table 14 reports the complete seed intervals.

D.6 Unclipped Dynamics Preserve the Horizon Ordering

Table 9 reports the theory-aligned panel. With clipping removed and every realized step satisfying $\eta\lambda_{\max} < 0.637$, 142 of 160 runs had negative absorption-time association, with mean -0.446. Gradient norm and Residual E improved maximum-gap reduction over random by 0.161 and 0.162, with intervals $[0.098, 0.226]$ and $[0.100, 0.224]$. Initial gradient energy had mean within-seed rank association 0.259 with early absorption, with interval $[0.088, 0.408]$, matching Equation 65. Student-Fisher residual selection improved maximum-gap reduction by 0.086, with interval $[0.038, 0.143]$.

Table 8: Teacher-trajectory accounting for the selected-query comparison. Actual protocol totals include eight shared warmup trajectories, 32 candidate screens, sixteen heldout target trajectories, and online training calls. Minimal cost removes screening from the random selector and excludes shared warmup and evaluation.

Schedule	Selected	Actual screen	Online calls	Actual total	Minimal method cost
Residual E, fixed prefix	4	32	0	56	32
Random, fixed prefix	16	32	0	56	16
Residual E, on policy	4	32	256	312	288
Random, on policy	16	32	256	312	256

Table 9: Theory-aligned $q = 4$ fixed-prefix panel without gradient clipping. Values are means over sixteen training seeds.

Query design	Early absorption	Late absorption	Max reduction	ρ_t
Random	0.0010	0.0002	0.133	-0.286
Gradient norm	0.0082	0.0003	0.294	-0.580
Residual E	0.0131	0.0003	0.295	-0.608
Teacher Fisher E	0.0014	0.0003	0.169	-0.396
Teacher Fisher residual	0.0016	0.0003	0.174	-0.401
Student Fisher residual	0.0022	0.0003	0.219	-0.460

The unclipped panel strengthens the mechanistic connection in three ways. First, the temporal decline persists under the scalar SGD update analyzed by Theorem B.4. Second, the realized stability products remain below one, so the sign ordering of contraction factors used in the discrete proof applies to every sketched operator. Third, the initial gradient-energy association with early absorption is positive across selectors after ranking within seed. These checks connect the theorem directly to a neural local regime.

The pure Fisher criteria again optimize their declared spectra. Teacher-Fisher E has maximum-gap reduction 0.169 and raises its mean minimum teacher-Fisher eigenvalue to 1.172. Student-Fisher E has reduction 0.134 and raises its mean minimum student-Fisher eigenvalue to 1.459. Student-Fisher residual improves reduction to 0.219. The finite-horizon ranking therefore moves toward residual-weighted designs even when clipping is removed, consistent with the realized-gap term in Theorem B.14.

D.7 Horizon-Aware Selection Transfers to a Larger Student

Table 10 reports the 4B-student scale transfer. Gradient norm improved maximum-gap reduction over random by 0.038 under fixed prefixes and 0.043 on-policy, with intervals $[0.017, 0.060]$ and $[0.018, 0.067]$. Teacher-Fisher residual selection improved it by 0.023 and 0.025, with intervals $[0.002, 0.049]$ and $[0.007, 0.043]$. Absorption-time association was negative in 46 of 64 runs, with means -0.265 fixed and -0.137 on-policy, preserving both horizon-aware gains at scale.

The 4B student produces maximum-gap reductions of 0.041 for random selection, 0.080 and 0.084 for Gradient norm, and 0.064 and 0.066 for Teacher-Fisher residual under fixed and regenerated occupancy. The corresponding paired gains have the same sign under both occupancy rules, transferring the residual-gradient and teacher-curvature effects to the larger student.

D.8 The Starvation Signature Transfers from Mathematics to Science

Table 11 reports the 128-run ARC-Challenge matrix across a distinct reasoning domain. We report starvation strength $s = -\rho_t$, so positive values indicate progressively decreasing per-update absorption. Under fixed-prefix SGD, the four selector means range from 0.460 to 0.582. Under fully on-policy AdamW, they range from 0.307 to 0.478. All eight seed-bootstrap intervals are bounded away from zero; their lower endpoints range from 0.080 to 0.449.

Table 10: Scale transfer from a Qwen3.5 4B student to a 9B teacher over eight seeds. Reductions are initial minus final teacher gap after 128 updates, and ρ_t is the absorption-time rank association.

	Fixed prefix			On policy		
Query design	**Max red.**	**Held red.**	ρ_t	**Max red.**	**Held red.**	ρ_t
Random	0.041	0.006	-0.339	0.041	0.011	0.003
Gradient norm	0.080	0.007	-0.092	0.084	0.013	-0.280
Teacher Fisher residual	0.064	0.007	-0.313	0.066	0.013	-0.125

Table 11: Cross-domain temporal starvation on ARC-Challenge. Starvation strength is $s = -\rho_t$, where ρ_t is the rank association between checkpoint interval and per-update absorption. Entries are means and 95 percent seed-bootstrap intervals over sixteen seeds.

Query design	**Fixed prefix, SGD**	**On policy, AdamW**
Random	0.494 $[0.379, 0.613]$	0.420 $[0.207, 0.616]$
Gradient norm	0.582 $[0.449, 0.711]$	0.307 $[0.080, 0.526]$
Gradient E	0.460 $[0.348, 0.578]$	0.348 $[0.166, 0.540]$
Residual E	0.542 $[0.415, 0.664]$	0.478 $[0.344, 0.613]$

The result covers Random, Gradient norm, Gradient E, and Residual E under both occupancy rules. The same temporal phenomenon therefore appears in mathematical word problems and multiple-choice science questions, under static cached trajectories and student-regenerated trajectories. This cross-domain agreement supports spectral starvation as a property of heterogeneous OPD geometry rather than a GSM8K-specific prompt effect.

D.9 Occupancy Drift Alters the Frozen Spectrum Signature

Table 12 reports the matched AdamW occupancy comparison. Regeneration changed the trajectory on 97.9 to 99.8 percent of repeated query visits and shifted absorption-time association by 0.372 for random, 0.358 for gradient norm, and 0.402 for teacher-Fisher residual selection; their intervals were $[0.173, 0.529]$, $[0.209, 0.519]$, and $[0.200, 0.583]$. With optimizer and schedule fixed, these changes expose the operator drift in Theorem B.18.

The matched comparison changes the interpretation of the primary fixed-versus-on-policy curves. With AdamW on both sides, regeneration makes temporal association less negative for Random, Gradient norm, and Teacher-Fisher residual. This is the direction predicted when a changing A_t repeatedly injects new fast components or changes which residual is measured by the frozen reference.

Trajectory turnover is near one for all four selectors, so its between-run association with absolute endpoint deviation is close to zero. Table 13 resolves the geometry hidden by this binary event. At the exact archived update-256 adapters, current-occupancy student-Fisher spectral drift from the common start ranges from 0.984 to 0.989 across selectors. Replacing regenerated trajectories by the original prefixes at identical current parameters still leaves fixed-prefix drift between 0.950 and 0.965, while the isolated occupancy increment ranges from 0.975 to 1.007. The maximum principal angle is approximately 89 degrees for every selector. Parameter evolution changes the Jacobians, and occupancy replacement contributes another order-one perturbation after parameters are held fixed.

The measured replay trajectory shows that this replacement occurs early. Pooled current-occupancy student-Fisher spectral drift is 0.976 at update 64, 0.984 at update 128, and 0.986 at the exact archived endpoint. Every selected trajectory has changed by the first measured checkpoint. The operator path has therefore left the frozen local regime well before the final interval, directly instantiating the perturbation sequence in Proposition B.20.

Fig. 6 joins Random, Gradient norm, and Teacher-Fisher residual to their matched-AdamW fixed-prefix counterparts. Across the 48 runs, occupancy-induced gradient drift has rank association 0.334 with the absolute on-policy versus fixed-prefix late-absorption deviation, with seed-cluster interval $[0.028, 0.566]$. This relationship directly connects the operator perturbation in Proposition B.20 to departure from the frozen training path.

Table 12: Matched-optimizer $q = 4$ occupancy comparison over sixteen seeds. $\Delta\rho_t$ is on-policy regeneration minus fixed-prefix reuse under the same AdamW update rule. Turnover is the fraction of consecutive revisits whose generated trajectory hash changes.

Query design	$\boldsymbol{\Delta\rho_t}$	**Turnover**
Random	0.372	0.979
Gradient norm	0.358	0.998
Teacher Fisher residual	0.402	0.995

Table 13: Direct operator drift at update 256. Fixed/init., Current/init., and Occupancy/fixed report relative spectral-norm drift for the student-Fisher operator. Traj. is trajectory-change fraction. Angle is the maximum current-occupancy versus common-start principal angle. $\Delta\rho_t$ is on-policy minus matched fixed-prefix absorption-time association; NA marks the selector without a matched-optimizer fixed-prefix run.

Selector	**Traj.**	**Fixed/init.**	**Current/init.**	**Occupancy/fixed**	**Angle / $\Delta\rho_t$**
Random	1.000	0.960	0.984	0.986	89.3° / 0.372
Gradient norm	1.000	0.965	0.989	1.007	89.1° / 0.358
Residual E	1.000	0.964	0.986	0.986	89.1° / NA
Teacher Fisher residual	1.000	0.950	0.987	0.975	89.1° / 0.402

D.10 Empirical Hierarchy of the Predictions

The experiments support a hierarchy rather than one scalar score. The most stable prediction is temporal: fixed-operator absorption declines in 158 of 160 q1 runs, 158 of 160 q4 runs, all sixteen q16 runs, and 142 of 160 unclipped q4 runs; all eight ARC selector and mode conditions have positive starvation strength. The second prediction is local and quantitative: A_0 predicts the first sixteen updates, and measured tangent drift tracks its subsequent error. The third concerns immediate residual value: Gradient norm and Residual E improve finite-horizon maximum-gap reduction in the main, unclipped, and scale-transfer panels. The fourth concerns occupancy drift: matched regeneration attenuates the frozen temporal signature and produces order-one operator replacement.

These levels explain why different selectors dominate different geometric objectives. Gradient norm is strong when the horizon rewards present residual energy. Residual E becomes especially valuable when four selected training queries can combine complementary high-amplitude directions. Pure Fisher E succeeds on minimum-eigenvalue coverage and supplies the long-horizon reference. Teacher-Fisher residual transfers consistently to the 4B model pair. Each outcome corresponds to a different term in the theory.

The compute ledger also clarifies scale. The study contains 1,088 training runs, 64 early tangent measurements, and 64 direct on-policy operator measurements. Fixed-prefix 0.8B runs take about 397 to 433 accelerator seconds on average depending on panel, while fully on-policy runs take about 1,488 seconds because they regenerate student trajectories and teacher targets at every update. Scale-transfer fixed runs average 265 seconds for 128 updates and on-policy runs average 979 seconds. Teacher-forward accounting is reported separately because fixed-prefix reuse and on-policy regeneration have different supervision costs.

E Implications

E.1 Spectral Starvation Is a Residual-Composition Effect

The rate identities change how a slow OPD curve should be read. A decreasing fractional absorption rate does not imply that teacher supervision has disappeared. It can arise because the residual has changed composition. Strong directions are learned first, their coefficients shrink, and weak directions receive an increasing share of the remaining gap. The exact system makes this distinction visible: the uniform design begins with substantial absorption but approaches the rate of its smallest active eigenvalue, whereas the balanced design maintains nearly constant fractional absorption. The

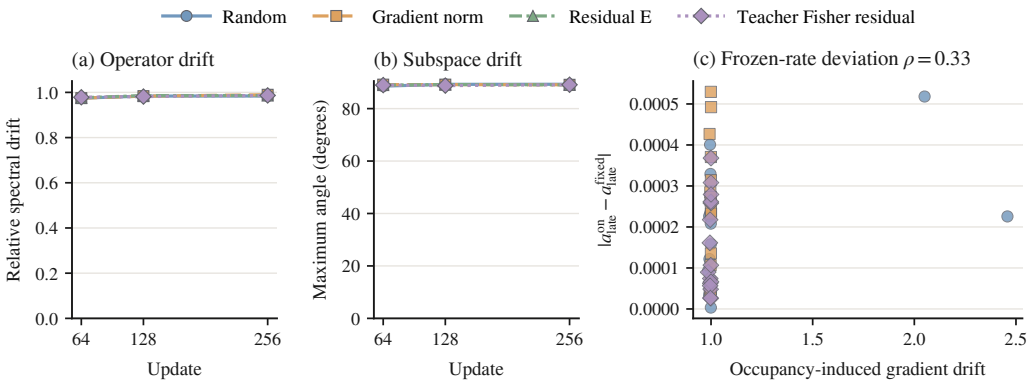


Fig. 6. Direct operator drift under changing occupancy. The first two panels use the full natural scales for relative drift and principal angle, showing that both quantities are already near their maximal reference values at update 64. The final panel joins the exact archived-adapter measurement to 48 matched AdamW runs. Occupancy-induced gradient drift has rank association 0.334 with absolute late-absorption deviation; the seed-cluster interval is $[0.028, 0.566]$.

neural fixed-prefix matrices show the same temporal ordering across almost every selector and seed. ARC-Challenge reproduces positive starvation strength for every tested selector under both fixed and regenerated trajectories, extending the signature from mathematics to science reasoning. Monitoring only total gap therefore loses the mechanism. A useful diagnostic reports the gap together with its active spectral mean, variance, smallest eigenvalue, and kernel component.

This interpretation also separates learning-rate tuning from query design. A scalar learning rate changes the effective time axis and can optimize the worst contraction over a fixed spectral interval, but it cannot remove the interval itself. A poorly conditioned query operator remains poorly conditioned after one scalar rescaling. Query selection changes the operator by adding, removing, or reweighting trajectory Jacobians. Preconditioning and adaptive optimizers can further change the coordinate geometry, yet the matched AdamW experiment shows that occupancy regeneration still changes the absorption signature when the optimizer is held fixed.

E.2 The Frozen Spectrum Has a Measured Validity Window

The initial student-Fisher operator is a quantitative local forecast. It explains the next sixteen updates with pooled $R^2 = 0.906$ and rank correlation 0.925 across 160 neural runs. This calibration turns the spectral account into a prediction of future teacher-gap reduction rather than a retrospective description of a declining curve. The forecast uses only the common-checkpoint gradients, curvature factors, and declared query schedule.

The matched tangent replay identifies where that local description expires. Relative spectral drift rises steadily through update 64, while absolute forecast error rises with it. Their pooled rank correlation is 0.504 with seed-cluster interval $[0.392, 0.623]$, and the seven checkpoint means have rank correlation 0.929. The resulting claim has two coupled parts: local spectral heterogeneity predicts the starvation signature, and measured operator drift predicts the breakdown of the frozen forecast.

E.3 Query Count Is an Incomplete Budget Description

One teacher query can generate many supervised token positions, so prompt count and supervision count already differ in OPD. The present results add a third quantity: the dimension and conditioning of the induced tangent design. At $q = 1$, one rollout contains enough heterogeneous token factors to produce spectral starvation in nearly every fixed-prefix run. Increasing to four selected training queries reliably improves Gradient norm and Residual E because these selectors combine additional relevant directions. Increasing a nested random design from four to sixteen selected queries expands fixed-prefix coverage, while the four-query Residual E schedule exceeds sixteen-query random selection under changing occupancy at equal online update count.

The operational consequence is that selected-set efficiency should be reported as an operator comparison at equal weight, not as performance divided by prompt count alone. Two prompt sets can have equal size, sequence length, teacher-token budget, and update count while differing in kernel dimension and smallest eigenvalue. A budget-aware selector should ask what new direction a candidate contributes to the existing set and whether the current teacher residual has energy in that direction. Proposition B.10 gives the qualitative test, while Equation 50 gives the quantitative consequence.

E.4 The Training Horizon Selects the Appropriate Geometry

The horizon theorem organizes the empirical ordering of selectors. Gradient norm measures immediate residual energy and is strong over the finite neural horizon. Residual E retains this amplitude while rewarding complementarity, which becomes valuable when the selected set grows from one to four training queries. Pure Fisher E designs increase the minimum eigenvalue of their declared curvature operator by more than an order of magnitude. These quantities distinguish protection of an unknown worst direction from reduction of the residual present at a finite time.

The appropriate query objective can therefore change during training. Early in optimization, the residual contains broad high-energy components and a gradient-energy rule can produce rapid progress. After those modes have been absorbed, the active spectral measure shifts toward weak directions and an E optimal criterion becomes increasingly relevant. A practical scheduler can estimate both $u^\top A_q u$ and the marginal change in the smallest active eigenvalue, then allocate queries according to the remaining horizon. The equal-trace reversal shows that no horizon-independent scalar ranking can recover both endpoints in general.

Residual-weighted E selection is one bridge between these endpoints. It does not require an explicit estimate of the eventual worst direction; the current gradient magnitudes weight the candidate factors, while the minimum eigenvalue rewards a complementary set. Its performance in the main and unclipped panels suggests that this compromise is useful when only one design is selected at the common checkpoint. Repeated redesign later in training would directly follow the moving spectral measure characterized by Theorem B.4.

E.5 Changing Occupancy Is a Structured Perturbation

Fully on-policy training changes the data distribution at the same time that it changes the model. The matched panel shows that this is not a small textual effect: regenerated trajectories differ on approximately 98 to 100 percent of repeated visits. With AdamW, learning rate, clipping, query subset, and update count matched, regeneration makes the absorption-time association less negative. New trajectories can reintroduce strong directions, remove previously occupied directions, or alter the teacher residual evaluated at a token position. These mechanisms break the fixed-operator modal ordering without erasing the aggregate tendency toward slower absorption.

Trajectory turnover and operator turnover play different roles. A hash change is a direct occupancy event, but it saturates in this experiment and has little power to rank endpoint deviations. The curvature-weighted Jacobian measurements show what the hashes omit. By update 64, current-occupancy student-Fisher spectral drift is already 0.976 and the maximum principal angle is approximately 89 degrees. At the exact archived endpoint, replacing regenerated trajectories by fixed prefixes at the same student parameters changes the operator by approximately one reference spectral norm. The textual replacement therefore introduces a different tangent design rather than a locally equivalent prefix collection.

The same measurement connects occupancy replacement to departure from the frozen path. Occupancy-induced gradient drift is associated with absolute late-rate deviation, with rank correlation 0.334 and seed-cluster interval $[0.028, 0.566]$. Proposition B.20 matches this structure: accumulated perturbations control the size of departure, while operator direction and residual alignment determine the evolving contraction.

E.6 Teacher and Student Curvature Address Different Local Questions

Teacher Fisher is the equality-limit geometry shared by smooth f-divergences. Student Fisher is the Gauss-Newton part of the exact cross-entropy Hessian at the current checkpoint. Their residual-

weighted versions perform differently across scale. Both are useful in the 0.8B main matrix, while teacher-Fisher residual transfers more consistently to the 4B student. This pattern indicates that the current student categorical curvature can emphasize directions that are locally strong for the student's own probabilities but less aligned with the teacher-gap reduction measured on the design pool.

The exact Hessian decomposition identifies the remaining term: a teacher-gap-weighted sum of network logit Hessians. As teacher and student approach agreement, this term vanishes and the two Fisher conventions approach the same categorical geometry. Earlier in training, its magnitude and sign can alter local curvature. Comparing teacher and student probes is therefore a diagnostic of where the current checkpoint lies between the global nonlinear regime and the shared local limit.

E.7 Operational Consequences

The spectral identities determine when OPD will exhibit diminishing fractional returns, which components are unreachable under the current trajectories, and why increasing selected prompt count can fail to improve conditioning. These diagnostics can reveal whether further updates repeat exhausted directions and whether a refreshed query design should target residual energy or the weakest active direction.

Cost accounting sharpens the same scope. Four selected training queries outperform sixteen selected random queries on maximum teacher-gap absorption at equal online update count. Constructing the spectral selector uses a 32-candidate teacher-supervised screen, giving minimal on-policy costs of 288 versus 256 teacher trajectories. This is selected-set efficiency with explicit screening cost and motivates amortized candidate representations across training horizons.

E.8 A Spectral Workflow for OPD

The theory and experiments suggest a concrete workflow. First, declare the trainable coordinates, candidate query pool, rollout policy, token weighting, and evaluation gap. Second, generate candidate trajectories at a common checkpoint and sketch their gradient or Fisher factors. Third, report kernel dimension, minimum active eigenvalue, residual energy, and spectral variance rather than query count alone. Fourth, select queries according to the available horizon, using gradient energy for immediate progress, E optimality for persistent weakest-direction coverage, or residual-weighted E selection between them. Fifth, refresh the measurement when trajectory or operator turnover indicates that the frozen design has become stale.

This workflow separates four decisions that are often merged: where the student visits, which tangent directions those visits excite, where the current teacher residual lies, and how long the selected design will be optimized. Each decision has an observable statistic and a corresponding theorem. The resulting analysis turns slow OPD absorption into a conditioning problem with separately reported screening, selected-set, online teacher, and token-position budgets.

F Additional Measurements

F.1 Teacher-Supervised Token Accounting

Table 14 reports sequence-length-aware costs. One teacher trajectory can contain from one to 64 supervised positions, so trajectory count and token-position count are separate budgets. The online q4 Residual E and q16 random schedules have nearly identical token counts, while screening contributes approximately 2,039 additional positions to the geometry-aware selector.

F.2 Direct Operator-Drift Checkpoints

Table 15 reports the student-Fisher operator measurement at every declared checkpoint. Updates 64 and 128 are states on the deterministic replay trajectory. Update 256 is measured after restoring the hash-verified archived adapter from the original training run. The replay-to-archive relative parameter distance averages 0.105 to 0.106 across the four selectors and is retained as a reproducibility diagnostic rather than used for the terminal association.

Table 14: Teacher-supervised token-position accounting over sixteen seeds. Each entry reports the mean and 95 percent seed-bootstrap interval.

Cost component	Mean positions	95 percent interval
Shared warmup	488.1	$[474.1, 501.0]$
32-candidate screen	2038.6	$[2021.1, 2047.8]$
Shared heldout targets	1020.6	$[1013.9, 1024.0]$
16 selected random targets	1019.0	$[1009.4, 1024.0]$
Residual E online training	16377.3	$[16364.4, 16384.0]$
Random online training	16370.4	$[16360.3, 16378.3]$

Table 15: Student-Fisher operator drift over checkpoints. Values pool the four selectors and sixteen seeds.

Checkpoint	Trajectory change	Fixed vs. initial	Current vs. initial	Occupancy increment
64	1.000	0.943	0.976	0.985
128	1.000	0.946	0.984	0.980
256	1.000	0.960	0.986	0.988

The current-occupancy operator is already separated from the common-start operator by update 64. Fixed-prefix drift at the same checkpoint is also large, showing that parameter evolution changes the Jacobian before occupancy replacement is introduced. The occupancy-increment column compares regenerated and fixed prefixes at identical current parameters and remains near one across the measured horizon.

F.3 Measured Compute

Table 16 decomposes accelerator-process use by training panel and mode. On-policy updates regenerate a student trajectory and evaluate the teacher at every update. Fixed-prefix updates reuse the precomputed sequence and teacher target.

Teacher-forward counts record online calls made during the training updates. Process hours sum elapsed time across concurrent accelerator processes and measure total device occupation rather than end-to-end wall time.

G Proofs

G.1 Proof of Theorem B.1

Proof. Use the convention $D_f(p, q) = \sum_a q_a f(p_a/q_a)$ with $f(1) = 0$. Let $q = p + \delta$ with $\mathbf{1}^\top \delta = 0$. A second order expansion at $q = p$ gives

$$D_f(p, p + \delta) = \frac{f''(1)}{2} \sum_a \frac{\delta_a^2}{p_a} + o(\|\delta\|_2^2). \tag{75}$$

The constant and first order terms vanish on the probability simplex. For a centered logit perturbation h, the probability perturbation is $\delta = C_p h + o(\|h\|_2)$, where $C_p = \mathrm{Diag}(p) - pp^\top$. Direct multiplication shows

$$C_p \,\mathrm{Diag}(p)^{-1} C_p = C_p. \tag{76}$$

Thus the quadratic form in Equation 75, restricted to probability perturbations induced by logits, reduces exactly to the categorical Fisher form. Substituting Equation 76 into Equation 75 yields the logit Hessian $f''(1)C_p$. Hence, up to the scalar $f''(1)$, every admissible smooth f-divergence induces the same local categorical Fisher geometry at agreement.

At teacher agreement, the first derivative with respect to the logits is zero, so the second derivative of the network map contributes no residual term. A parameter perturbation $\Delta\vartheta$ induces the first-order logit perturbation $h = J_s \Delta\vartheta$, and the second-order chain rule therefore pulls the logit Hessian back to

Table 16: Measured compute by panel. Process hours sum wall time over accelerator processes; teacher forwards count online teacher calls during updates.

Panel and mode	Runs	Mean seconds	Process hours	Teacher forwards
arc q4/fixed	64	622.7	11.07	0
arc q4/on policy	64	1672.3	29.73	16,384
q1/fixed	160	397.0	17.64	0
q1/on policy	160	1486.4	66.06	40,960
q16/fixed	16	399.7	1.78	0
q16/on policy	16	1493.1	6.64	4,096
q4 fixed adamw/fixed	64	408.9	7.27	0
q4 noclip/fixed	160	433.2	19.25	0
q4/fixed	160	398.9	17.73	0
q4/on policy	160	1488.0	66.14	40,960
scale q4/fixed	32	265.2	2.36	0
scale q4/on policy	32	979.4	8.71	4,096
Operator-drift measurement	64	1638.3	29.12	17,408
Early tangent measurement	64	346.4	6.16	0

$f''(1)J_s^\top C_p J_s$ at prefix s. Setting $p = \pi_{\mathrm{T}}(\cdot \mid s)$ and averaging under the occupied prefix distribution gives Equation 36. □

G.2 Proof of Proposition B.3

Proof. At the agreement point, $\nabla\mathcal{L}_\rho(\theta_\star;\theta) = 0$ and $H_\rho = \nabla^2\mathcal{L}_\rho(\theta_\star;\theta)$. The fundamental theorem of calculus along the segment gives

$$\nabla\mathcal{L}_\rho(\theta_\star+\delta;\theta) - H_\rho\delta = \int_0^1 \left[\nabla^2\mathcal{L}_\rho(\theta_\star+s\delta;\theta) - H_\rho\right]\delta\,\mathrm{d}s. \tag{77}$$

The Lipschitz assumption bounds the integrand norm by $Ms\|\delta\|_2^2$. Integrating s from zero to one yields Equation 38. □

G.3 Proof of Proposition B.2

Proof. The logit gradient of categorical cross entropy is $p_\vartheta(s) - p_{\mathrm{T}}(s)$, and its logit Hessian is $C_{p_\vartheta(s)}$. Applying the second order chain rule through $z_\vartheta(s)$ gives the Gauss Newton pullback plus one network Hessian for each logit, weighted by the corresponding first derivative. This is Equation 37. □

G.4 Proof of Theorem B.4

Proof. Diagonalize $A = V\,\mathrm{Diag}(\lambda_i)V^\top$. Repeated application of Equation 41 gives

$$u_t = \sum_i c_i(1-\eta\lambda_i)^t v_i, \tag{78}$$

which proves Equation 43. Define $b_i = c_i^2$ and $r_i = (1-\eta\lambda_i)^2$. Under the step size assumption, every r_i lies in $[0,1]$ and decreases with λ_i. Let $S_t = \sum_i b_i r_i^t$. Cauchy Schwarz applied to the vectors with coordinates $\sqrt{b_i}r_i^{t/2}$ and $\sqrt{b_i}r_i^{(t+2)/2}$ gives

$$S_{t+1}^2 \le S_t S_{t+2}. \tag{79}$$

Hence S_{t+1}/S_t is nondecreasing, and $a_t = 1 - S_{t+1}/S_t$ is nonincreasing. At any time with $S_{t+1} > 0$, equality occurs exactly when r_i is constant on the support of the time indexed spectral measure. Finally, the largest active r_i dominates the ratio S_{t+1}/S_t as t tends to infinity. It corresponds to the smallest active eigenvalue, which gives Equation 44.

For the discrete identity, $S_{t+1}/S_t = \mathbb{E}_{\nu_t}[r]$ and

$$\frac{S_{t+2}}{S_{t+1}} - \frac{S_{t+1}}{S_t} = \frac{\mathbb{E}_{\nu_t}[r^2]}{\mathbb{E}_{\nu_t}[r]} - \mathbb{E}_{\nu_t}[r] = \frac{\mathrm{Var}_{\nu_t}(r)}{\mathbb{E}_{\nu_t}[r]}. \tag{80}$$

The left hand side is $a_t - a_{t+1}$, proving Equation 45. □

For completeness, the continuous time identity follows from

$$E(t) = \sum_i c_i^2 \exp(-2\lambda_i t). \tag{81}$$

Differentiating $-\log E(t)$ once gives twice the mean of λ under ν_t. Differentiating again gives minus four times its variance, proving Equation 46.

G.5 Proof of Proposition B.5

Proof. With two active eigenvectors, Equation 43 becomes

$$\|u_t\|_2^2 = c_1^2 r_1^t + c_2^2 r_2^t. \tag{82}$$

Substitution into the definition of a_t gives Equation 47. Equality of the two modal energies requires $c_1^2 r_1^t = c_2^2 r_2^t$. Taking logarithms and using $r_2 > r_1$ gives Equation 48. After the crossover, the ratio of fast to slow energy is

$$\frac{c_1^2 r_1^t}{c_2^2 r_2^t} = \frac{c_1^2}{c_2^2} \left(\frac{r_1}{r_2} \right)^t, \tag{83}$$

which converges to zero. The absorption therefore approaches $1 - r_2$, the rate associated with the smaller eigenvalue. □

G.6 Proof of Corollary B.6

Proof. Decompose u_0 into the kernel, the μ eigenspace, and the remaining positive eigenspaces. The kernel energy is invariant. Every positive eigenvalue λ_i satisfies $\lambda_i \geq \mu$, and the step condition makes $0 \leq 1 - \eta\lambda_i \leq 1 - \eta\mu$. Hence

$$\sum_{i:\lambda_i>0} c_i^2 (1 - \eta\lambda_i)^{2t} \leq (1 - \eta\mu)^{2t} \sum_{i:\lambda_i>0} c_i^2. \tag{84}$$

Keeping only the coefficients in the μ eigenspace gives the lower bound. Orthogonality identifies the two coefficient sums with the projected norms in Equation 49. □

G.7 Proof of Corollary B.7

Proof. For a positive definite symmetric operator, the worst case norm contraction of $I - \eta A$ is the largest absolute value of $1 - \eta\lambda$ over the spectrum. On the interval $[\mu, L]$, the minimizing step equalizes the two endpoint magnitudes:

$$1 - \eta\mu = -(1 - \eta L). \tag{85}$$

Solving gives $\eta^\star = 2/(L + \mu)$ and substituting it at either endpoint gives Equation 51. □

G.8 Proof of Proposition B.8

Proof. For $0 < \eta < 2/L$ and every positive eigenvalue $\lambda_i \leq L$, we have $-1 < 1 - \eta\lambda_i < 1$, hence $0 \leq r_i < 1$. Kernel directions have $r_i = 1$. The proof of Theorem B.4 uses only nonnegativity of the r_i to establish log-convexity of $S_t = \sum_i c_i^2 r_i^t$, so the absorption sequence remains nonincreasing. The largest active r_i dominates S_{t+1}/S_t and gives Equation 52. If $\eta \leq 1/L$, the quantity $1 - \eta\lambda$ is nonnegative and decreases with λ, so the largest r_i corresponds to the smallest active eigenvalue. Above $1/L$, absolute contraction can increase again as $\eta\lambda$ approaches two. □

G.9 Proof of Proposition B.9

Proof. Because $A(w)$ is symmetric, its kernel and range are orthogonal invariant subspaces. Applying the kernel projector to one update gives

$$P_{\ker A(w)} u_{t+1} = P_{\ker A(w)} u_t - \eta P_{\ker A(w)} A(w) u_t = P_{\ker A(w)} u_t. \tag{86}$$

Induction makes this component equal to $P_{\ker A(w)} u_0$ for every t. Orthogonality then yields Equation 54. □

G.10 Proof of Proposition B.10

Proof. Every vector in the intersection of the supported query kernels is plainly in $\ker A(w)$. For the reverse inclusion, let $x \in \ker A(w)$. Positive semidefiniteness gives

$$0 = x^\top A(w)x = \sum_{q \in \mathcal{I}(w)} w_q x^\top A_q x. \tag{87}$$

Every summand is nonnegative and every supported weight is positive, so $x^\top A_q x = 0$ for each $q \in \mathcal{I}(w)$. For a positive semidefinite matrix, this equality is equivalent to $A_q^{1/2} x = 0$ and hence $A_q x = 0$. Thus x lies in every supported kernel. □

G.11 Proof of Proposition B.11

Proof. Let the copy weights be $w_{j,1}, \ldots, w_{j,m}$ with $\sum_k w_{j,k} = w_j$. Their contribution to the design operator is

$$\sum_{k=1}^{m} w_{j,k} A_j = w_j A_j. \tag{88}$$

Every other query contribution is unchanged, so the complete operator is identical. Kernel, spectrum, and the recurrence in Equation 41 are functions of this operator and therefore remain identical. □

G.12 Proof of Theorem B.12

Proof. On an invariant subspace U, diagonalize the restriction of $A(w)$. For a unit initial vector with coefficients c_i in this basis,

$$\|u_t\|_2^2 = \sum_i c_i^2 \exp(-2t\lambda_i). \tag{89}$$

The supremum over unit vectors puts all mass on the smallest restricted eigenvalue and equals the right hand side of Equation 59. Maximizing that eigenvalue therefore minimizes the worst case residual. The semidefinite constraint in Equation 58 is equivalent to $\lambda_{\min}(A(w)|_U) \geq \gamma$, which completes the claim. □

G.13 Proof of Proposition B.13

Proof. The Rayleigh characterization on U gives

$$\lambda_{\min}(A(w)|_U) = \inf_{\substack{x \in U \\ \|x\|_2 = 1}} \sum_q w_q x^\top A_q x. \tag{90}$$

For each fixed unit vector x, the expression is linear in w. Hence its pointwise infimum over unit vectors in U is concave, proving concavity of $w \mapsto \lambda_{\min}(A(w)|_U)$. Moreover, $P_U A(w) P_U \succeq \gamma P_U$ holds exactly when every unit $x \in U$ satisfies $x^\top A(w)x \geq \gamma$. Thus the semidefinite constraint in Equation 58 is equivalent to requiring every Rayleigh quotient on U to be at least γ. □

G.14 Proof of Theorem B.14

Proof. The solution of the symmetric linear flow is $u(\tau) = \exp(-\tau A(w))u_0$. Squaring its norm gives Equation 64. Differentiating the matrix exponential at zero yields Equation 65, so the initial decrease is determined by the realized quadratic energy $u_0^\top A(w)u_0$.

For the long-horizon statement, diagonalize $A(w)$ and subtract the invariant kernel energy. The remaining expression is

$$\sum_{i:\lambda_i > 0} c_i^2 \exp(-2\tau\lambda_i). \tag{91}$$

Among terms with $c_i \neq 0$, the one with the smallest positive eigenvalue decays slowest and therefore dominates as $\tau \to \infty$. Taking $-(2\tau)^{-1}$ times the logarithm then gives Equation 66. □

G.15 Verification of Proposition B.15

For the declared u_0, the initial energy derivative under A_1 is $-2(1.72)$ and under A_2 is -2. Thus A_1 gives the larger infinitesimal decrease. Both operators are positive definite and have trace two. Their smallest eigenvalues are 0.1 and 1, so Theorem B.14 gives asymptotic rates 0.1 and 1. Continuity of the two residual curves and their opposite short- and long-horizon ordering implies at least one finite reversal. Direct evaluation of the closed forms locates the first crossing at approximately 1.221, as recorded by the theorem certificate.

G.16 Proof of Proposition B.16

Proof. For the maximizing class,

$$z_{a^\star} + \delta_{a^\star} \geq z_{a^\star} - \|\delta\|_\infty. \tag{92}$$

Every competing class satisfies

$$z_a + \delta_a \leq \max_{b \neq a^\star} z_b + \|\delta\|_\infty. \tag{93}$$

Subtracting the second bound from the first gives a positive difference whenever Equation 69 holds. The maximizing class is therefore unchanged. □

G.17 Proof of Proposition B.17

Proof. Binary teacher cross entropy as a function of the gold log odds is

$$\ell_{\mathrm{T}}(\ell) = -\tau \log \sigma(\ell) - (1-\tau)\log(1-\sigma(\ell)). \tag{94}$$

Its derivative is $\sigma(\ell) - \tau = p - \tau$. Gradient descent gives $\ell^+ = \ell - \eta(p-\tau)$, which is Equation 70. The sign statements follow directly. □

G.18 Proof of Theorem B.18

Proof. Every eigenvalue of $I - \eta A_t$ on the declared subspace lies in $[0, 1-\eta\mu]$, so

$$\|u_{t+1}\|_2 \leq (1-\eta\mu)\|u_t\|_2 + \varepsilon_t. \tag{95}$$

Iterating this scalar recurrence gives Equation 72. □

G.19 Proof of Proposition B.19

Proof. Real symmetric commuting matrices admit a common orthonormal eigenbasis. In that basis, the update at time t multiplies coefficient i by $1 - \eta\lambda_{i,t}$. After T updates,

$$u_T = \sum_i c_i \left[\prod_{t=0}^{T-1}(1-\eta\lambda_{i,t})\right] v_i. \tag{96}$$

Orthogonality removes cross terms when the squared norm is taken, giving Equation 73. □

G.20 Proof of Proposition B.20

Proof. Let $e_t = u_t - \bar{u}_t$. Subtracting the frozen recurrence from the changing-operator recurrence gives

$$e_{t+1} = (I - \eta A)e_t - \eta(A_t - A)u_t + r_t. \tag{97}$$

The reference map has operator norm at most one, so

$$\|e_{t+1}\|_2 \leq \|e_t\|_2 + \eta\|A_t - A\|_{\mathrm{op}}\|u_t\|_2 + \|r_t\|_2. \tag{98}$$

Because $e_0 = 0$, iterating this inequality proves Equation 74. If $\|I - \eta A\|_{\mathrm{op}} = c < 1$, retaining the factor c at each recursion weights the summand from time t by c^{T-1-t}. □