WHEN CAN A ZERO BE DELETED? EXACT STATE COMPRESSION FOR GATED LOW-RANK TRAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

A low-rank component can leave the predictor while still determining the optimizer's next accepted update. We solve the resulting state-elimination problem for anchored gated low-rank training with one shared Euclidean ball. A current-checkpoint certificate keeps selected gates at zero over a declared horizon; one evolving radial statistic then replaces their factor directions in every projection, complete objective, and line-search decision, including rejected trials. An accepted-update family establishes why this radial information cannot all be discarded. The same construction separates *when memory is reclaimed* from *which learning rule is executed*: immediate, staged, and delayed release under one license preserve the same real training observations. A unified loss budget supplies computable certificates; a sharp cross-entropy bound admits 35 rather than 28 of 118 zero candidates in nonlinear confirmation. In a quality-first shared-backbone study, eight independently initialized adapters all admit release, removing 18 of 77 learned zero slots and 10.43% of their aggregate live state. All three release schedules match the full reference's trial signatures over 200 accepted blocks per adapter. Regression continuations retain 49.9–71.1% live-state savings, and directed rounding validates a complete nonlinear continuation. These results turn structural inactivity into a storage decision without adding a new real update rule.

1 INTRODUCTION

A sparse predictor need not have a small training state. When a low-rank gate reaches zero, its prediction term vanishes, but its factors may still consume a shared constraint and contribute regularization energy. Deleting their arrays can change the next accepted update. Conversely, keeping every such array until training ends pays for directions the predictor has stopped using. The missing interface is a constructive answer to *what may disappear while the chosen learning process remains unchanged*.

Low-rank adaptation and adaptive rank methods select economical representations (Hu et al., 2022; Zhang et al., 2023; Ding et al., 2023). We address the execution that follows that selection. The distinction matters when several adapters share a frozen backbone: memory can be reclaimed immediately, in batches, or when a system has room to rebuild arrays. These systems choices should not become additional learning hyperparameters. Our result gives that separation for an explicit drift-controlled continuation problem.

Three obstacles, one construction. Equal predictions need not give equal projected candidates. Equal candidates need not pass the same complete acceptance test. A gate that is zero now can revive. We resolve all three together: a checkpoint certificate establishes the future zero face, and a radial state carries the removed directions' entire required influence through the shared ball and original controller. The output is a smaller *executable* state, not just a smaller prediction formula.

Contributions. We establish a certificate-authorized exact reduction, its necessary radial information, and release-schedule independence; a unified future-loss budget with sharp nonlinear bounds and a largest certifiable set; and learning, numerical, and ownership evidence for the same construction. Our independent adapter workload tests actual continued learning on shared data and backbone, while schedule experiments isolate the storage decision. Deterministic state savings and measured execution costs are reported separately.

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.

2 THE CONTINUATION PROBLEM

A slot and a zero. A frozen matrix $W_0 \in \mathbb{R}^{d_o \times d_i}$ receives a gated low-rank update

$$W = W_0 + sB\,\mathrm{Diag}(\gamma)A, \quad s = \alpha/r_0, \quad A \in \mathbb{R}^{r_0 \times d_i},\ B \in \mathbb{R}^{d_o \times r_0}. \tag{1}$$

Gate $\gamma_j \in [0,1]$ weights the rank-one term $s\gamma_j b_j a_j^\top$. We call this triple a *slot*; the number of active slots bounds, but need not equal, matrix rank. The original scale s stays fixed after deletion.

Concatenate factors when adapting several layers. After a training-only warm stage, fix an anchor Θ_w and optimize

$$F(\Theta,\gamma) = \mathcal{L}(\Theta,\gamma) + \frac{\mu}{2}\|\Theta - \Theta_w\|^2 + \lambda^\top\gamma, \qquad \|\Theta - \Theta_w\| \le R, \quad \gamma \in [0,1]^{r_0}. \tag{2}$$

Here $\mathcal{L} \ge 0$, $\mu > 0$, and $\lambda \ge 0$. The anchor penalizes drift from the warm state; the shared ball places a hard budget on that drift. Appendix B gives the nonnegative MCP/cost extension used by the frozen-feature classification experiment.

Reference updates and time. Let $f = F - \lambda^\top\gamma$. A *sweep* updates all factors jointly and then all gates. A factor trial at inverse step $M > 0$ is $\Theta^+ = \Pi_{\|\cdot-\Theta_w\|\le R}(\Theta - \nabla_\Theta f/M)$. A gate centered at x solves

$$z^* = \arg\min_{z\in[0,1]}\{qz + MD_h(z,x)\}, \quad q = \partial_\gamma f + \lambda, \quad D_h(z,x) = h(z) - h(x) - h'(x)(z-x). \tag{3}$$

With displacement d, the factor distance is $D_\Theta = \|d\|^2/2$; the gate distance sums D_h. Either block accepts exactly when

$$f^+ \le f + \langle \nabla f, d\rangle + (1-\nu)MD + a_{\mathrm{sl}}\max(1,|f|), \quad 0 < \nu < 1. \tag{4}$$

A rejection increases M without committing the candidate. H counts accepted factor or gate blocks: one completed sweep consumes two, a rejection consumes none. No trial for block $H+1$ is authorized by an H-block certificate.

Assumption 1 (Execution class)**.** *The factor anchor is isotropic and the factor constraint is one joint Euclidean ball. Gradients and the separable prox are exact in real arithmetic; h is strictly convex with finite endpoint derivatives. Accepted factor trials have $M \ge \mu$. Original scales, anchors, oracle/data and tie rules are fixed. The controller retains inverse steps, phase, counters and stopping budget, and uses only retained variables and reconstructible observations. Raw removed-gate gradients are not among its inputs.*

A zero that cannot simply be erased. Take retained coordinate $u = 0$, zero anchors, $R = 1$, $\mu = 1/2$ and $\mathcal{L}(u) = 2 - 2u \ge 0$ on the ball. Removed coordinates $v = 0$ and $v = 1/2$ have the same prediction-related loss. At $M = 1$ their unprojected points are $(2,0)$ and $(2,1/4)$, giving retained projections 1 and $8/\sqrt{65} \simeq 0.992278$. Both are accepted at $\nu = 1/4$, since the exact anchor Taylor remainder is $\mu D_\Theta \le (1-\nu)MD_\Theta$. Survivor-only state cannot reproduce both next accepted updates.

3 WHAT A ZERO MUST LEAVE BEHIND

Split factors into retained u and removed v, with anchors (a,b). On the face where the selected gates are zero, $f = \bar{f}(u,\gamma_S) + \mu\|v-b\|^2/2$. The reduced state keeps retained factors, their anchors, surviving gates, controller fields and metadata, together with

$$\zeta = \|v - b\|^2. \tag{5}$$

The discarded arrays are v and b. View their residual as one virtual coordinate $r = \sqrt{\zeta}$ with zero anchor: it participates in the constraint and quadratic penalty, not the predictor. Its *trial* value may be signed; the stored state remains ζ. Metadata retain original slot order, scales and the absolute license endpoint. Oracle/data and cold recovery are separate.

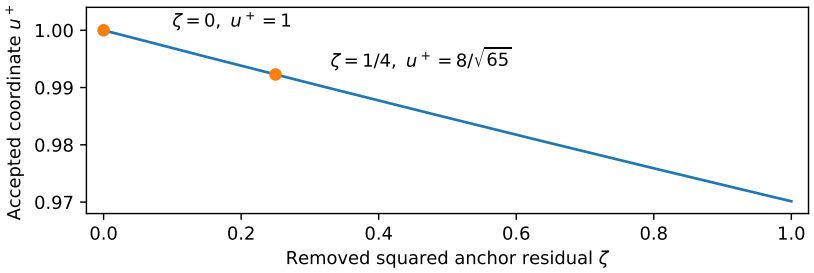

Figure 1: The same predictor, a continuum of different accepted updates. For the feasible family $v^2 = \zeta \in [0, 1]$, the next retained coordinate is strictly decreasing in the removed energy. All displayed trials are accepted; the labeled points recover the two-state obstruction.

Every trial closes radially. Let $g_u = \nabla_u \bar{f}$ include the retained anchor gradient. For any $M > 0$,

$$c_M = 1 - \mu/M, \quad p = u - g_u/M, \quad \eta = \min\{1, R/(\|p - a\|^2 + c_M^2 \zeta)^{1/2}\},$$
$$\rho = \eta c_M, \qquad u^+ = a + \eta(p - a), \qquad \zeta^+ = \rho^2 \zeta. \tag{6}$$

Set $\eta = 1$ at a zero denominator. The complete comparison quantities are

$$f^+ = \bar{f}(u^+, \gamma_S) + \tfrac{\mu}{2}\zeta^+, \qquad f = \bar{f}(u, \gamma_S) + \tfrac{\mu}{2}\zeta,$$
$$2D_\Theta = \|u^+ - u\|^2 + (\rho - 1)^2 \zeta, \qquad \langle \nabla f, d \rangle = \langle g_u, u^+ - u \rangle + \mu(\rho - 1)\zeta. \tag{7}$$

Indeed, the removed loss gradient is zero, so its unprojected anchor residual is $c_M(v - b)$. Joint projection yields $v^+ - b = \rho(v - b)$, and subtraction gives $v^+ - v = (\rho - 1)(v - b)$. Norms and the gradient inner product give Eq. (7). This holds also for rejected trials with $c_M < 0$.

The certificate interface. At the current checkpoint, a certificate for a zero set C and horizon H must give a computable bound $G_j(C, H)$ on $-\partial_{\gamma_j} \mathcal{L}$ at every reachable gate center *while the prefix has stayed on the zero face*. It passes when $G_j(C, H) \le \lambda_j$ for all $j \in C$. Section 4 constructs such bounds from current energy and anchor segments, including for changing nonlinear features.

Theorem 2 (Checkpoint-to-execution reduction)**.** *Under Assumption 1, let the current exact-zero set C pass the certificate interface for H accepted blocks. At current count k_0, the absolute endpoint is $K = k_0 + H$; earlier licenses must cover K. Remove C's live factors and anchors, add their squared residuals to ζ, and retain the original controller. Equations* (6)–(7) *then recover every specified retained candidate, complete objective, accept/reject decision and reconstructible control transition for the next at most H accepted blocks, including all intervening rejected trials. A common earlier stop is preserved.*

Proof. Radial identities recover every factor comparison on a zero prefix. Accepted multipliers lie in $[0, 1]$, preserving its certificate region. The certificate gives $q_j \ge 0$ at each removed zero gate, so finite-endpoint prox keeps every trial zero. Its displacement and Bregman term vanish; its raw loss derivative need not. Matched comparisons give matched controls and exclude a first revival. Induct over trials and blocks, stopping before $H + 1$. Appendix A gives the full argument. □

Corollary 3 (Necessary and sufficient radial information)**.** *Fix retained state, data/configuration and permitted controller inputs, and consider checkpoints individually certified to the same horizon. Include the current complete objective in the observations, but not certificate file identifiers. Their specified execution observations agree if and only if their ζ values agree.*

Equal radii suffice by the theorem; $F_1 - F_2 = \mu(\zeta_1 - \zeta_2)/2$ gives necessity. Even without observing F, the feasible family in Figure 1 has $u^+(\zeta) = 4/\sqrt{16 + \zeta}$ with strictly negative derivative. Every first trial is accepted, so a reproducing summary must distinguish these radii (Appendix A).

Corollary 4 (Memory scheduling does not change the real updates)**.** *Let C be licensed at block k_0 through the absolute endpoint $K = k_0 + H$. At accepted-block boundaries before K, release any still-stored subsets of C, adding their* current *squared anchor residuals to ζ exactly once. Keep the original controller and the license for C. Immediate, staged, and delayed release produce the same observations on the common survivors, complete objectives and trial decisions through K.*

Each release is an observation-preserving representation change; the theorem applies between changes. Finite induction proves the claim. A jointly licensed set need not have every physical subset independently pass the numerical bound: the original license covers all of C, including the still-stored part. Appendix A gives the proof. Section 6.1 tests the result by changing storage schedules while keeping one license and endpoint.

One correction for the complete acceptance margin. Write $\mathfrak{m} = f - f^+ + \langle \nabla f, d\rangle + (1-\nu)MD_\Theta + a_{\mathrm{sl}} \max(1, |f|)$, so $\mathfrak{m} \geq 0$ accepts. Let $\bar{\mathfrak{m}}$ use only retained terms *at the same correctly projected candidate*. Equation (7) simplifies to

$$\mathfrak{m} - \bar{\mathfrak{m}} = \tfrac{1}{2}[(1-\nu)M - \mu](\rho - 1)^2\zeta + \Delta_{\mathrm{sl}}, \tag{8}$$
$$\Delta_{\mathrm{sl}} = a_{\mathrm{sl}}\big[\max(1, |\bar{f} + \mu\zeta/2|) - \max(1, |\bar{f}|)\big].$$

Without slack, omission can cause false rejection when $M > \mu/(1-\nu)$ or false acceptance when $\mu \leq M < \mu/(1-\nu)$, provided the correction crosses zero. Candidate reconstruction alone is therefore insufficient.

4 Certifying the future from the checkpoint

Reach zero, then keep it. The scalar objective in Eq. (3) has increasing derivative $q + M(h'(z) - h'(x))$, giving

$$z^* = 0 \iff q \geq M(h'(x) - h'(0)); \qquad x = 0,\ q \geq 0 \implies z^* = 0. \tag{9}$$

The implemented entropy-core kernel has quadratic tails and finite endpoint slopes (Appendix B); this allows finite-step entry to an exact zero. Ordinary entropy's infinite endpoint slope does not.

A loss budget and a one-dimensional future region. At issuance let F_0 be the complete objective and $C_0 = \max(1, \sum_j \lambda_j)$ with the original slot count. Prox optimality and the unchanged acceptance rule give $F_{k+1} \leq F_k - \nu M_k D_k + a_{\mathrm{sl}} \max(1, |f_k|)$. Since $|f_k| \leq \max(F_k, \sum_j \lambda_j)$,

$$F_k \leq E_H = (1 + a_{\mathrm{sl}})^H (F_0 + C_0) - C_0, \qquad 0 \leq k \leq H. \tag{10}$$

On a zero prefix all selected anchor residuals receive the same accepted multiplier $\rho \in [0, 1]$. Consequently

$$a_j(t) = a_{w,j} + t\Delta a_j, \quad b_j(t) = b_{w,j} + t\Delta b_j, \qquad t \in [0, 1]. \tag{11}$$

These bounds concern the zero prefix, not a contraction assumption on the full nonconvex optimizer. The certificate prevents its first exit.

Regression and nonlinear cross-entropy share one interface. For squared regression with $\|a\|_X = \|Xa\|/\sqrt{n}$, Cauchy–Schwarz gives the inexpensive bound

$$G_j^{\mathrm{reg}} = |s|\sqrt{2E_H}\, \max(\|a_j\|_X, \|a_{w,j}\|_X) \max(\|b_j\|, \|b_{w,j}\|). \tag{12}$$

For cross-entropy $\ell = -\log p_y$, define the sharp universal constant

$$\|p - e_y\|^2 \leq c_* \ell, \qquad c_* = \max_{\ell > 0} \frac{2(1 - e^{-\ell})^2}{\ell} = 0.814528755\ldots. \tag{13}$$

Its maximizer is the unique positive solution of $e^{\ell_*} = 1 + 2\ell_*$; binary probabilities attain equality. For a nonlinear adapted layer, assume $\|h_i\| \leq H_i$ and $\|J_i\|_{\mathrm{op}} \leq L_i$ uniformly over the feasible region,

where J_i maps that layer's output to logits. With endpoint maxima $A_j = \max(\|a_j\|, \|a_{w,j}\|)$ and $B_j = \max(\|b_j\|, \|b_{w,j}\|)$,

$$G_j^{\mathrm{sharp}} = |s| A_j B_j \sqrt{c_* E_H \, n^{-1} \sum_i (H_i L_i)^2} = \sqrt{c_*/2}\, G_j^{(2)}. \tag{14}$$

The ratio is $0.638173\ldots$. The bound improves release coverage while retaining the same optimizer and feasible region. The chain rule gives $\partial_{\gamma_j} \ell_i = s(a_j^\top h_i)\langle p_i - e_{y_i}, J_i b_j\rangle$; Eq. (13) and samplewise Cauchy–Schwarz prove Eq. (14). The bounded tanh network uses analytic global H_i, L_i (Appendix D), not the Jacobian at one observed iterate.

Use the budget in the dangerous direction. For fixed-feature logits define $\kappa_{ij}(t) = \max\{0, \max_{\ell \neq y_i} s x_i^\top a_j(t)[b_{j,y_i}(t) - b_{j,\ell}(t)]\}$. The forcing obeys $-\partial_{\gamma_j} \mathcal{L} \le n^{-1} \sum_i \kappa_{ij}(t)(1 - p_{i,y_i})$. The same loss budget restricts these unknown probabilities through

$$\Psi_E(k) = \max_{0 \le z_i < 1} \left\{ \frac{1}{n} \sum_i k_i z_i : \frac{1}{n} \sum_i -\log(1 - z_i) \le E \right\}. \tag{15}$$

Its water-filling solution spends budget only where the dangerous coefficient is large. On cells $I_q = [l_q, u_q]$, quadratic extrema give $K_{ijq} = \sup_{t \in I_q} \kappa_{ij}(t)$, hence $G_j^{\mathrm{dir}} = \max_q \Psi_{E_H}(K_{\cdot jq})$. Appendix B derives the unique water level and a computable partition error bound. These are bounds on future gate forcing, not a search through future training trajectories.

The retained energy also buys a stronger certificate. For a proposed simultaneous set C, let $S_C = \zeta_{\mathrm{old}} + \sum_{j \in C} (\|\Delta a_j\|^2 + \|\Delta b_j\|^2)$; old certificates must cover the same endpoint. Its anchor energy is $\mu S_C t^2/2$, leaving the interval budget

$$E_q(C) = E_H - \tfrac{\mu}{2} S_C l_q^2, \qquad G_j^{\mathrm{joint}}(C) = \max_q \Psi_{E_q(C)}(K_{\cdot jq}). \tag{16}$$

At a valid checkpoint $E_H \ge F_0 \ge \mu S_C/2$, so these budgets are nonnegative. The same subtraction strengthens other energy-based bounds.

Proposition 5 (Largest feasible certificate portfolio)**.** *For common-horizon valid bounds with $A \subseteq C \Rightarrow G_j^{(l)}(A) \ge G_j^{(l)}(C)$, their minimum $G_j^{\mathrm{mix}} = \min_l G_j^{(l)}$ is valid and has the same monotonicity. Starting with all current zero candidates and iterating $C_{k+1} = \{j \in C_k : G_j^{\mathrm{mix}}(C_k) \le \lambda_j\}$ gives the largest feasible set in this bound family. It contains every constituent's largest feasible set.*

An element failing in C_k fails in every subset containing it; elimination therefore preserves every feasible subset. The finite fixed point is itself feasible. A cheap-first implementation short-circuits *within each current set*; set-dependent successes must be re-evaluated after that set shrinks.

For nonnegative per-slot array costs b_j and fixed representation overhead B_0, this largest set also maximizes $\sum_{j \in C} b_j - B_0$ among nonempty feasible sets of the same certificate family. Choose no release when this maximum is negative. This is an array-recovery consequence, not a runtime-optimality claim (Appendix B). The water level quantifies how joint energy subtraction changes the bound; its sensitivity and partition analysis are collected in Appendix B.

5 Connecting the reduction to numerical execution

The preserved object is the real recurrence in Section 2. We distinguish its direct floating expression, a cancellation-aware remainder expression, and directed-rounding validation. A matched floating trace checks an implementation; a validated trace additionally encloses the real state.

Evaluate the remainder, not cancelling objectives. For a factor displacement let $D = \|d\|^2/2$ and $\mathcal{R}_{\mathcal{L}} = \mathcal{L}(\Theta + d) - \mathcal{L}(\Theta) - \langle \nabla \mathcal{L}, d\rangle$. The same real acceptance margin is

$$\mathfrak{m} = [(1 - \nu)M - \mu]D - \mathcal{R}_{\mathcal{L}} + a_{\mathrm{sl}} \max(1, |f|). \tag{17}$$

Algorithm 1 License once; schedule storage independently

Input: original checkpoint/controller, fixed configuration, and horizon H.
Inspect exact zeros; evaluate a valid current certificate for a set C and absolute endpoint K. Every earlier license must cover K.
At a chosen accepted-block boundary before K, sum the current squared residuals of a still-stored subset of C into ζ; rebuild owning arrays. Keep the original C license, surviving anchors, scales and IDs.
Reconstruct every factor trial by Eqs. (6)–(7). A rejection updates only controls; an acceptance commits state and block count. Use the original prox for every stored gate.
Stop at the common stopping rule or K. Verified arithmetic suspends on an unresolved guard before commit; it does not reinterpret it as a rejection.

For $p_i = \mathrm{softmax}(z_i)$ and $\delta z_i = z_i^+ - z_i$,

$$\mathcal{R}_{\mathcal{L}} = \frac{1}{n} \sum_i \left\{ D_{\mathrm{KL}}(p_i \| p_i^+) + (p_i - e_{y_i})^\top (\delta z_i - J_{\Theta,i} d) \right\}. \tag{18}$$

The second term retains the nonlinear parameterization. The floating implementation uses centered elementary functions and extended accumulation. The interval implementation intersects valid direct and remainder bounds for the same margin. Appendix 4 proves both identities, the gate form and inclusion-preserving intersections.

Proposition 6 (Validated execution prefix)**.** *Given valid initial-state and certificate enclosures, inclusion-preserving trial evaluation and state propagation make every submitted branch identical to the real reference branch. The propagated state encloses that reference through the submitted prefix. Unresolved guards authorize no further transition.*

MPFR directed endpoints enclose the original oracle, projection and gate prox (MPFR Project, 2025). A nonnegative lower margin accepts; a negative upper margin rejects; a sign-spanning guard suspends. Induction proves the proposition. Exact decoded inputs are checked separately for feasibility: a point proved outside the ball is invalid, not a low-precision unresolved case. Starting with warm factors equal to their anchors also provides a feasible source for validated detection and release (Appendix 4). State intervals propagate throughout; higher precision restarts at the same source, never at a midpoint.

6 Experiments: storage choices for continued learning

We test three links: a learned zero is licensed; the smaller state executes the same prescribed transitions; and physical storage is reduced. A full reference and radial implementations start from one checkpoint and controller. Trial signatures record phase, inverse step and decision; every retained candidate, complete comparison and controller field is checked separately in the schedule study. Repetitions and nested trials are not independent tasks.

6.1 Eight independent adapters, one shared backbone

Our application is a working set of distinct adapters continuing from their own warm states on a common frozen $64 \to 32 \to 10$ tanh network. They share the original digit split and one backbone, not eight replicas of a saved optimization state. The target is a fixed 10° image rotation; both layers' factors and gates continue updating. Each branch retains its own anchor, shared factor ball and original controller.

Quality first, then evaluate release. Before confirmation, two development initializations compare $\lambda \in \{.003, .01, .03\}$ with unchanged $R = .5$, $\mu = .001$ and the original prox. A configuration must reach validation accuracy at least 85% and loss no more than its own warm loss plus .02. Among passing configurations, mean validation loss selects $\lambda = .003$; release counts and test data do not enter this choice. Eight preassigned, different adapter initializations then use this fixed configuration. Checks every five sweeps select the first nonempty certificate whose decoded point is exactly feasible,

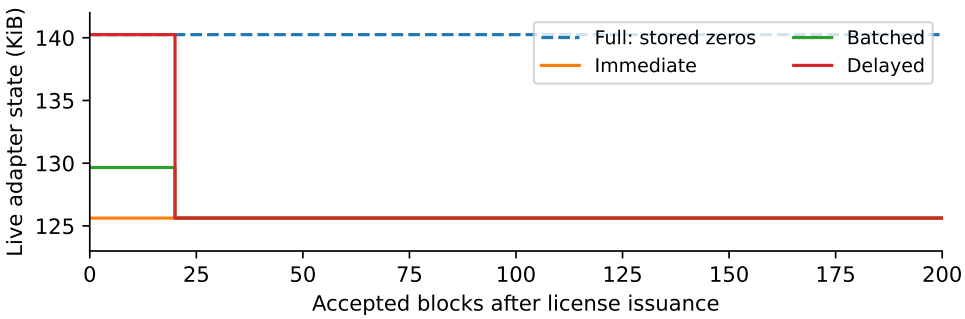

Figure 2: One learning rule, different storage schedules. All policies finish with 128,640 live bytes and matched signatures for all eight adapters. Immediate, staged and delayed release reduce byte–block residence by 10.43%, 10.14% and 9.39%, respectively, over the same 200-block tail. Residence uses algorithmic time, not wall-time memory pressure.

with $H = 200$. All empty and unsuccessful checks are retained. Appendix C gives the source, protocol and complete per-branch results.

All eight branches obtain nonempty release, authorizing 18 of 77 current zero slots (128 originally allocated). Directed-rounding checks independently verify the eight issuance points and selected inequalities for $H = 200$; these checks do not certify the subsequent floating trajectories. At the end all eight meet the predeclared validation quality criterion. Mean full-reference test accuracy is 84.63%, and all three radial schedules give identical test classes and accuracies. All 800 accepted factor steps activate the shared drift constraint. Seven branches have maximum tail logit changes between 1.83×10^{-3} and 7.28×10^{-2}; one is nearly stationary (4.63×10^{-9}). The aggregate live adapter state falls from 143,616 to 128,640 bytes, a 10.43% reduction. Every branch has distinct state arrays; shared backbone and data are counted once.

6.2 Change memory scheduling, not the learning rule

For each successful branch, fix one license C and one absolute endpoint. Compare immediate release; staged release (the first half of IDs immediately, the rest after 20 accepted blocks); and all-at-once release delayed by 20 blocks. The staged policy coincides with immediate release for singleton sets. No policy re-certifies a smaller subset, reselects a checkpoint, or extends the endpoint.

Across eight full-reference continuations there are 1,741 trials, 141 rejections and 1,600 accepted blocks. All three schedules match every trial signature and controller observation, with no observed licensed-gate revival. The largest retained-candidate absolute discrepancy is 1.17×10^{-15} and the largest logit discrepancy is 1.78×10^{-14}. These finite floating checks test Corollary 4; its real-arithmetic conclusion follows from the proof, not a tolerance test.

6.3 Mechanism and certificate strength

In the both-active projection regime $D_0 = \|p-a\|^2 > R^2$, put $\chi = c_M^2 \zeta / D_0$. Full and naive retained residuals are collinear, giving

$$\|u_0^+ - u^+\|/R = 1 - (1+\chi)^{-1/2}. \tag{19}$$

Appendix A proves all regimes and the bound $R[1 - (1 + c_M^2\zeta/R^2)^{-1/2}]$. Feasible accepted both-active trials have $\chi < 1$. The 217 active projections in the longer regression traces occupy $[1.04, 2.65] \times 10^{-4}$, a measured coupling regime rather than a large synthetic separation. In rank-four regression, naive deletion changes 16 sweep signatures and reaches prediction error 1.86×10^{-4}; full/radial prediction error is at most 1.12×10^{-15} across the 20 admitted runs.

On ten fixed-protocol 20° initializations, changing only the cross-entropy coefficient from 2 to c_* authorizes 28 versus 35 of the same 118 zeros: seven extra slots, a 25% relative increase, with unchanged domain, horizon and Jacobian bounds. They represent 35/160 initially allocated slots, not 35 independent tasks. The fixed-gate reference has 67.86% mean accuracy. Eight of ten additional

Continuation cohort	Release runs	Slots / zeros	Quality	Live saved
Regression, rank 2	10/10	57/57	.01316 loss	71.1%
Regression, rank 4	10/10	40/40	.01343 loss	49.9%
Regression, rank 8	0/10	0/0	.01387 loss	0%
Nonlinear, 20°	10/10	35/118	67.78% acc.	3.8–36.5%
Shared backbone, 10°	8/8	18/77	84.63% acc.	10.43%

Table 1: Learning and release, with distinct run and candidate denominators. Quality is full-reference test loss or accuracy; savings concern live adapter state, not backbone or process peak. Regression entries are group means, the 20° entry is a run range, and the shared-backbone entry is a working-set aggregate. The cohorts have separate protocols, not a common ranking task.

Input and evaluation	Completed evidence	What is checked
18 saved checkpoints, remainder	200 blocks each; 5,288 trials	Full/radial signatures; maximum logit difference 4.35×10^{-14}
Saved boundary, 1024-bit intervals	200 blocks; 300 decisions, each representation	Contained real states and proved decisions
Warm equals anchor, 1024-bit intervals	30 detection blocks; nonempty release; 200 tail blocks	Feasible-source detection and radial continuation
8 shared-backbone issuance points	Eight domains and $H = 200$ release inequalities	Validated authorization; subsequent floating tails tested separately

Table 2: Evidence on the declared input and arithmetic. Signature agreement is not a bound on floating-state distance to the real reference. The archived warm run has 230 accepted blocks and 346 decisions, distinct from the saved boundary.

radius, curvature, width and initialization controls admit release; both equal-amplitude empty controls remain in the denominator. On frozen ConvNeXt features for Fashion-MNIST, a separate 66-zero pool admits 0/0/2/2/2 under norm, centered sharp norm, directional, joint directional and portfolio bounds. Joint subtraction is below .417% of its energy budget; Appendix B gives a realizable case with strict joint gain.

6.4 Connect floating traces to a real reference

A saved boundary checkpoint exposes why arithmetic must be specified. Direct full/radial computation first gives margins -2.22×10^{-16} and zero, followed by 32 different sweep signatures. Its geometric correction is only 1.31×10^{-26}. The certified ideal margin at that trial instead lies in $[1.4228865130424542, 1.4228865130424544] \times 10^{-16} > 0$. All 300 remainder full/radial trial signatures match the certified reference; naive also matches in this low-coupling case. Table 2 separates this arithmetic test from the geometric counterexample.

The saved boundary needs 1024 bits to complete: 256-bit propagation stops after 60 accepted blocks at an unresolved comparison. A six-block screen certifies 13 of the 18 decoded inputs; five are provably outside the ball and are kept as invalid inputs. The feasible warm-source run licenses two of 12 exact zeros at block 30 and completes its 200-block tail, with final retained-state width 2.66×10^{-79}. No midpoint reset or input repair is used. The original 1024-bit saved and warm runs cost approximately 895 and 1,038 seconds, respectively; their validation is separate from fast execution costs.

7 What the smaller state buys

State reduction and execution speed answer different questions. Removing a slot reclaims its factors, anchors, gate and ID; the shared backbone, data and dense-network work remain. Our working-set experiment starts with eight independently warmed adapters in one process, sharing one oracle. Timed full and immediate-release arms symmetrically skip zero prediction and factor-data terms and use the same remainder comparison and copy-on-write strategy.

Eight-adapter working set	Full, zero terms skipped	Immediate radial release
Final live state, bytes	143,616	128,640
Completion time, s	14.883	15.326
Matched complete signatures	9/9 pairs	
Paired net time, s	-0.425 [-0.569, -0.349]	
Completion ratio	0.9726	

Table 3: Same arithmetic and endpoint, with all additional inspection and rebuilding inside the timer. Centers are medians of three batch medians; the bracket is the range of those medians, not a confidence interval. Trial logging is symmetric; interval issuance validation is a separate evaluation.

Charge the complete storage decision. Three fresh process batches each perform three random-order pairs from warm state to the same eight predeclared endpoints. The release arm pays every scheduled analytic check, empty/failed check, exact point-domain guard, license creation and reconstruction. Common initial file loading, cloning and prior source/warm training are outside both arms. No interval validation job runs concurrently. Thus the measured ratio concerns *warm-start-to-endpoint completion*, not source-to-finish training or strictly verified execution. All nine pairs match complete signatures, but all three batch net centers are negative: the smaller state does not produce a completion-time speedup in this workload. The cold decisions and shared dense computation are part of the measured cost, not removed from its denominator.

Distinguish final state from switching peak. Owning reconstruction frees the old arrays; allocating its replacement can still increase transient memory. With N equal-size states, whole-list rebuilding can hold $N(B_F + B_R)$ array bytes, whereas one-at-a-time commits need at most $NB_F + B_R$. This is an array-lifetime bound, not a process-residency identity (Appendix D). In the two 256-copy scale tests, sequential rather than batch reconstruction reduces observed array peaks from 7.152 to 4.386 MiB and 19.121 to 10.409 MiB. Single-copy live reductions remain 6,544 and 6,672 bytes; these replicas are not independent tasks. For the eight independent adapters, three isolated workers per schedule confirm owning buffers and distinct states. Post-run private-resident medians are 145,000 KiB for the full reference and 145,304 KiB for immediate release (a difference of +304 KiB). These process quantities are reported separately from the exact 14,976-byte live-state reduction; they do not establish a larger concurrent capacity.

The earlier six-case matched-arithmetic timing study contains 180 pairs. All 90 remainder pairs match signatures, but five of its six net centers are negative and every batch range crosses zero. Direct arithmetic matches 75/90; its 15 boundary-case failures do not count as trajectory-preserving speedups. These results and the positive charged regression tails (21.0/11.2 ms in their collection environment) identify different operating regimes; the underlying state-elimination guarantee is unchanged.

$$\Delta T(H) = T_F(H) - T_R(H) - C, \qquad S_{\mathrm{e2e}} = \frac{T_{\mathrm{pre}} + T_F(H)}{T_{\mathrm{pre}} + C + T_R(H)}. \tag{20}$$

All costs used on a path belong to C. With proved per-block saving $d > 0$, positive amortization requires $\lfloor C/d \rfloor + 1 \le H \le \min\{H_{\mathrm{safe}}, H_{\mathrm{remaining}}, H_{\mathrm{prior}}\}$ (Appendix D); noisy fitted times do not establish that lower bound.

Related guarantees and conclusion. LoRA, AdaLoRA, SoRA and statistical rank inference select representation or capacity (Hu et al., 2022; Zhang et al., 2023; Ding et al., 2023; Gao & Tan, 2026). Safe and dynamic screening already operate within iterative solvers (Ndiaye et al., 2017; Bonnefoy et al., 2015); exact lumping studies general state reduction (Horstmeyer & Atay, 2016). We construct the missing certificate and sufficient state for the specified jointly constrained training transition. Current authorization removes high-dimensional directions; radial reconstruction preserves their required influence; release-schedule independence makes memory timing a systems choice. Independent-adapter learning and matched execution measurements test that separation without replacing the chosen real update rule.

AI USE STATEMENT

Generative AI tools were used to aid manuscript restructuring, language editing, and documentation. We have reviewed all AI-assisted work. The mathematical results, proofs, computational checks, and reported numbers were produced and verified by the authors, who take full responsibility for the final content of the paper.

REPRODUCIBILITY STATEMENT

The appendices contain the core proofs and protocols. The artifact supplies all new shared-backbone, issuance and release-schedule inputs, source and raw results, plus the byte-verified surviving historical material. An availability ledger identifies historical aggregates whose original raw records are absent from the source archive; these are not replaced by another same-named snapshot. Clean and marked PDFs share one source. Archive manifests, independent-build checks and complete ordered volume instructions accompany the release.

REFERENCES

E. Alpaydin and C. Kaynak. Optical recognition of handwritten digits. UCI Machine Learning Repository, 1998. URL https://doi.org/10.24432/C50P49.

Antoine Bonnefoy, Valentin Emiya, Liva Ralaivola, and Rémi Gribonval. Dynamic screening: Accelerating first-order algorithms for the lasso and group-lasso. *IEEE Transactions on Signal Processing*, 2015. doi: 10.1109/TSP.2015.2447503. URL https://arxiv.org/abs/1412.4080.

Ning Ding, Xingtai Lv, Qiaosen Wang, Yulin Chen, Bowen Zhou, Zhiyuan Liu, and Maosong Sun. Sparse low-rank adaptation of pre-trained language models. In *Proceedings of EMNLP*, pp. 4133–4145, 2023. doi: 10.18653/v1/2023.emnlp-main.252. URL https://aclanthology.org/2023.emnlp-main.252/.

Yihang Gao and Vincent Y. F. Tan. Statistical inference for rank allocation in low-rank adaptation. *arXiv preprint arXiv:2607.20205*, 2026. URL https://arxiv.org/abs/2607.20205.

Leonhard Horstmeyer and Fatihcan M. Atay. Characterization of exact lumpability for vector fields on smooth manifolds. *Differential Geometry and its Applications*, 48:46–60, 2016. doi: 10.1016/j.difgeo.2016.06.001. URL https://arxiv.org/abs/1607.01237.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

MPFR Project. GNU MPFR 4.2.2: The multiple precision floating-point reliable library, 2025. URL https://www.mpfr.org/mpfr-current/mpfr.html.

Eugene Ndiaye, Olivier Fercoq, Alexandre Gramfort, and Joseph Salmon. Gap safe screening rules for sparsity enforcing penalties. *Journal of Machine Learning Research*, 18(128):1–33, 2017. URL https://jmlr.org/papers/v18/16-577.html.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Nikos Karampatziakis, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. AdaLoRA: Adaptive budget allocation for parameter-efficient fine-tuning. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=lq62uWRJjiY.

A EXECUTION, NECESSARY INFORMATION, AND NUMERICAL VALIDATION

Complete proof of Theorem 2. On the licensed zero face, $\nabla_v f = \mu(v - b)$. The unprojected removed residual is $c_M(v-b)$; shared projection gives $v^+ - b = \rho(v-b)$ and $v^+ - v = (\rho-1)(v-b)$. Squared norms and the gradient inner product prove Eqs. (6)–(7), including rejected $c_M < 0$ trials and the zero-denominator convention. At a removed gate, $q_j + M[h'(z) - h'(0)]$ is increasing and nonnegative at zero because the certificate gives $q_j \geq 0$. Strict convexity makes zero the unique prox solution. Its displacement, Bregman term and gradient inner product vanish, although its raw loss derivative may not.

Initially, retained arrays, ζ and controller agree. Every reconstructed observation, including complete f in the slack, gives the same proposal and comparison. Rejection changes only matched controls; acceptance commits matching state. Accepted $\rho \in [0, 1]$ preserves the common segment, and the energy bound below keeps the certificate valid on the zero prefix. The gate argument excludes a first exit. Induction over trials and accepted blocks proves equality through the common stop or endpoint, with no trial for block $H + 1$. Later releases add current residuals once; all licenses must cover that endpoint.

For one correctly projected candidate, full minus retained-only margin equals $[\mu(1 - \rho^2)/2 + \mu(\rho - 1) + (1 - \nu)M(\rho - 1)^2/2]\zeta + \Delta_{\text{sl}}$. Using $1 - \rho^2 + 2(\rho - 1) = -(\rho - 1)^2$ proves Eq. (8). For gate blocks, the non-slack radial terms cancel but complete f remains necessary for the relative slack.

Necessary information and virtual coordinate. The theorem proves sufficiency in Corollary 3; $F_1 - F_2 = \mu(\zeta_1 - \zeta_2)/2$ proves necessity. For the family in Section 2, projection gives $u^+ = 4/\sqrt{16 + \zeta}$ and $du^+/d\zeta = -2/(16 + \zeta)^{3/2} < 0$. The linear loss has zero remainder, so its complete margin is $D_\Theta/4 \geq 0$: all first trials are accepted. The inverse $\zeta = 16[(u^+)^{-2} - 1]$ proves the next-update obstruction. Generally, at fixed M, $D_0 = \|p - a\|^2 > 0$, active projection and $c_M \neq 0$, $\|u^+ - a\|^2 = R^2 D_0/(D_0 + c_M^2\zeta)$ is injective in ζ. The virtual coordinate $r = \sqrt{\zeta}$ gives $r^+ = \rho r$, $(r^+ - r)^2 = (\rho - 1)^2\zeta$ and $\mu r(r^+ - r) = \mu(\rho - 1)\zeta$. It explains the same reconstruction; replacing a negative trial r^+ by $|r^+|$ would change these quantities. Accepted $M \geq \mu$ gives $r^+ \geq 0$.

Proof of release-schedule independence. The license establishes zero persistence of the mathematical set C through K, independently of which arrays remain. At an event, converting only the newly removed current residuals into ζ preserves common survivors, complete objective and controller. Between events the main theorem applies. Induction over the finite events and intervening trials proves Corollary 4. No energy is counted twice; earlier licenses also cover K. A physical subset $D \subset C$ may have $G_j(D) > G_j(C)$; its authority is the original C license, not a newly asserted subset bound.

All deletion-error regimes. Writing $D_0 = \|p - a\|^2$ and $e = c_M^2\zeta$, collinearity gives

$$\|u_0^+ - u^+\| = \begin{cases} 0, & D_0 + e \leq R^2, \\ \sqrt{D_0}[1 - R/\sqrt{D_0 + e}], & D_0 \leq R^2 < D_0 + e, \\ R[1 - (1 + e/D_0)^{-1/2}], & D_0 > R^2. \end{cases} \tag{21}$$

For fixed e, the middle derivative is $[1 - Re/(D_0 + e)^{3/2}]/(2\sqrt{D_0}) > 0$; the last branch decreases. Thus the maximum is $R[1 - (1 + e/R^2)^{-1/2}]$. Feasible accepted trials have $e \leq R^2$, hence $\chi < 1$ in the last branch. Integration of $\frac{1}{2}(1 + \chi)^{-3/2}$ gives bounds $R\chi/[2(1 + \chi_{\max})^{3/2}]$ and $R\chi/2$.

Remainders and validated prefixes. Expanding the anchor gives $f(\Theta + d) - f(\Theta) - \langle \nabla f, d\rangle = \mathcal{R}_{\mathcal{L}} + \mu\|d\|^2/2$, proving Eq. (17). The log-softmax identity gives $\ell(z + \delta z) - \ell(z) - (p - e_y)^\top \delta z = \log \sum_c p_c e^{\delta z_c} - p^\top \delta z = D_{\text{KL}}(p\|p^+)$. Adding and subtracting $(p - e_y)^\top \delta z$ proves Eq. (18). For the linear-penalty gate block the anchor is constant, giving $\mathfrak{m} = (1 - \nu)MD_h - \mathcal{R}_{\mathcal{L}}(\gamma, \Delta\gamma) + a_{\text{sl}} \max(1, |f|)$; MCP adds its own Taylor remainder.

Directed endpoints enclose arithmetic extrema and monotone elementary functions, with zero-crossing square lower bounds and nonzero-divisor guards (MPFR Project, 2025). Outward matrix products, projection and verified inverse kernel branches preserve inclusion; ambiguous branches suspend. Spectral proposals pass only if interval $LDL^\top$ has positive pivots for $b^2 I - A^\top A$ (or the smaller

Gram matrix). The positive sharp-constant root is bracketed in $[1.256431208626, 1.256431208627]$; interval evaluation bounds c_* above by $814529/10^6$. These operations validate energy and certificate inequalities, not just point-SVD estimates. Exact-input ball slack is classified as feasible, invalid or unresolved by its endpoints. For Proposition 6, inclusion determines every submitted margin sign. Rejection updates matched controls and acceptance encloses the original map; induction gives the claimed prefix. Unresolved signs submit nothing; higher precision restarts from the same source, with exact-zero margins not guaranteed to resolve finitely.

Intersections and feasible acquisition. For $T(e, D) = e \min\{1, R^2/(D+e)\}$, $T(0, 0) = 0$, active-branch derivatives are $\partial_e T = R^2 D/(D+e)^2 \geq 0$ and $\partial_D T = -R^2 e/(D+e)^2 \leq 0$. Continuity and the inactive branch $T = e$ give rectangular range $[T(e_-, D_+), T(e_+, D_-)]$. Intersect its outward enclosure with $[\rho]^2[\zeta]$, retaining signed ρ elsewhere. Taylor's integral formula and $4 \leq h''_\delta \leq L_\delta$ bound the gate distance by $[2d^2, L_\delta d^2/2]$. Intersect it with direct Bregman evaluation. Two valid margin bounds for the same real trial can likewise be intersected; an empty intersection is an error, not a selectable sign. Warm factors equal to their anchors define a feasible exact source. Original factor projection stays in the ball and gate blocks leave factors fixed. Containing propagation preserves that feasible reference through detection; a passing certificate authorizes release and the main theorem applies thereafter. Induction proves the acquisition–release chain. Its containing box need not lie wholly in the ball; its reference does. No midpoint replacement or repair of unrelated exterior point inputs is used.

B ZERO PERSISTENCE AND LOSS-BUDGET PROOFS

Endpoint prox, penalty and energy. The strictly increasing prox derivative gives $h'(z^*) = \mathrm{clip}_{[h'(0), h'(1)]}(h'(x) - q/M)$ and $z^* = 0 \iff q \geq M[h'(x) - h'(0)]$. For $b_0(x) = x \log x + (1 - x) \log(1 - x)$ and $0 < \delta < 1/2$, let $L_\delta = [\delta(1-\delta)]^{-1}$ and $T_a(x) = b_0(a) + b'_0(a)(x - a) + L_\delta(x - a)^2/2$. The implemented h_δ equals b_0 on $[\delta, 1 - \delta]$, T_δ below, and $T_{1-\delta}$ above. Value, slope and curvature match, so it is C^2, has $4 \leq h''_\delta \leq L_\delta$ and finite endpoint slopes. Ordinary entropy cannot reach zero from the interior at finite q/M. We use $\delta = .1$ (Fashion: .05).

Fashion uses $p_{\lambda,\eta}(t) = \lambda t - t^2/(2\eta)$ for $t \leq \eta\lambda$ and $\eta\lambda^2/2$ above. With $P_2 = \lambda t - p_{\lambda,\eta}(t)$ and $f = \mathcal{L} + \mu\|\Theta - \Theta_w\|^2/2 - P_2(\gamma) + \beta c^\top \gamma$, $F = f + \lambda^\top \gamma \geq 0$, $P'_2(0) = 0$ and the zero threshold is $\lambda_j + \beta c_j$. Factor geometry is unchanged. Prox optimality against its center gives $\langle \nabla f, d\rangle + MD \leq 0$ for factors and $\langle \nabla f + \lambda, d\rangle + MD \leq 0$ for gates. Combining with Eq. (4) yields $F_{k+1} \leq F_k - \nu M_k D_k + a_{\mathrm{sl}} \max(1, |f_k|)$. Since $0 \leq \lambda^\top \gamma \leq \Lambda = \sum_j \lambda_j$, $|f_k| \leq \max(F_k, \Lambda)$, so $F_{k+1} + C_0 \leq (1 + a_{\mathrm{sl}})(F_k + C_0)$ for $C_0 = \max(1, \Lambda)$. Induction proves Eq. (10); its stable evaluation is $F_0 + (F_0 + C_0)\,\mathrm{expm1}(H\,\mathrm{log1p}(a_{\mathrm{sl}}))$.

Common segment and sharp nonlinear bound. On a zero prefix, accepted multipliers $\rho \in [0, 1]$ have a common product t, proving Eq. (11). Regression has $\partial_{\gamma_j}\mathcal{L} = (s/n)\langle \widehat{Y} - Y, (Xa_j)b_j^\top\rangle_F$. Cauchy–Schwarz and norm convexity give Eq. (12); the endpoint prox excludes the first exit. For cross-entropy, off-class squared probability mass is at most $(1 - p_y)^2$, hence $\|p - e_y\|^2 \leq 2(1 - p_y)^2$. The derivative of $r(\ell) = 2(1 - e^{-\ell})^2/\ell$ is $2(1 - e^{-\ell})[2\ell e^{-\ell} - (1 - e^{-\ell})]/\ell^2$. The function $e^\ell - 1 - 2\ell$ decreases then increases, with one positive root. Since r vanishes at both ends, that root gives the global maximum c_*, attained by binary probabilities. The chain rule and uniform bounds give

$$|\partial_{\gamma_j}\mathcal{L}| \leq |s|\|a_j(t)\|\|b_j(t)\| \sqrt{\left(n^{-1}\sum_i (H_i L_i)^2\right)\left(n^{-1}\sum_i \|p_i - e_{y_i}\|^2\right)} \leq G_j^{\mathrm{sharp}}. \tag{22}$$

The last step uses the loss budget and endpoint maxima. For linear logits, $(p - e_y)^\top \mathbf{1} = 0$ also allows b to be centered by $I - \mathbf{1}\mathbf{1}^\top/C_{\mathrm{cls}}$.

Directional water filling and sensitivity. Expansion gives $-\partial_{\gamma_j}\mathcal{L} \leq n^{-1}\sum_i \kappa_{ij}(t)(1 - p_{i,y_i})$; $z_i = 1 - p_{i,y_i}$ is feasible in Eq. (15). The Lagrangian upper bound is $\tau E + n^{-1}\sum_i [k_i z_i + \tau \log(1 - z_i)]$, $\tau \geq 0$. For $\tau > 0$ its coordinate maximizer is $z_i = (1 - \tau/k_i)_+$ ($z_i = 0$ for $k_i = 0$). When $E > 0$,

$k \neq 0$, the unique water level solves

$$n^{-1} \sum_{k_i > \tau} \log(k_i/\tau) = E, \qquad \Psi_E(k) = n^{-1} \sum_{k_i > \tau} (k_i - \tau). \tag{23}$$

The left side decreases continuously from infinity to zero. These maximizers satisfy the budget and attain their Lagrangian upper bound, proving optimality; zero budget or coefficients give zero value. With active-set size m, $\log \tau = (\sum_{i \le m} \log k_{(i)} - nE)/m$, $\partial_E \Psi_E = \tau$, and $\partial_E \tau = -n\tau/m$. Continuity across active-set changes and integration give

$$\tau(E)\delta \le \Psi_E(k) - \Psi_{E-\delta}(k) \le \tau(E-\delta)\delta. \tag{24}$$

The relaxation need not equal the network-reachable probability set. Each signed class coefficient is quadratic; endpoints and concave interior vertices give K_{ijq}, and coordinate monotonicity proves the directional bound.

Joint energy, largest sets and partition resolution. Nonnegative terms yield $\mathcal{L}(t) \le E_H - \mu S_C t^2/2$; $E_H \ge F_0 \ge \mu S_C/2$ ensures nonnegative budgets. For fixed envelopes, $A \subseteq C$ implies $G_j(A) \ge G_j(C)$. A feasible $S \subseteq C_k$ cannot contain an element failing in C_k, since $G_j(S) \ge G_j(C_k) > \lambda_j$ would contradict feasibility. Peeling therefore retains every feasible subset; its finite fixed point is itself feasible and largest. Taking a minimum of valid monotone bounds preserves both properties and contains each constituent feasible family, proving Proposition 5.

For $G^* = \sup_t \Psi_{E(t)}(\kappa(t))$, left-endpoint evaluations give $L_\Delta \le G^*$ and interval maxima/budgets give $G^* \le U_\Delta$. Because $0 \le z_i \le 1$, Ψ is $1/n$-Lipschitz in coefficient 1-norm; its budget derivative $\tau \le \|k\|_\infty$. If κ_i is L_i-Lipschitz, $\kappa_i \le \bar{\kappa}$, and $E(t) = E_H - \mu S_C t^2/2$, then $U_\Delta - L_\Delta \le K\Delta$, $K = n^{-1}\sum_i L_i + \mu S_C \bar{\kappa}$. Nested refinement decreases the exact upper bound. Certify if $U \le \lambda$; $L > \lambda$ fails this relaxation; otherwise refine. A strict gap δ resolves when $\Delta < \delta/K$; equality need not resolve. Floating bounds additionally require validated evaluation errors.

A realizable strict improvement. Take one input $x = 1$, two classes, label one, zero gate/anchors, $a = \sqrt{.9}$, $b = (\sqrt{.9}, 0)^\top$, $s = 10/9$, $\mu = 1$, $R = \sqrt{1.8}$, $\lambda = .3$. Frozen logits give $p_y = e^{-.1}$, so $F_0 = E_H = 1$ at zero slack and $\kappa(t) = t^2$. The uncorrected bound is $1 - e^{-1} > .3$; writing $x = t^2$ gives $\max_{[0,1]} x[1 - e^{-(1-.9x)}] \le \max_{[0,1]} x(1 - .9x) = 5/18 < .3$. With $K = 2 + 1.8$, a 256-cell upper bound is $5/18 + 3.8/256 < .3$.

Additive byte recovery. For fixed $b_j \ge 0$ and common nonempty overhead B_0, every feasible $C \subseteq C^*$ satisfies $\sum_C b_j - B_0 \le \sum_{C^*} b_j - B_0$. Thus the largest certificate set maximizes this byte ledger even for unequal layer sizes; compare with zero when declining release is allowed. This does not optimize nonadditive workspaces, rebuilding or runtime.

C PROTOCOLS, DENOMINATORS AND EVIDENCE AVAILABILITY

Regression and frozen features. Regression uses a 64×64 map, eight slots, 256/512/512 train/validation/test examples, noise .02 and target ranks 2/4/8. Training-only SVD warmup and seeds 801–802 select $(R, \lambda) = (.5, .01)$ before seeds 901–910 per rank; $s = 1$, $\mu = .001$, $M_{\min} = .1$, $h_{.1}$. Detection lasts at most 100 sweeps then 200 tail sweeps; a check at t uses $H = 2(300 - t)$. The restart study uses ranks 2/4, seeds 901/902, $H = 600$, splitting after four sweeps: 4,807 trials, 217 active projections, 76 mutation checks. Fashion uses frozen ConvNeXt-Tiny features (1,500/500/1,000), rank 16, $s = .125$, $\mu = 10^{-4}$, $\beta = .00375$, $\lambda = .01$, MCP shape 3, $h_{.05}$. Two of ten runs release; full/radial test accuracies are 86.0/85.4% and net seconds $-.108/-.066$. The 66-zero pool uses $H = 200$ and partitions 1/4/16/64; cheap-first agrees with the full portfolio at 16.

Original nonlinear confirmation. Digits (Alpaydin & Kaynak, 1998) use a stratified 1,100/300/397 split (seed 20260926), pixels divided by 16, 20° rotation with linear interpolation and zero padding. A width 32 tanh source trains for 400 full-batch Adam steps at .02, then freezes. Two eight-slot adapters have unit gates/scale and .05-times exchangeable log-normal amplitudes (log-standard-deviation 1); 250 factor-warmup steps define the anchors. Adaptation uses $R = .5$, $\mu = .001$, $\nu = .5$, $a_{\mathrm{sl}} = 10^{-12}$, $M_{\min} = .01$, $h_{.1}$. Development seeds 2601–2602 choose $\lambda = .01$ from .003/.01/.03 by nonempty

release, then validation loss. Confirmation seeds 2801–2810 use checks every five sweeps through 200, first sharp admission and $H = 200$. This historical selection differs from the quality-first study below. The fixed-gate comparator shares warm state and factor-gradient budget. Eight of ten controls

Seed	2801	2802	2803	2804	2805	2806	2807	2808	2809	2810
Coeff. 2 / sharp	1/2	3/3	4/4	3/3	4/4	5/5	1/2	1/4	6/7	0/1
Zero candidates	10	12	14	10	12	15	12	8	13	12
Accuracy (%)	64.74	67.51	71.79	62.22	72.29	71.79	66.75	67.25	69.02	64.48
Live saved (%)	7.66	16.31	15.33	16.31	15.33	23.98	7.66	15.33	36.45	3.83
Direct differences	0	0	0	0	0	0	32	0	0	0
Remainder differences	0	0	0	0	0	0	0	0	0	0

Table 4: All original nonlinear initializations. Differences are full/radial sweep-signature mismatches within each stated arithmetic backend.

(width 64/128, radius.25/.75, curvature.0005/.002, equal-amplitude starts) admit; both empty starts stay in the 20-run denominator. The 18 admitted snapshots store arrays, anchors, gates, inverse steps and masks. Their 200-block floating replay does not repeat detection. Interval screening certifies 13 six-block prefixes and rejects five exact ball-exterior inputs; a broad enumeration additionally checks three calibration points (one valid, two invalid), excluded from 13/18. Natural and intersected 256-bit bounds both stop at 60 blocks on the boundary case. Validated warm acquisition fixes that case's warm=anchor source,1024 bits, five-sweep checks through 50 and a 200-block post-release tail. Published totals are 230 blocks/346 decisions, with no reset.

Historical costs and availability. The six-case cost study has three fresh processes per case, five random-order pairs per arithmetic, three warmup sweeps and 100 timed sweeps (180 pairs). Cold checks and reconstruction are charged. It uses one pinned BLAS thread, but interval work on another CPU and shared-host load are uncontrolled. Remainder net centers in case order 2801/2802/2803/2901/2911/2807 are $-4.51, -16.71, -41.08, -6.24, -9.56, +1.98$ms; all batch ranges cross zero. Memory uses two inputs,1/16/64/256 replicas, three roles and three workers: 72 workers, sharing only the oracle. Array ownership and boundary-sampled peaks are separate from RSS/private and historical high-water marks; explicit trim and one subsequent remainder sweep are logged. At 256 replicas, sequential-minus-full post-sweep private medians are $+660/-636$KiB for width 32/128. The supplied canonical source was truncated after 638 complete entries. Surviving files and published aggregates are preserved; missing original raw timing, worker and warm-acquisition records are identified in the availability ledger, not reconstructed from another same-named version. All experiments below have complete new raw records.

Independent adapters and quality-first selection. One common frozen backbone comes from the boundary case; the original digit split receives a fixed 10° rotation. Eight-slot adapters use the same warmup and update constants above. Development seeds 31001–31002 run 100 sweeps per λ. Eligibility requires validation accuracy $\geq .85$ and $\mathcal{L}_{\mathrm{val}} \leq \mathcal{L}_{\mathrm{warm,val}} + .02$. Minimum mean validation loss selects.003 (tie: smallerλ), without release counts or test data. Confirmation seeds 31101–31108 and checks every five sweeps through 100 are fixed before development. The first nonempty analytic certificate at an exactly feasible decoded point fixes $H = 200$. Independent 256-bit validation checks all eight issuance domains and selected inequalities only, with zero continuation blocks. Immediate removes all IDs at issuance; staged removes the first $\lceil |C|/2 \rceil$ then the rest 20 blocks later; delayed removes all at 20. Single-slot staged runs equal immediate. Current energy is added once; the original license and absolute endpoint remain fixed. Instrumented trials compare common-survivor candidates, full comparison quantities and controls. Performance uses the original uninstrumented solver, not that observer. Completion costs use three new processes, three random-order pairs each, from eight warm states through detection and tails in round-robin order. All additional checks, domain guards and rebuilding are timed; common loading, cloning and prior source/warm training are excluded. No concurrent experiment runs; host frequency/load remain uncontrolled. Twelve isolated memory workers (three per physical schedule including full) retain only their chosen arm. A timed-run interruption is preserved and that entire batch was rerun unchanged; its incomplete output is excluded, not selected by speed.

Adapter	Issue sweep	Slots / zeros	Val. acc.	Test acc.	Live bytes
31101	40	1/12	87.67%	83.38%	17,264
31102	30	1/7	85.67%	81.86%	17,264
31103	35	4/12	91.67%	88.92%	14,336
31104	30	3/11	88.33%	84.89%	15,024
31105	30	2/8	88.00%	83.88%	16,576
31106	40	1/6	88.00%	84.13%	17,264
31107	30	5/10	89.00%	86.40%	13,648
31108	35	1/11	87.67%	83.63%	17,264

Table 5: Every independent adapter starts with 17,952 live bytes. All three schedules match full-reference accuracy; validation eligibility is not a test accuracy threshold.

D GLOBAL NONLINEAR BOUNDS AND USABLE HORIZONS

For a two-layer tanh network of width m, second-layer inputs satisfy $\|h_i\| \le \sqrt{m}$ and the output Jacobian is the identity. First-layer inputs satisfy $H_i = \|x_i\|$ and its downstream Jacobian norm is at most $\|W_2\|_{\mathrm{op}}$, since tanh is 1-Lipschitz. For scale one, write $W_2 = W_{20} + (B_w + \Delta B)\,\mathrm{Diag}(\gamma)(A_w + \Delta A)$ and let $a = \|A_w\|_{\mathrm{op}}$, $b = \|B_w\|_{\mathrm{op}}$. With $x = \|\Delta A\|_F$, $y = \|\Delta B\|_F$, $x^2 + y^2 \le R^2$ and $\|\,\mathrm{Diag}(\gamma)\|_{\mathrm{op}} \le 1$,

$$\|W_2\|_{\mathrm{op}} \le \|W_{20}\|_{\mathrm{op}} + ab + bx + ay + xy \le \|W_{20}\|_{\mathrm{op}} + ab + R\sqrt{a^2 + b^2} + R^2/2. \tag{25}$$

Cauchy–Schwarz bounds the linear terms and $xy \le (x^2 + y^2)/2$ bounds the product. The constants are uniform on the whole shared ball, including all operating widths; frozen biases preserve the tanh bounds.

Safety versus amortization. For fixed candidates and bounds, $H_1 \le H_2$ gives $E_{H_1} \le E_{H_2}$, so $C^*(H_2) \subseteq C^*(H_1)$. If F_0 passes and $a_{\mathrm{sl}} > 0$, solving $E_H \le E_{\mathrm{crit}}$, the greatest closed admissible budget, gives $H \le \log((E_{\mathrm{crit}} + C_0)/(F_0 + C_0))/\log(1 + a_{\mathrm{sl}})$. Floor and verify the bound, intersect prior endpoints and remaining blocks. Zero slack gives constant energy. A deterministic bound $\Delta T \ge Hd - C$ gives

$$\lfloor C/d \rfloor + 1 \le H \le \min\{H_{\mathrm{safe}}, H_{\mathrm{remaining}}, H_{\mathrm{prior}}\}. \tag{26}$$

For layer output $Z = HW_0^\top + s[(HA^\top)\,\mathrm{Diag}(\gamma)]B^\top$, $d\mathcal{L} = \langle G, dZ\rangle_F$ and cyclic trace identities give the loss derivatives below; the original anchor gradients are added afterwards.

$$\nabla_B \mathcal{L} = sG^\top[(HA^\top)\,\mathrm{Diag}(\gamma)], \qquad \nabla_A \mathcal{L} = s[(GB)\,\mathrm{Diag}(\gamma)]^\top H,$$
$$\nabla_H \mathcal{L} = GW_0 + s[(GB)\,\mathrm{Diag}(\gamma)]A, \qquad \partial_{\gamma_j} \mathcal{L} = s\sum_i (HA^\top)_{ij}(GB)_{ij}.$$

Ownership accounting. After k sequential commits, construction of the next copy holds $(N - k)B_F + kB_R + B_R \le NB_F + B_R$ because $B_R \le B_F$. Each map reads only its own state, so ordering changes neither output nor controller. This proves the equal-size array bound. For unequal groups, $\sum_i B_{F,i} + \max_i B_{R,i} + W_{\mathrm{rebuild}}$ includes reconstruction workspace. Containers and allocator caches are additional. A failed sequential operation reports its committed prefix, not an atomic batch. The bound concerns live arrays, not a process's historical high-water mark.