FINITE-STEP ADAPTATION DECISIONS: LOSS RESPONSES AND RETAINED DIRECTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Scalar loss agreement can hide the effects of directions omitted by finite adaptation. We study when loss observations certify those effects. For radial updates containing ordinary and normalized gradient descent, we derive a response law at every finite derivative order. Its leading matrix depends affinely on the initial gain slope. This law transfers explicit direction-containment tests across the slope interval $[0, 1]$ and explains realizable loss cancellations. A normal-gradient correction at a saved state instead produces a positive energy measurement. We characterize four-prefix observation designs by an exact rank condition and give finite transfer bounds for unobserved steps. Residual propagation then converts these observations into target-bias guarantees, including execution errors and changing batches. An optional reference linearization tightens the guarantee at additional derivative cost. For a logistic head on fixed features, the resulting intervals distinguish adequate restricted updates from inadequate ones and zero update.

1 INTRODUCTION

Adapting a pretrained model often means choosing which parameter directions may change and how many updates to perform. A small update space reduces the variables that must be represented or optimized. It also excludes responses that a full update could follow. This creates a finite decision problem: is the chosen restriction adequate for the outcome of this adaptation? The answer depends on the loss, the execution budget and the evaluation target, as well as on the dimension of the update space. Matching the initial gradient does not settle this question. After the first step, curvature changes the gradient. Later steps can introduce further directions through higher derivatives. Nor does matching the terminal training loss settle it. The loss is a scalar summary of a high-dimensional state. Different omitted effects can cancel in that summary, while remaining visible to another execution or to a validation objective. A useful certificate must distinguish a small omitted effect from an effect hidden by the chosen observation.

We study this distinction through scalar loss responses. An auditor may receive paired terminal losses, or the loss decrease produced by a controlled correction at a saved state. The executor constructs the compared updates or correction direction. The observation cost includes this construction and the loss evaluations. When residual norms are already available, they provide a more direct input to the same finite certificate. Our question is which information suffices for a prescribed target tolerance, and what must be paid to obtain it.

The analysis starts with a fixed retained space and a fixed loss. We derive how omitted directions enter the terminal response of ordinary descent, normalized descent and smooth normalizations. Their leading response matrices share an affine dependence on one gain parameter. This structure explains both cancellation and the response designs that exclude it. We then change the observation: a short correction along the omitted gradient produces positive energy. The resulting checkpoint measurements detect all relevant omitted combinations exactly when an explicit coefficient matrix has full rank. Its conditioning and Taylor remainder govern finite transfer. Residual bounds then propagate through two specified finite executions, including shared stochastic batches, refreshed spaces and update defects. A reference linear correction can preserve information lost by norm-only propagation, at extra derivative cost. The frozen-feature example shows how this directional information changes the certification decision on the same trajectories.

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 Related work

Low-dimensional adaptation motivates the restriction being audited. LoRA represents a weight update through two thin factors while freezing the pretrained weights (Hu et al., 2022). Intrinsic-dimension studies ask how many search directions can suffice for learning (Li et al., 2018). Gradient compression instead reduces the representation used during optimization (Kasiviswanathan, 2021). GaLore projects gradients into low-rank spaces (Zhao et al., 2024). Flora relates LoRA dynamics to random projection and resamples the projection to permit high-rank accumulated updates (Hao et al., 2024). Gradient alignment relates these restricted updates to full updates. LoRA-GA initializes the factors to approximate the first full gradient (Wang et al., 2024). For linear models, LoRA-One studies alignment with singular subspaces of the initial full gradient and derives a theory-guided initialization (Zhang et al., 2025). LoRA-Pro adjusts factor gradients to reduce the discrepancy in the induced first-order weight update (Wang et al., 2025). For longer executions, GoLore gives stochastic convergence guarantees for randomized subspace momentum updates (He et al., 2025); PESO formulates guarantees for induced updates in the full parameter space (Lou et al., 2026). A finite target certificate compares two coupled executions at a prescribed stopping point, possibly on a different loss.

This comparison also has an information requirement. Classical observability concerns whether outputs distinguish internal states (Hermann & Krener, 1977). Data informativity asks whether data suffice for a desired analysis or control property without first identifying a unique system (van Waarde et al., 2020). Our observations test containment of the directions generated by finite adaptation. A zero terminal response can remain compatible with nonzero omissions. The relevant design problem is therefore to exclude those compatible alternatives. Gram matrices express this ambiguity through inner products, connecting the analysis to quadratic lifting and positive semidefinite separation (Candès et al., 2013; Vandenberghe & Boyd, 1996). Derivative expansions describe the directions that these observations must detect. B-series organize derivatives of composed updates, while nonlinear moment matching relates local responses to projection spaces (McLachlan et al., 2017; Benner & Breiten, 2015). We use this established calculus to identify the directions generated by finite adaptation. The specific structural result is that the leading omitted-direction response is affine in the initial gain slope. This permits explicit tests to transfer across gains. Convergence under adaptive normalization has also been analyzed (Yang et al., 2025).

A finite certificate must additionally control accumulated error. GeoLoRA is a close comparison: its Theorem 3 bounds low-rank iterates against full gradient flow under smoothness and a uniform near-tangency bound around that flow (Schotthöfer et al., 2025). Here controlled loss observations supply bounds on the omitted force, and the comparator is a specified discrete execution. A separate target is handled through classical output-oriented residual estimation (Becker & Rannacher, 2001; Janon et al., 2016). Smoothness bounds and bounded-error set-membership estimation supply further supporting tools (Bubeck, 2015; Milanese & Vicino, 1991). The contribution lies in linking the response law and observation design to these finite guarantees with explicit information and execution costs. The comparison requires specified programs, observations and a target.

3 Background and the certification problem

An adaptation starts from a fixed parameter vector and applies a finite sequence of updates to a training loss. We translate that starting point to zero. Let $F : \mathbb{R}^d \to \mathbb{R}$ be the loss and let u_t denote the full parameter displacement after t updates. A restricted execution has displacement $\bar{u}_t$. It keeps updates in a chosen linear subspace.

The orthogonal projector $W = W^\top = W^2$ selects that subspace. The identity I acts on the full parameter space, and $P = I - W$ selects the omitted directions. Thus every gradient splits as $\nabla F = W\nabla F + P\nabla F$. For ordinary gradient steps with positive step sizes η_t,

$$u_0 = \bar{u}_0 = 0, \qquad u_{t+1} = u_t - \eta_t \nabla F(u_t), \qquad \bar{u}_{t+1} = \bar{u}_t - \eta_t W \nabla F(\bar{u}_t).$$

The gradients are evaluated at their respective states. Even when the first steps agree, later states and gradients can differ. We compare the programs after a prescribed number N of updates.

Let V be the target used to evaluate their outcome. For example, F may be a training loss and V a validation loss. The signed target bias and the adequacy decision are

$$\beta_V = V(\bar{u}_N) - V(u_N), \qquad |\beta_V| \le \tau,$$

where $\tau > 0$ is the chosen tolerance. A certificate must enclose β_V using the allowed information. An interval entirely inside $[-\tau, \tau]$ certifies adequacy. An interval entirely outside it certifies a violation. An interval crossing its boundary leaves the decision unresolved.

The distinction between an observed loss and the target is already visible in two dimensions. Take

$$F(a,b) = -a + \tfrac{1}{2}a^2 + ab + 7b^2, \qquad W = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \eta_0 = \eta_1 = \tfrac{1}{4}.$$

The full program updates both coordinates; the restricted program retains only a. Direct evaluation gives

$$u_1 = \bar{u}_1 = (1/4, 0), \qquad u_2 = (7/16, -1/16), \qquad \bar{u}_2 = (7/16, 0),$$
$$F(u_2) = F(\bar{u}_2) = -175/512, \qquad V(\bar{u}_2) - V(u_2) = 1/16 \quad \text{for } V(a,b) = b.$$

Thus equal terminal training losses leave a nonzero target bias. A different loss observation reveals the omission. Copy the saved state $y = \bar{u}_1$ and correct its omitted gradient:

$$r = P\nabla F(y) = (0, 1/4), \qquad y^- = y - \tfrac{1}{16}r = (1/4, -1/64), \qquad F(y) - F(y^-) = 9/4096 > 0.$$

The correction acts on a copy, so it does not change the reference execution being audited.

This example defines the two observation types used below. A paired response is the terminal difference $\Delta(B) = F(\bar{u}_N(B)) - F(u_N(B))$ for a prescribed budget B. A checkpoint report is the decrease $D(y) = F(y) - F(y - \sigma P\nabla F(y))$ at a saved state, with correction length $\sigma > 0$. The analyst may receive only these scalar values. The executor still pays for the compared routes or correction direction. If the normal norm is already available, it avoids the extra probe losses.

Two questions must be separated. The structural question asks whether a response that vanishes as $B \downarrow 0$ forces particular directions to be retained. The finite question asks whether intervals measured at nonzero budgets certify $|\beta_V| \le \tau$. The first identifies informative observations. The second must also account for remainders, numerical error and propagation along the execution.

4 Response structure and cancellation

The example shows that a terminal loss can hide an omitted direction. We next vary the total step budget. At small budgets, different powers of the budget separate effects that enter at different derivative orders. This identifies which directions a vanishing response can certify. We hold the loss and retained space fixed to isolate that structure. Write $\nabla F(0) = -\mu v$, where $\mu > 0$ and $\|v\| = 1$. The initial descent direction is v. Set $r = \operatorname{rank} W$ and $q = \operatorname{rank} P$. Norms are Euclidean; a multilinear norm is the supremum over unit inputs.

A positive profile $\alpha = (\alpha_0, \dots, \alpha_{N-1})$ satisfies $\sum_t \alpha_t = 1$. It partitions a total step mass B. A positive smooth gain ϕ, normalized by $\phi(1) = 1$, scales the gradient according to its squared norm. The two programs become

$$\begin{aligned} u_0 &= \bar{u}_0 = 0, \\ u_{t+1} &= u_t - \frac{B\alpha_t}{\mu}\phi(\|\nabla F(u_t)\|^2/\mu^2)\nabla F(u_t), \\ \bar{u}_{t+1} &= \bar{u}_t - \frac{B\alpha_t}{\mu}\phi(\|W\nabla F(\bar{u}_t)\|^2/\mu^2)W\nabla F(\bar{u}_t). \end{aligned} \tag{1}$$

Ordinary GD has $\phi \equiv 1$ and $\eta_t = B\alpha_t/\mu$. Unit normalization has $\phi(s) = s^{-1/2}$. Smooth normalization has $\phi_\delta(s) = \sqrt{(1+\delta)/(s+\delta)}$, $\delta \ge 0$. When $Wv = v$, all three start with the same speed. The local path length is bounded by cB for a gain-dependent constant c; it equals B for unit normalization. Changing N at fixed B changes the budget partition.

Denote the terminal response by $\Delta^\phi_\alpha(B) = F(\bar{u}_N) - F(u_N)$. The subscript N alone denotes the equal profile $(1/N, \dots, 1/N)$. Each update evaluates the gradient at a new state. A Taylor expansion

describes which additional directions this creates. The Hessian $Q = D^2F(0)$ gives the first change in the gradient; higher derivatives describe how that change itself varies. Define the vector-valued derivative contractions by

$$\langle \mathsf{T}_3[a,b],c\rangle = D^3F(0)[a,b,c], \qquad \langle \mathsf{T}_4[a,b,c],e\rangle = D^4F(0)[a,b,c,e].$$

Write $[B^k]$ for the coefficient of B^k in a finite Taylor expansion and set $u^{[k]} = [B^k]u(B)$. Then

$$\begin{aligned}
[B]\nabla F(u(B)) &= Qu^{[1]},\\
[B^2]\nabla F(u(B)) &= Qu^{[2]} + \tfrac{1}{2}\mathsf{T}_3[u^{[1]},u^{[1]}],\\
[B^3]\nabla F(u(B)) &= Qu^{[3]} + \mathsf{T}_3[u^{[1]},u^{[2]}] + \tfrac{1}{6}\mathsf{T}_4[u^{[1]},u^{[1]},u^{[1]}].
\end{aligned}$$

The first two directions form $E_2 = \operatorname{span}\{v, Qv\}$. The next coefficients introduce Q^2v and $\mathsf{T}_3[v,v]$. A term $a^2b/2$, for instance, can introduce a b component through the third derivative even when all Q^kv remain on the a axis. The first two nonlinear spaces are therefore

$$\begin{aligned}
E_3 &= \operatorname{span}\{v, Qv, Q^2v, \mathsf{T}_3[v,v]\},\\
E_4 &= E_3 + \operatorname{span}\{Q^3v, Q\mathsf{T}_3[v,v], \mathsf{T}_3[v,Qv], \mathsf{T}_4[v,v,v]\}.
\end{aligned} \tag{2}$$

The certificate tests whether the retained space contains every vector in this span.

Higher orders follow the same rule: apply a loss derivative to directions already generated at lower orders. We call each nested derivative operation a word. The layer index records derivative nesting; the execution length counts updates. Let $\mathsf{T}_{\ell+1} = D^\ell\nabla F(0)$ and $\mathcal{W}_1 = \{v\}$. A derivative word is one nested contraction in the finite list

$$\mathcal{W}_k = \bigcup_{\substack{\ell\ge 1,\ k_i\ge 1\\ 1+\sum_i k_i = k}} \{\mu^{-1}\mathsf{T}_{\ell+1}[w_1,\dots,w_\ell] : w_i \in \mathcal{W}_{k_i}\}, \qquad E_k = \operatorname{span}\bigcup_{i\le k}\mathcal{W}_i. \tag{3}$$

Symmetric permutations are listed once; linear dependencies are allowed. This recovers the displayed E_3, E_4. The relevant gain parameter is $\vartheta = -2\phi'(1)$: it is 0 for GD, 1 for normalization and $1/(1+\delta)$ for smooth normalization. The theorem identifies exactly how it enters the leading response. For a vector tuple w, its Gram matrix has entries $\operatorname{Gram}(w)_{ik} = \langle w_i, w_k\rangle$. It is positive semidefinite and vanishes exactly when every w_i does. Thus it records the size and alignment of all omitted directions at once.

The first part of the theorem maps this Gram matrix to a scalar response. The remaining parts reverse the implication for carefully chosen profiles. Here $o(B^p)$ means that the response divided by B^p tends to zero. The cases $j = 2, 3$ concern the next directions beyond E_2 and E_3, and appear at orders B^5 and B^7, respectively.

Theorem 1 (Graded radial response and transferable profiles)**.** *Fix $j \ge 2$. Assume $F \in C^{j+1}$, $\phi \in C^{j+1}$, and $E_j \subseteq \operatorname{ran} W$. Let $w^{[j]} = (\mu Pu : u \in \mathcal{W}_{j+1})$ and $G^{[j]} = \operatorname{Gram}(w^{[j]})$. There are profile-polynomial matrices satisfying*

$$\begin{aligned}
\Delta^\phi_\alpha(B) &= \frac{B^{2j+1}}{\mu}\operatorname{tr}(C_{\alpha,j}(\vartheta)G^{[j]}) + O(B^{2j+1}m_j(B)),\\
C_{\alpha,j}(\vartheta) &= (1-\vartheta)C_{\alpha,j}(0) + \vartheta C_{\alpha,j}(1),
\end{aligned} \tag{4}$$

where $m_j(B) = B + \omega_{j+1}(cB)$ and $\omega_s(h) = \sup_{\|x\|\le h}\|D^sF(x) - D^sF(0)\|$. The constants are uniform over finite positive profiles. If E_{j+1} is retained, the remainder is at most $CB^{2j+1}m_j(B)^2$.

For the following converses, assume $Wv = v$ and $0 \le \vartheta \le 1$. With $F \in C^3$ and a C^4 gain,

$$\Delta^\phi_2 = o(B^5), \quad \Delta^\phi_5 = o(B^5) \iff E_3 \subseteq \operatorname{ran} W \iff \sup_{N,\alpha}|\Delta^\phi_\alpha(B)| = o(B^5). \tag{5}$$

With $F \in C^4$, put $\alpha^a = (12,7,58,21,2)/100$ and $\alpha^b = (2,9,1,1,35,2)/50$. Then

$$\max\{|\Delta^\phi_{\alpha^a}|, |\Delta^\phi_{\alpha^b}|\} = o(B^7) \iff E_4 \subseteq \operatorname{ran} W \iff \sup_{N,\alpha}|\Delta^\phi_\alpha(B)| = o(B^7). \tag{6}$$

Both statements hold in a fixed positive metric after coordinate transformation.

The matrix $C_{\alpha,j}$ contains the weights imposed by the execution; $G^{[j]}$ contains the unknown omitted effects. Their trace pairing can vanish with $G^{[j]} \neq 0$. The selected profiles rule out that ambiguity.

The order B^{2j+1} comes from two omitted factors and one step. For the full update map Φ_t in (1), put $g_t = \nabla F(\bar{u}_t)$. The terminal sensitivity propagated backward along the reference is

$$\lambda_N = g_N, \qquad \lambda_t = D\Phi_t(\bar{u}_t)^\top \lambda_{t+1}.$$

Retaining E_j removes the normal coefficients below degree j:

$$Pg_t = O(B^j), \qquad P\lambda_{t+1} = O(B^j), \qquad B\langle Pg_t, P\lambda_{t+1}\rangle = O(B^{2j+1}).$$

The leading forward coefficients are gain-independent. The backward recurrence is affine in ϑ after lower retained directions are projected out. This yields the matrix law (4); the derivative recurrences and remainder bounds appear in Appendix B.3.

At the first nonlinear layer, after $PQv = 0$, write

$$X = PQ^2v/\mu, \quad Y = P\mathsf{T}_3[v,v], \qquad G_3 = \mathrm{Gram}(X,Y) \succeq 0.$$

After E_3 is retained, the next omitted tuple is

$$w = (PQ^3v/\mu^2,\ PQ\mathsf{T}_3[v,v]/\mu,\ P\mathsf{T}_3[v,Qv]/\mu,\ P\mathsf{T}_4[v,v,v]), \qquad G_4 = \mathrm{Gram}(w).$$

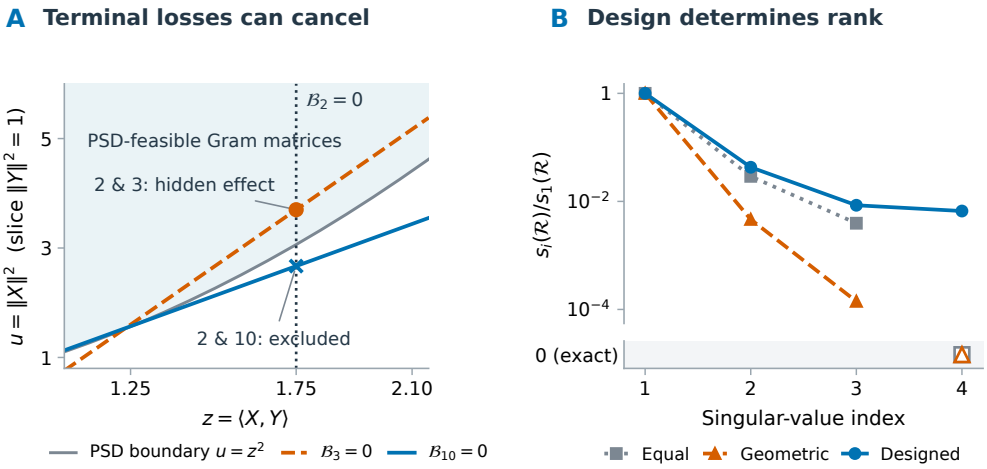

Figure 1: Equal observation counts can reveal different information. (A) On $\|Y\|^2 = 1$, feasible Gram matrices satisfy $u = \|X\|^2 \geq z^2$, where $z = \langle X, Y\rangle$. The common zero of $\mathcal{B}_2, \mathcal{B}_3$ is feasible; that of $\mathcal{B}_2, \mathcal{B}_{10}$ is excluded. With one omitted dimension, only the boundary is feasible. (B) Singular values of the four-prefix matrix $\mathcal{R}$. Equal and geometric $(1,2,4,8)/15$ steps leave one combination invisible. The design $(14,4,11,1)/30$ detects all four combinations. Exact zeros occupy a separate strip below the logarithmic scale.

In these bases, denote the response matrices by A_α and C_α. They inherit the affine law (4). The designs in the theorem give

$$5^5 A_5 + 131\, 2^5 A_2 \succ 0, \qquad -40A_a + 123A_b \succ 0, \qquad -3C_a + 14C_b \succ 0 \quad (0 \leq \vartheta \leq 1).$$

For a positive-definite combination H and an omitted Gram matrix G,

$$\text{zero responses} \implies 0 = \mathrm{tr}(HG) \geq \lambda_{\min}(H)\,\mathrm{tr}\,G \implies G = 0.$$

The cubic response first removes PQv. The positive combinations then remove G_3 and G_4 successively. Their endpoint calculations are in Appendix B. Each profile specifies a response function of B. The containment tests use the asymptotic order of these functions.

Figure 1A makes the cancellation concrete. At the normalized endpoint, define the equal-step leading coefficient $\mathcal{B}_N = \mu N^5 \lim_{B\downarrow 0} \Delta_N(B)/B^5$ after $PQv = 0$. On the slice $\|Y\|^2 = 1$, put $u = \|X\|^2$ and $z = \langle X, Y\rangle$. Positive semidefiniteness requires $u \geq z^2$. The two zero-response lines for $N = 2, 3$ meet inside this region. When $r, q \geq 2$, those observations are compatible with a nonzero omitted pair. For $N = 2, 10$, the intersection lies outside it, so the same ambiguity is excluded. With one omitted dimension only its boundary is realizable. The exact integer and dimension classification appears with its proof in Corollary 1.

5 Checkpoint observations and finite transfer

The first theorem resolves cancellation by combining terminal responses from multiple executions as functions of the budget. A saved state allows a more local question: how much can the loss decrease if we briefly move along the omitted gradient? The answer bounds the omitted force at that state directly. Put $r(y) = P\nabla F(y)$ and $y^- = y - \sigma r(y)$. If $\|D^2 F\| \leq L$ along this correction segment, then

$$\sigma(1 - L\sigma/2)\|r(y)\|^2 \leq D(y) := F(y) - F(y^-) \leq \sigma(1 + L\sigma/2)\|r(y)\|^2, \qquad 0 < \sigma < 2/L. \tag{7}$$

This is the standard descent estimate (Bubeck, 2015, Lemma 3.4). Its positive lower coefficient makes a small measured decrease imply a small omitted gradient. Unlike a terminal response, this observation cannot vanish through a signed cancellation.

A prefix is the reference state after the first t steps. Let $s_t = \sum_{i<t} \alpha_i$. Let $e_{t,j}$ be the degree-j elementary symmetric polynomial in $\alpha_0, \ldots, \alpha_{t-1}$. After the lower directions are retained, the new normal rows are

$$a_t = (-e_{t,2}, s_t^2/2), \qquad f_t = (e_{t,3}, -\tfrac{1}{2}\sum_{i<t} \alpha_i s_i^2, -s_t e_{t,2}, s_t^3/6). \tag{8}$$

Specifically, for $r_t(B) = P\nabla F(\bar{u}_t(B))$,

$$E_2 \subseteq \operatorname{ran} W \implies r_t(B) = B^2 a_t \cdot (X, Y) + o(B^2),$$
$$E_3 \subseteq \operatorname{ran} W \implies r_t(B) = B^3 f_t \cdot w + o(B^3).$$

The rows are independent of the gain. Stacking four rows $f_1, \ldots, f_4$ gives a matrix $\mathcal{R}$. Its nullspace consists of omitted combinations invisible to all four prefixes. The theorem asks whether those observations leave any such combination undetected. The polynomial condition below is precisely the full-rank test for this matrix.

Theorem 2 (Checkpoint design by normal-word rank)**.** *Use the retained program* (1)*. Assume a positive C^4 gain with $\phi(1) = 1$. All correction segments must have a common admissible length σ. Initial containment may be given or inferred from the energy premise.*

For $F \in C^3$, two positive steps and their two noninitial prefixes satisfy

$$\max_{t=1,2} D(\bar{u}_t(B)) = o(B^4) \quad \Longleftrightarrow \quad E_3 \subseteq \operatorname{ran} W.$$

For $F \in C^4$, use four positive fractions (a, b, c, d) summing to one. In the dimension-free worst case,

$$\max_{1\leq t\leq 4} D(\bar{u}_t(B)) = o(B^6) \quad \Longleftrightarrow \quad E_4 \subseteq \operatorname{ran} W$$

holds exactly when

$$a^3bd - a^3c^2 + a^2b^2d - 2a^2bc^2 + a^2bcd + ab^3c - 2ab^2c^2 + b^4c \neq 0. \tag{9}$$

Four prescribed energies are necessary in this observation class, even when E_3 is already retained. The profile $(14, 4, 11, 1)/30$ satisfies the condition.

The first two rows have determinant $a^3b/2$. The four-row matrix $\mathcal{R}$ has determinant $a^4bcd/12$ times (9). Appendix D expands this determinant and gives the realizability proof. Figure 1B compares three designs with the same four observations. Equal and geometric steps each have a zero singular value. Both designs retain a blind combination despite their different step spacings. The full-rank design removes that combination. Its smallest singular value also controls how strongly finite errors are amplified and hence the tightness of finite transfer.

The same matrix gives a finite transfer. Fix one loss and one projector with E_3 retained. Calibrate the four prefixes at budget b. Suppose their scalar reports give normal bounds R_i^{cal}. Assume a certified forward remainder $\|r_t(B) - B^3 f_t \cdot w\| \le C_n B^4$. Appendix D gives C_n from explicit derivative and gain bounds. For a target profile at budget B, set $\ell_t = f_t^{\mathrm{tar}} \mathcal{R}^{-1}$. The calibration and target remainders obey $\|\rho_i^{\mathrm{cal}}\| \le C_n b^4$ and $\|\rho_t^{\mathrm{tar}}\| \le C_n B^4$. Eliminating the common omitted tuple gives

$$
\begin{aligned}
r^{\mathrm{cal}} &= b^3 \mathcal{R} w + \rho^{\mathrm{cal}}, \qquad w = b^{-3} \mathcal{R}^{-1} (r^{\mathrm{cal}} - \rho^{\mathrm{cal}}), \\
r_t^{\mathrm{tar}} &= B^3 f_t^{\mathrm{tar}} w + \rho_t^{\mathrm{tar}} = (B/b)^3 \ell_t (r^{\mathrm{cal}} - \rho^{\mathrm{cal}}) + \rho_t^{\mathrm{tar}}.
\end{aligned}
$$

Taking norms yields

$$
\|r_t^{\mathrm{tar}}\| \le (B/b)^3 \sum_i |\ell_{t,i}| R_i^{\mathrm{cal}} + C_n B^4 + C_n B^3 b \|\ell_t\|_1 =: \widehat{R}_t. \tag{10}
$$

These bounds can replace target-step probes in the next certificate. They pay explicitly for conditioning and finite Taylor error. The loss and projector must stay fixed so that the tuple w is shared. Changing batches requires new residual bounds. Let c_{ref} and c_{probe} be the costs of a calibration step and a normal probe. Replacing N target probes reduces added audit work only if $4c_{\mathrm{ref}} + 4c_{\mathrm{probe}} < N c_{\mathrm{probe}}$.

6 Finite target certification

The observation design has now supplied a bound on the omitted gradient. Propagating this local discrepancy through later updates gives a bound on the final target error. For ordinary GD, the preceding fixed-loss transfer supplies these local bounds. With changing batches or refreshed subspaces, we instead obtain a bound at each stage. Let F_t be the batch loss at stage t. The full comparator uses ordinary SGD. A reference can refresh its projector or use an inexact update:

$$
\begin{aligned}
x_{t+1} &= x_t - \eta_t \nabla F_t(x_t), \\
y_{t+1} &= y_t - \eta_t \Pi_t \nabla F_t(y_t) + \xi_t, \qquad \|\xi_t\| \le d_t.
\end{aligned} \tag{11}
$$

Both programs use the same realized batches and learning rates. The orthogonal projector Π_t can depend on reference history. The evaluation objective V may differ from every F_t. The proposition takes residual bounds and known smoothness as inputs. It returns a state-error bound E_t and a target-error bound U_V. In particular, $U_V \le \tau$ certifies adequacy for this execution.

Proposition 1 (Finite scalar input–output certificate)**.** *Assume Hessian bounds L_t on the comparison and probe domains. At each probed reference state, obtain an upper endpoint D_t^+ for the normal-correction loss decrease. Use $0 < \sigma_t < 2/L_t$ and define*

$$
R_t = \sqrt{\frac{\max(0, D_t^+)}{\sigma_t (1 - L_t \sigma_t / 2)}}, \qquad E_{t+1} = q_t E_t + \eta_t R_t + d_t. \tag{12}
$$

The inputs satisfy $E_0 \ge \|x_0 - y_0\|$. The number q_t bounds the Lipschitz constant of $\Phi_t(x) = x - \eta_t \nabla F_t(x)$. At unprobed stages supply any valid omitted-gradient bound R_t. If the propagated balls remain in the certified domains, then

$$
\|x_t - y_t\| \le E_t, \qquad |V(x_N) - V(y_N)| \le G_V E_N + L_V E_N^2 / 2 =: U_V, \tag{13}
$$

where $G_V \ge \|\nabla V(y_N)\|$ and $L_V \ge \sup \|D^2 V\|$. A terminal scalar correction using V can supply G_V. Inexact directions use (66)*. Measurement error is added to the final interval.*

One may use $q_t = 1 + \eta_t L_t$. When $m_t I \preceq D^2 F_t \preceq L_t I$, the tighter choice is $\max\{|1 - \eta_t m_t|, |1 - \eta_t L_t|\}$.

The propagation follows directly from the local discrepancy. Put $e_t = x_t - y_t$ and $n_t = (I - \Pi_t) \nabla F_t(y_t)$. Subtracting the updates gives

$$
\begin{aligned}
e_{t+1} &= \Phi_t(x_t) - \Phi_t(y_t) - \eta_t n_t - \xi_t, \\
\|e_{t+1}\| &\le q_t \|e_t\| + \eta_t \|n_t\| + \|\xi_t\| \le q_t E_t + \eta_t R_t + d_t = E_{t+1}, \\
E_N &= E_0 \prod_{k<N} q_k + \sum_{t<N} (\eta_t R_t + d_t) \prod_{t<k<N} q_k.
\end{aligned}
$$

At the terminal state, Taylor integration yields

$$V(x_N) - V(y_N) = \langle \nabla V(y_N), e_N \rangle + \int_0^1 (1-s) e_N^\top D^2 V(y_N + s e_N) e_N \, ds,$$
$$|V(x_N) - V(y_N)| \le G_V \|e_N\| + \tfrac{1}{2} L_V \|e_N\|^2 \le G_V E_N + \tfrac{1}{2} L_V E_N^2.$$

This argument holds for each shared batch stream. Appendix E includes the scalar, inexact-direction and measurement-error terms.

The norm bound discards the direction of each discrepancy. It can therefore remain too large even when every residual is measured accurately. A reference linearization retains that direction. Let $A_t = D\Phi_t(y_t)$ and $r_t = \Phi_t(y_t) - y_{t+1}$. Compute a linear correction with error at most δ_t:

$$z_0 = 0, \qquad z_{t+1} \approx A_t z_t + r_t, \qquad c_t = y_t + z_t.$$

All derivative bounds must cover $[y_t, c_t]$ and the radius-w_t balls about c_t, including the target bounds at the final stage. For $\|D^2\Phi_t\| \le m_t^\Phi$, propagate

$$w_0 \ge \|x_0 - c_0\|, \qquad w_{t+1} = q_t w_t + \tfrac{1}{2} m_t^\Phi \|z_t\|^2 + \delta_t. \tag{14}$$

Indeed the local linearization error satisfies

$$\|\Phi_t(c_t) - c_{t+1}\| \le \|\Phi_t(y_t + z_t) - \Phi_t(y_t) - A_t z_t\| + \delta_t \le \tfrac{1}{2} m_t^\Phi \|z_t\|^2 + \delta_t,$$
$$\|x_{t+1} - c_{t+1}\| \le q_t \|x_t - c_t\| + \|\Phi_t(c_t) - c_{t+1}\| \le w_{t+1}.$$

The signed bias has center $V(y_N) - V(c_N)$ and radius $(G_V + L_V\|z_N\|)w_N + L_V w_N^2/2$. For SGD this costs a reference Hessian–vector action per step and an extra terminal loss. It also needs a third-derivative bound. Appendix F proves the corrected residual recurrence. These derivative actions are additional to the scalar-only reports.

For bilinear adapters, ξ_t accounts for the factor-induced metric and cross-increment discrepancies; Appendix E gives their decomposition.

7 Finite evidence and execution costs

The empirical question is whether more informative residual propagation changes the adequacy decision for the same restricted execution. We compare three certificates on identical stored trajectories: scalar correction reports, direct residual norms and centered linear corrections. The first comparison isolates the conversion from a loss decrease to a norm. The second retains directional information and pays for additional derivative actions.

We adapt a logistic head on fixed RoBERTa-large features. There are 64 training and 32 separate validation examples. Each of two batch streams uses 64 SGD steps with batch size eight. Candidate dimensions are $8, 16, 32, 48$. Their bases refresh every eight steps from fixed principal directions and two past reference gradients. The tolerance is fixed before the centered calculation, with target values available at selection. The feature dimension is 1024. The training row space has dimension at most 64; its numerical rank is 63. A quadratic regularizer gives $\frac{1}{2} I \preceq D^2 F_t \preceq I$ globally. The validation Hessian is bounded by $0.26I$. With $\eta = 1/8$, the comparator contracts by $15/16$. These global bounds make every comparison domain valid.

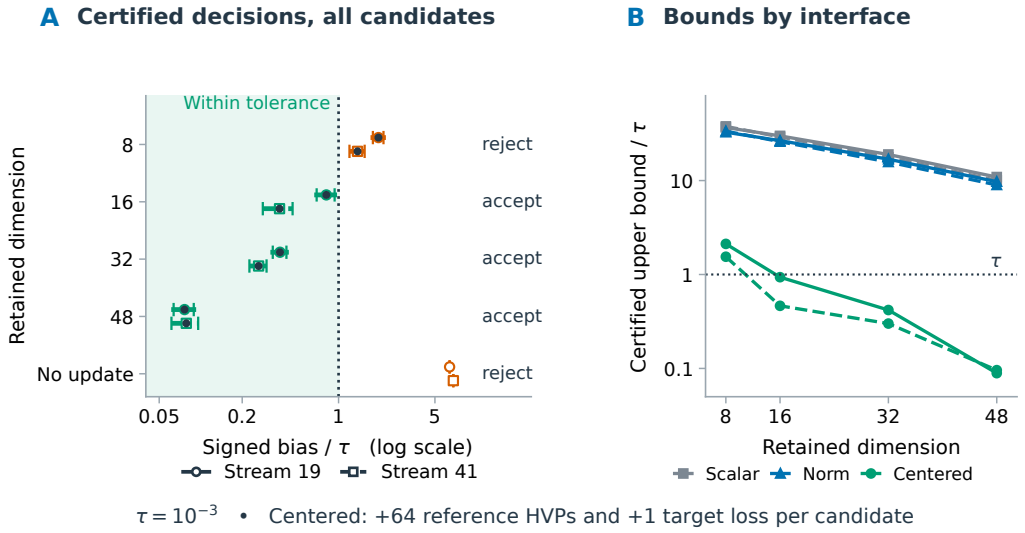

Figure 2: Directional propagation changes the certification decision. (A) Deterministic numerical enclosures of the target bias. Circles/squares distinguish batch streams; dark dots show stored target values. Intervals inside the shaded tolerance region certify adequacy; intervals entirely to its right certify failure. Zero-update intervals use the same certified full-loss enclosures. (B) Upper bounds divided by $\tau = 10^{-3}$. Solid/dashed lines denote streams 19/41. The norm-based bounds all exceed one. The figure lists the added reference Hessian–vector product (HVP) and loss costs.

Figure 2B compares two inputs to the same norm-based propagation bound. The scalar and direct-norm bounds differ modestly and all exceed $\tau = 10^{-3}$. The centered calculation retains the propagated direction and evaluates the target at the corrected state. In panel A, both streams' intervals accept dimensions 16, 32 and 48; they reject dimension 8 and zero update. Thus 16 is the smallest certified dimension among the tested candidates. Rejecting zero update confirms that the test requires preserving the target effect of adaptation.

The certificate uses reference states, scalar probes and the linear correction. Outward intervals enclose nonlinear losses, and arithmetic bounds account for correction errors. Appendix G gives the derivative bounds, numerical enclosures and cost counts. Each candidate requires 64 reference gradients, eight basis constructions, 64 normal probes and one validation probe. The centered additions are listed in Figure 2. Direct norm access avoids the extra trial loss evaluations. Stagewise probes supply the residual bounds for these changing-batch executions.

8 Conclusion and Future Directions

Finite adaptation makes a choice about which directions to retain. Assessing that choice requires observations that reveal the omitted effects and bounds that connect them to the desired target. The response law explains why terminal losses can cancel and how designed profiles exclude that ambiguity. Checkpoint energies give a positive alternative whose information is determined by rank. Finite residual propagation completes the connection to a tolerance decision. The stored-data example shows why preserving directional information can matter even after residual norms are known.

Future work can optimize calibration profiles jointly with their preparation cost. Tighter propagation under weaker contraction and for multi-step factor updates would broaden the range of certifiable executions. Comparing these extensions at a common target tolerance would quantify their tradeoff against direct norm access.

AI USE STATEMENT

We used generative AI tools to assist with manuscript writing and experimental code development. Manuscript assistance covered organization, language revision, literature and reference checks, and presentation of numerical results. Mathematical assistance included developing and checking derivations and writing proof arguments. Coding assistance covered experiment implementation, numerical verification scripts, and figure-generation code. We have reviewed and verified all AI-assisted content, including the manuscript and experimental code. Checks included explicit derivations, verification against original references, and numerical consistency tests. The authors retain responsibility for the final text, mathematical claims, code, and reported results.

ETHICS STATEMENT

This work concerns finite-step adaptation theory and numerical certificates. The numerical study reuses SST-2 examples and fixed RoBERTa-large features. Reuse of these resources should preserve attribution and follow their access and redistribution conditions. Applications of the certificates should retain the stated comparison and assumptions. Fairness, privacy, and deployment safety require evaluation appropriate to the downstream task.

REPRODUCIBILITY STATEMENT

Sections 3–7 specify the update rules, observation interfaces, and target quantities. Appendices A–F and H give the assumptions, derivations, and finite-error bounds. Appendix G describes the feature preprocessing, batch streams, subspace refresh, derivative constants, interval arithmetic, and reference linearization used in Figure 2. The scalar, direct-norm, and centered certificates share the same trajectories, fixing the comparison for the reported tolerance decisions.

REFERENCES

Roland Becker and Rolf Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica*, 10:1–102, 2001. doi: 10.1017/S0962492901000010.

Peter Benner and Tobias Breiten. Two-sided projection methods for nonlinear model order reduction. *SIAM Journal on Scientific Computing*, 37(2):B239–B260, 2015. doi: 10.1137/14097255X.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3–4):231–357, 2015. doi: 10.1561/2200000050. URL https://arxiv.org/abs/1405.4980.

John C. Butcher, Taketomo Mitsui, Yuto Miyatake, and Shun Sato. On the B-series composition theorem. *BIT Numerical Mathematics*, 65(2), 2025. doi: 10.1007/s10543-025-01063-y. URL https://link.springer.com/article/10.1007/s10543-025-01063-y. Article 22.

Emmanuel J. Candès, Thomas Strohmer, and Vladislav Voroninski. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013. doi: 10.1002/cpa.21432.

Yongchang Hao, Yanshuai Cao, and Lili Mou. Flora: Low-rank adapters are secretly gradient compressors. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17554–17571. PMLR, 2024. URL https://proceedings.mlr.press/v235/hao24a.html.

Yutong He, Pengrui Li, Yipeng Hu, Chuyan Chen, and Kun Yuan. Subspace optimization for large language models with convergence guarantees. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 22468–22522. PMLR, 2025. URL https://proceedings.mlr.press/v267/he25i.html.

Robert Hermann and Arthur J. Krener. Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, 22(5):728–740, 1977. doi: 10.1109/TAC.1977.1101601.

Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, 2 edition, 2002. doi: 10.1137/1.9780898718027.

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://arxiv.org/abs/2106.09685.

Alexandre Janon, Maëlle Nodet, and Clémentine Prieur. Goal-oriented error estimation for the reduced basis method, with application to sensitivity analysis. *Journal of Scientific Computing*, 68 (1):21–41, 2016. doi: 10.1007/s10915-015-0127-z.

Shiva Prasad Kasiviswanathan. SGD with low-dimensional gradients with applications to private and distributed learning. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 1905–1915. PMLR, 2021. URL https://proceedings.mlr.press/v161/kasiviswanathan21a.html.

Chunyuan Li, Heerad Farkhoor, Rosanne Liu, and Jason Yosinski. Measuring the intrinsic dimension of objective landscapes. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1804.08838.

Yuchen Lou, Zeqi Ye, and Minshuo Chen. Parameter-efficient subspace optimization for LLM fine-tuning. *arXiv preprint arXiv:2512.02216*, 2026. URL https://arxiv.org/abs/2512.02216v2. Version 2.

Robert I. McLachlan, Klas Modin, Hans Munthe-Kaas, and Olivier Verdier. Butcher series: A story of rooted trees and numerical methods for evolution equations. *Asia Pacific Mathematics Newsletter*, 7(1):1–11, 2017. URL https://arxiv.org/abs/1512.00906.

M. Milanese and Antonio Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991. doi: 10.1016/0005-1098(91)90134-N.

Mark Schmidt, Nicolas Le Roux, and Francis Bach. Convergence rates of inexact proximal-gradient methods for convex optimization. In *Advances in Neural Information Processing Systems*, volume 24, pp. 1458–1466, 2011. URL https://proceedings.neurips.cc/paper/2011/hash/8f7d807e1f53eff5f9efbe5cb81090fb-Abstract.html.

Steffen Schotthöfer, Emanuele Zangrando, Gianluca Ceruti, Francesco Tudisco, and Jonas Kusch. GeoLoRA: Geometric integration for parameter efficient fine-tuning. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/7ccae68f81c1bd9104c304a9d8967048-Abstract-Conference.html.

Henk J. van Waarde, Jaap Eising, Harry L. Trentelman, and M. Kanat Camlibel. Data informativity: A new perspective on data-driven analysis and control. *IEEE Transactions on Automatic Control*, 65(11):4753–4768, 2020. doi: 10.1109/TAC.2020.2966717.

Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996. doi: 10.1137/1038003.

Shaowen Wang, Linxi Yu, and Jian Li. LoRA-GA: Low-rank adaptation with gradient approximation. In *Advances in Neural Information Processing Systems*, volume 37, pp. 54905–54931. Curran Associates, Inc., 2024. doi: 10.52202/079017-1741. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/62c4718cc334f6a0a62fb81c4a2095a1-Abstract-Conference.html.

Zhengbo Wang, Jian Liang, Ran He, Zilei Wang, and Tieniu Tan. LoRA-Pro: Are low-rank adapters properly optimized? In *International Conference on Learning Representations*, pp. 93787–93808, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/ea184f920a0f0f8d8030aa1bd7ac9fd4-Abstract-Conference.html.

Yufeng Yang, Erin E. Tripp, Yifan Sun, Shaofeng Zou, and Yi Zhou. Adaptive gradient normalization and independent sampling for (stochastic) generalized-smooth optimization. *Transactions on Machine Learning Research*, July 2025. URL https://openreview.net/forum?id=KKSQQMlEfw.

Yuanhe Zhang, Fanghui Liu, and Yudong Chen. LoRA-One: One-step full gradient could suffice for fine-tuning large language models, provably and efficiently. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 75513–75574. PMLR, 2025. URL https://proceedings.mlr.press/v267/zhang25ax.html.

Jiawei Zhao, Zhenyu Zhang, Beidi Chen, Zhangyang Wang, Anima Anandkumar, and Yuandong Tian. GaLore: Memory-efficient LLM training by gradient low-rank projection. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 61121–61143. PMLR, 2024. URL https://proceedings.mlr.press/v235/zhao24s.html.

APPENDIX

Appendix A supplies the common estimates. Appendices B–D prove the response and observation results. Appendices E–G give the finite target bounds and numerical calculations. Appendix H treats paired terminal-loss observations.

APPENDIX CONTENTS

A Taylor remainders and residual identities . . . 14
A.1 Uniform forward and adjoint remainders . . . 14
A.2 A signed residual identity . . . 16
A.3 The normalized defect and its leading order . . . 17
B Radial responses and transferable certificates . . . 18
B.1 Forward and backward coefficients . . . 18
B.2 The signed response and the low-regularity remainder . . . 20
B.3 The gain-affine law at every finite order . . . 21
B.4 Successive elimination of the omitted layers . . . 23
B.5 Fixed metrics and linear coordinates . . . 24
C Sharp classification at the normalized endpoint . . . 25
C.1 Coefficients and the integer-pair kernel . . . 25
C.2 The equal-step kernel and a thirteen-step alternative . . . 29
C.3 Realization by a scalar potential . . . 30
D Checkpoint rank and finite observation transfer . . . 31
D.1 Correction energy and the normal gradient . . . 31
D.2 The exact rank criterion . . . 31
D.3 A finite bound at unobserved target steps . . . 33
E Target-bias bounds from normal residuals . . . 34
E.1 From scalar residuals to target bias . . . 34
E.2 Inexact correction directions . . . 35
E.3 The actual bilinear update . . . 35
E.4 Execution costs . . . 36
F A centered finite certificate . . . 36
F.1 Linear correction and the remaining error . . . 37
F.2 Derivative inputs and the tolerance decision . . . 37
G Target certificates for the feature-head experiment . . . 38
G.1 Feature model and derivative bounds . . . 38
G.2 Stored programs and scalar intervals . . . 39
G.3 Verified reference linearization . . . 39
H Finite bounds from paired terminal losses . . . 40
H.1 Observation intervals and the finite guarantee . . . 40
H.2 Common derivative and remainder bounds . . . 41
H.3 The defect inferred from a lower-order response . . . 43
H.4 The higher-order response remainder . . . 43
H.5 Proof of the finite input–output guarantee . . . 45
H.6 Initial containment from scalar observations . . . 46
H.7 A rank certificate from 53-bit interval observations . . . 47

A TAYLOR REMAINDERS AND RESIDUAL IDENTITIES

The response expansions use uniform Taylor remainders. The finite target bounds use the signed residual identity in Section A.2. Both estimates allow the reference and comparator to use different updates.

A.1 UNIFORM FORWARD AND ADJOINT REMAINDERS

B-series provide the elementary-differential calculus for composed updates (Butcher et al., 2025; McLachlan et al., 2017). The lemma below derives the finite-smoothness remainders and uniform constants used here.

Lemma 1 (Uniform forward and adjoint remainders)**.** *Let $p \geq 2$. On a convex ball of radius R, assume $f, f_{\mathrm{a}} \in C^{p+1}$, $V \in C^{p+2}$, and*

$$\max\{\sup \|D^j f\|, \sup \|D^j f_{\mathrm{a}}\|\} \leq D_j \quad (0 \leq j \leq p+1),$$
$$\sup \|D^j V\| \leq M_j \quad (1 \leq j \leq p+2).$$

For $\alpha_t > 0$, $\sum_t \alpha_t = 1$, consider $x_{t+1} = x_t + B\alpha_t f(x_t)$, $x_0 = 0$. The adjoint uses $I + B\alpha_t Df_{\mathrm{a}}(x_t)$ along these states. The fields may differ. The degree-p forward and adjoint expansions have $O(B^{p+1})$ remainders. The constants below are uniform in the number of steps.

Proof. Write $[z^j]H$ for a polynomial coefficient and $[H]_{\leq p}$ for truncation. Define

$$a_1 = D_0, \quad a_j = [z^{j-1}] \sum_{k=1}^{j-1} \frac{D_k}{k!} \left(\sum_{i=1}^{j-1} a_i z^i \right)^k, \quad A(z) = \sum_{j=1}^{p} a_j z^j, \quad s_t = \sum_{i<t} \alpha_i.$$

Let $T_k f$ be the Taylor polynomial at zero. Set $Z_0 = 0$ and $Z_{t+1} = [Z_t + B\alpha_t T_{p-1} f(Z_t)]_{\leq p}$. For $Z_{t,j} = [B^j] Z_t$, the coefficient induction is

$$Z_{t,1} = s_t f(0), \qquad \|Z_{t,1}\| \leq s_t a_1,$$
$$Z_{t+1,j} - Z_{t,j} = \alpha_t \sum_{k=1}^{j-1} \frac{1}{k!} \sum_{\substack{i_1+\cdots+i_k=j-1 \\ i_\ell \geq 1}} D^k f(0)[Z_{t,i_1}, \ldots, Z_{t,i_k}],$$
$$\|Z_{t+1,j} - Z_{t,j}\| \leq \alpha_t \sum_{k=1}^{j-1} \frac{D_k}{k!} \sum_{i_1+\cdots+i_k=j-1} \prod_{\ell=1}^{k} a_{i_\ell} = \alpha_t a_j,$$
$$\|Z_{t,j}\| \leq \sum_{i<t} \alpha_i a_j = s_t a_j \leq a_j.$$

Choose $B_0 > 0$ with $D_0 B_0 \leq R$ and $A(B_0) \leq R$. Put $K = A(B_0)/B_0$. For a nonnegative-coefficient polynomial H, set

$$\mathcal{T}_{p+1}(H) = \sum_{j \geq p+1} [z^j] H(z) B_0^{j-p-1}.$$

For $0 < B \leq B_0$,

$$\|x_t\| \leq B D_0 s_t \leq R, \qquad \|Z_t\| \leq A(B) \leq KB \leq R,$$
$$\|f(Z_t) - T_{p-1} f(Z_t)\| \overset{\text{Taylor}}{\leq} \frac{D_p}{p!} \|Z_t\|^p \leq \frac{D_p K^p}{p!} B^p.$$

Define

$$D_{\mathrm{poly}} = \mathcal{T}_{p+1}\left(z \sum_{k=0}^{p-1} \frac{D_k}{k!} A(z)^k \right), \qquad D_{\mathrm{step}} = D_{\mathrm{poly}} + D_p K^p / p!.$$

The one-step residual and state error $e_t = \|x_t - Z_t\|$ satisfy

$$\|Z_{t+1} - Z_t - B\alpha_t f(Z_t)\| \le \alpha_t D_{\mathrm{poly}} B^{p+1} + B\alpha_t \frac{D_p K^p B^p}{p!} = \alpha_t D_{\mathrm{step}} B^{p+1},$$

$$e_{t+1} \le (1 + B\alpha_t D_1) e_t + \alpha_t D_{\mathrm{step}} B^{p+1},$$

$$e_t \le D_{\mathrm{step}} B^{p+1} \sum_{i<t} \alpha_i \prod_{i<k<t} (1 + B\alpha_k D_1)$$

$$\le D_{\mathrm{step}} B^{p+1} \sum_{i<t} \alpha_i \exp\left(B D_1 \sum_{i<k<t} \alpha_k \right) \le e^{D_1 B_0} D_{\mathrm{step}} B^{p+1}.$$

Thus

$$C_Z = e^{D_1 B_0} D_{\mathrm{step}}, \qquad \max_t \|x_t - Z_t\| \le C_Z B^{p+1}. \tag{15}$$

For the terminal polynomial $[T_p V(Z_N)]_{\le p}$,

$$|V(x_N) - [T_p V(Z_N)]_{\le p}| \le |V(x_N) - V(Z_N)| + |V(Z_N) - T_p V(Z_N)| + |T_p V(Z_N) - [T_p V(Z_N)]_{\le p}|$$

$$\le \left\{ M_1 C_Z + \frac{M_{p+1} K^{p+1}}{(p+1)!} + \mathcal{T}_{p+1}\left(\sum_{k=1}^{p} \frac{M_k}{k!} A(z)^k \right) \right\} B^{p+1}.$$

Denote the coefficient in braces by $E_{p,V}$:

$$E_{p,V} = M_1 C_Z + \mathcal{T}_{p+1}\left(\sum_{k=1}^{p} \frac{M_k}{k!} A(z)^k \right) + \frac{M_{p+1} K^{p+1}}{(p+1)!}. \tag{16}$$

For the adjoint, let

$$\lambda_N = \nabla V(x_N), \qquad \lambda_t = (I + B\alpha_t D f_{\mathrm{a}}(x_t))^\top \lambda_{t+1}.$$

Its coefficients $L_{t,j}$ are defined by terminal Taylor expansion and

$$L_{t,j} = L_{t+1,j} + \alpha_t \sum_{k=0}^{j-1} H_{t,k}^\top L_{t+1,j-1-k}, \qquad H_{t,k} = [B^k] D f_{\mathrm{a}}(Z_t).$$

Set

$$H(z) = \sum_{k=0}^{p} \frac{D_{k+1}}{k!} A(z)^k, \quad T(z) = \sum_{k=0}^{p} \frac{M_{k+1}}{k!} A(z)^k, \quad h_j = [z^j]H, \quad t_j = [z^j]T,$$

$$l_0 = M_1, \qquad l_j = t_j + \sum_{k=0}^{j-1} h_k l_{j-1-k}, \qquad L(z) = \sum_{j=0}^{p} l_j z^j, \quad A_\lambda = L(B_0).$$

The backward coefficient induction gives

$$\|L_{N,j}\| \le t_j, \qquad \|H_{t,k}\| \le h_k,$$

$$\|L_{t,j}\| \le \|L_{N,j}\| + \sum_{i=t}^{N-1} \alpha_i \sum_{k=0}^{j-1} \|H_{i,k}\| \|L_{i+1,j-1-k}\|$$

$$\le t_j + \sum_{i=t}^{N-1} \alpha_i \sum_{k=0}^{j-1} h_k l_{j-1-k} \le l_j.$$

For $\Lambda_t = \sum_{j=0}^{p} L_{t,j} B^j$, define

$$C_A = \mathcal{T}_{p+1}\left(z \sum_{k=0}^{p-1} \frac{D_{k+1}}{k!} A(z)^k L(z) \right) + D_{p+1} K^p A_\lambda / p!,$$

$$C_T = M_2 C_Z + \mathcal{T}_{p+1}(T) + M_{p+2} K^{p+1} / (p+1)!. \tag{17}$$

Taylor's integral remainder and the discarded coefficient sums give

$$\|\Lambda_t - (I + B\alpha_t Df_{\mathrm{a}}(Z_t))^\top \Lambda_{t+1}\| \le \alpha_t C_A B^{p+1}.$$

At the terminal state,

$$\begin{aligned}\|\lambda_N - \Lambda_N\| &\le M_2\|x_N - Z_N\| + \left\{\mathcal{T}_{p+1}(T) + \frac{M_{p+2}K^{p+1}}{(p+1)!}\right\} B^{p+1} \\ &\le C_T B^{p+1}.\end{aligned}$$

With $\varepsilon_t = \|\lambda_t - \Lambda_t\|$,

$$\begin{aligned}\varepsilon_t &\le (1 + B\alpha_t D_1)\varepsilon_{t+1} + \alpha_t(C_A + B_0 D_2 A_\lambda C_Z)B^{p+1}, \\ \varepsilon_t &\le \prod_{k=t}^{N-1}(1 + B\alpha_k D_1)\varepsilon_N \\ &\qquad + B^{p+1}(C_A + B_0 D_2 A_\lambda C_Z)\sum_{i=t}^{N-1}\alpha_i \prod_{k=t}^{i-1}(1 + B\alpha_k D_1) \\ &\le e^{D_1 B_0}(C_T + C_A + B_0 D_2 A_\lambda C_Z)B^{p+1}.\end{aligned}$$

All sums use $\sum_t \alpha_t = 1$; none contains a reciprocal step fraction. □

A.2 A SIGNED RESIDUAL IDENTITY

Let $x_{t+1} = \Phi_t(x_t)$ and let y_t be computed reference states. On the comparison balls assume $\|D\Phi_t\| \le a_t$, $\|D^2\Phi_t\| \le m_t$, and $\|D^2 V\| \le L_V$. Define

$$e_t = x_t - y_t, \quad r_t = \Phi_t(y_t) - y_{t+1}, \quad A_t = D\Phi_t(y_t), \qquad E_{t+1} = a_t E_t + \|r_t\|, \quad E_0 \ge \|e_0\|.$$

The products below use the identity matrix when the index set is empty. The comparison tube follows from

$$\begin{aligned}e_{t+1} &= \Phi_t(y_t + e_t) - \Phi_t(y_t) + r_t, \\ \|e_{t+1}\| &\le \int_0^1 \|D\Phi_t(y_t + se_t)e_t\|\,ds + \|r_t\| \le a_t\|e_t\| + \|r_t\|, \\ \|e_t\| &\le E_0 \prod_{k<t} a_k + \sum_{i<t}\|r_i\| \prod_{i<k<t} a_k = E_t. \end{aligned} \tag{18}$$

Taylor's formula in integral form is

$$\begin{aligned}e_{t+1} &= A_t e_t + r_t + \rho_t, \qquad \rho_t = \int_0^1 (1-s)D^2\Phi_t(y_t + se_t)[e_t, e_t]\,ds, \\ V(x_N) - V(y_N) &= \nabla V(y_N)^\top e_N + \rho_V, \qquad \rho_V = \int_0^1 (1-s)D^2 V(y_N + se_N)[e_N, e_N]\,ds, \\ \|\rho_t\| &\le m_t E_t^2/2, \qquad |\rho_V| \le L_V E_N^2/2. \end{aligned} \tag{19}$$

For any vectors $\widehat{\lambda}_t$, set $b_t = A_t^\top \widehat{\lambda}_{t+1} - \widehat{\lambda}_t$ and $b_V = \nabla V(y_N) - \widehat{\lambda}_N$. Then

$$\begin{aligned}\widehat{\lambda}_{t+1}^\top e_{t+1} - \widehat{\lambda}_t^\top e_t &= \widehat{\lambda}_{t+1}^\top(A_t e_t + r_t + \rho_t) - \widehat{\lambda}_t^\top e_t \\ &= \widehat{\lambda}_{t+1}^\top(r_t + \rho_t) + b_t^\top e_t, \\ \widehat{\lambda}_N^\top e_N - \widehat{\lambda}_0^\top e_0 &= \sum_{t=0}^{N-1}(\widehat{\lambda}_{t+1}^\top e_{t+1} - \widehat{\lambda}_t^\top e_t) \\ &= \sum_{t=0}^{N-1}\{\widehat{\lambda}_{t+1}^\top r_t + \widehat{\lambda}_{t+1}^\top \rho_t + b_t^\top e_t\}.\end{aligned}$$

Consequently the single identity used below is

$$\begin{aligned} V(y_N) - V(x_N) = & -\widehat{\lambda}_0^\top e_0 - \sum_t \widehat{\lambda}_{t+1}^\top r_t \\ & - \sum_t (\widehat{\lambda}_{t+1}^\top \rho_t + b_t^\top e_t) - b_V^\top e_N - \rho_V. \end{aligned} \tag{20}$$

If $\|r_t - \widehat{r}_t\| \le \delta_t^r$, $\|b_t\| \le \delta_t^\lambda$ and $\|b_V\| \le \delta_V$, define $\widehat{C} = -\widehat{\lambda}_0^\top e_0 - \sum_t \widehat{\lambda}_{t+1}^\top \widehat{r}_t$. The complete error estimate is

$$\begin{aligned} |V(y_N) - V(x_N) - \widehat{C}| &\le \sum_t \|\widehat{\lambda}_{t+1}\|(\|r_t - \widehat{r}_t\| + \|\rho_t\|) + \sum_t \|b_t\|\|e_t\| + \|b_V\|\|e_N\| + |\rho_V| \\ &\le \sum_t \left[\|\widehat{\lambda}_{t+1}\| \left(\delta_t^r + \frac{m_t E_t^2}{2} \right) + \delta_t^\lambda E_t \right] + \delta_V E_N + \frac{L_V E_N^2}{2}. \end{aligned} \tag{21}$$

The derivation above gives a finite discrete residual estimate. Its adjoint weighting follows the goal-oriented estimation principle of Becker & Rannacher (2001). It also covers joint parameter–optimizer states when their maps satisfy the displayed derivative bounds.

A.3 The normalized defect and its leading order

Let $F \in C^3$ on a convex ball of radius r_{dom}. Assume $\nabla F(0) = -\mu v$, $\|v\| = 1$, $Wv = v$, $\|D^2 F\| \le L$, $\|D^3 F\| \le M_3$. Let $g_* > 0$ bound both gradient norms below. Use

$$h_t = B\alpha_t, \quad \sum_t h_t = B \le r_{\mathrm{dom}}, \quad G = \mu + L r_{\mathrm{dom}}, \quad p_t = W\nabla F(\bar{u}_t), \quad r_t = P\nabla F(\bar{u}_t),$$

$$q_t = \|p_t\|, \quad \nu_t = \sqrt{q_t^2 + \|r_t\|^2}, \quad \mathsf{f} = -\nabla F/\|\nabla F\|, \quad \bar{\mathsf{f}}_t = -p_t/q_t.$$

Here r_t denotes the normal gradient; the map residual in (20) is $h_t\delta_t$, where

$$\begin{aligned} \delta_t = \mathsf{f}(\bar{u}_t) - \bar{\mathsf{f}}_t &= -\frac{p_t + r_t}{\nu_t} + \frac{p_t}{q_t} = \left(\frac{q_t}{\nu_t} - 1 \right) \bar{\mathsf{f}}_t - \frac{r_t}{\nu_t}, \\ \|\delta_t\|^2 = 2(1 - q_t/\nu_t) &= \frac{2\|r_t\|^2}{\nu_t(\nu_t + q_t)} \le \frac{\|r_t\|^2}{g_*^2}. \end{aligned} \tag{22}$$

For $\mathcal{U}(g) = g/\|g\|$, $n = g/\nu$, and $z^\perp = (I - nn^\top)z$,

$$\begin{aligned} D\mathcal{U}(g)z &= z^\perp/\nu, \\ D^2\mathcal{U}(g)[z_1, z_2] &= -\nu^{-2}\{\langle n, z_1\rangle z_2^\perp + \langle n, z_2\rangle z_1^\perp + n\langle z_1^\perp, z_2^\perp\rangle\}, \\ \|D\mathsf{f}\| &\le \|D\mathcal{U}\|\|D^2 F\| \le L/g_* =: K_1, \\ \|D^2\mathsf{f}\| &\le \|D\mathcal{U}\|\|D^3 F\| + \|D^2\mathcal{U}\|\|D^2 F\|^2 \\ &\le M_3/g_* + 3L^2/g_*^2 =: K_2. \end{aligned}$$

Define

$$\lambda_N = \nabla F(\bar{u}_N), \qquad \lambda_t = (I + h_t D\mathsf{f}(\bar{u}_t))^\top \lambda_{t+1}. \tag{23}$$

Put

$$\begin{aligned} \mathcal{C}_W &= \sum_t h_t \left[\frac{\langle P\lambda_{t+1}, r_t\rangle}{\nu_t} + (1 - q_t/\nu_t)\langle \lambda_{t+1}, \bar{\mathsf{f}}_t\rangle \right], \\ S_W &= g_*^{-1} \sum_t h_t \|r_t\|, \quad C_{\mathrm{rem}} = \tfrac{1}{2} e^{2K_1 r_{\mathrm{dom}}} [L + GK_2 r_{\mathrm{dom}} e^{K_1 r_{\mathrm{dom}}}]. \end{aligned} \tag{24}$$

Lemma 2 (Normalized signed correction)**.** *For $J = F(u_N)$ and $J_W = F(\bar{u}_N)$,*

$$|J_W - J - \mathcal{C}_W| \le C_{\mathrm{rem}} S_W^2. \tag{25}$$

Proof. The required substitutions in (18) and (21) are

$$\Phi_t(x) = x + h_t \mathsf{f}(x), \quad a_t = 1 + h_t K_1, \quad m_t = h_t K_2, \quad \widehat{\lambda}_t = \lambda_t, \quad \delta_t^r = \delta_t^\lambda = \delta_V = 0, \quad e_0 = 0.$$

Substitution gives

$$\begin{aligned}
E_t &= \sum_{i<t} h_i \|\delta_i\| \prod_{i<k<t} (1 + h_k K_1) \le e^{K_1 B} S_W, \\
\|\lambda_t\| &\le G \prod_{i=t}^{N-1} (1 + h_i K_1) \le G e^{K_1 B}, \\
-\sum_t h_t \lambda_{t+1}^\top \delta_t &= \sum_t h_t \left[\frac{\langle P\lambda_{t+1}, r_t \rangle}{\nu_t} + (1 - q_t/\nu_t) \langle \lambda_{t+1}, \bar{\mathsf{f}}_t \rangle \right] = \mathcal{C}_W, \\
|J_W - J - \mathcal{C}_W| &\le \tfrac{1}{2} L E_N^2 + \tfrac{1}{2} \sum_t \|\lambda_{t+1}\| h_t K_2 E_t^2 \\
&\le \tfrac{1}{2} L e^{2K_1 B} S_W^2 + \tfrac{1}{2} G K_2 e^{3K_1 B} S_W^2 \sum_t h_t \le C_{\mathrm{rem}} S_W^2.
\end{aligned}$$

□

For equal steps $h = B/N$, suppose $r_t = h^j a_t + O(h^{j+1})$ and $P\lambda_{t+1} = h^j b_{t+1} + O(h^{j+1})$. At fixed N, substitution gives

$$\begin{aligned}
\nu_t^{-1} &= \mu^{-1} + O(h), & \langle \lambda_{t+1}, \bar{\mathsf{f}}_t \rangle &= -\mu + O(h), \\
1 - q_t/\nu_t &= \frac{\|r_t\|^2}{\nu_t(\nu_t + q_t)} = \frac{h^{2j}\|a_t\|^2}{2\mu^2} + O(h^{2j+1}), & S_W^2 &= O(h^{2j+2}), \\
\mathcal{C}_W &= \frac{h^{2j+1}}{\mu} \sum_t \left(\langle a_t, b_{t+1} \rangle - \frac{1}{2}\|a_t\|^2 \right) + O(h^{2j+2}).
\end{aligned}$$

Therefore

$$J_W - J = \frac{h^{2j+1}}{\mu} \left[\sum_t \langle a_t, b_{t+1} \rangle - \frac{1}{2} \sum_t \|a_t\|^2 \right] + O(h^{2j+2}). \tag{26}$$

B Radial responses and transferable certificates

This section proves Theorem 1. The coefficient calculations determine the leading response matrices. The remainder bounds then extend the polynomial calculation to the stated smoothness class. Set $\widehat{F} = F/\mu$, $H = Q/\mu$, $T = \mathsf{T}_3/\mu$, and $U = \mathsf{T}_4/\mu$. Then $\nabla \widehat{F}(0) = -v$. The rescaled vector field is $f(g) = -\phi(\|g\|^2)g$. A loss difference for $\widehat{F}$ is multiplied by μ to recover the physical loss difference.

B.1 Forward and backward coefficients

Let

$$s_t = \sum_{i<t} \alpha_i, \qquad e_{t,0} = 1, \qquad e_{0,k} = 0 \ (k > 0), \qquad e_{t+1,k} = e_{t,k} + \alpha_t e_{t,k-1}.$$

For the Taylor-polynomial loss, write $u_t^{[k]} = [B^k]\bar{u}_t$ and $g_t^{[k]} = [B^k]\nabla \widehat{F}(\bar{u}_t)$. Assume the lower words in each displayed formula are retained. Put $h_v = \langle v, Hv \rangle$ and $\vartheta = -2\phi'(1)$. The first two coefficients follow by multiplication:

$$\begin{aligned}
u_t^{[1]} &= s_t v, & g_t^{[1]} &= s_t H v, \\
[B]\phi(\|W g_t\|^2) &= \phi'(1)\, 2\langle -v, s_t H v \rangle = \vartheta s_t h_v, \\
u_{t+1}^{[2]} - u_t^{[2]} &= -\alpha_t (s_t H v - \vartheta s_t h_v v), \\
u_t^{[2]} &= -\left(\sum_{i<t} \alpha_i s_i \right) H v + \vartheta h_v \left(\sum_{i<t} \alpha_i s_i \right) v \\
&= -e_{t,2} H v + \vartheta h_v e_{t,2} v.
\end{aligned}$$

For $\gamma_t = [B^2]\phi(\|Wg_t\|^2)$, the next coefficient is

$$\begin{aligned} g_t^{[2]} &= Hu_t^{[2]} + \tfrac{1}{2}T[u_t^{[1]}, u_t^{[1]}] \\ &= -e_{t,2}H^2v + \vartheta h_v e_{t,2}Hv + \tfrac{1}{2}s_t^2 T[v,v], \\ u_{t+1}^{[3]} - u_t^{[3]} &= -\alpha_t\{g_t^{[2]} + \vartheta s_t h_v g_t^{[1]} - \gamma_t v\} \\ &= \alpha_t e_{t,2}H^2v - \tfrac{1}{2}\alpha_t s_t^2 T[v,v] - \vartheta h_v \alpha_t(e_{t,2}+s_t^2)Hv + \alpha_t\gamma_t v. \end{aligned}$$

Using $\sum_{i<t}\alpha_i e_{i,2} = e_{t,3}$ gives

$$\begin{aligned} u_t^{[2]} &\equiv -e_{t,2}Hv \pmod{\operatorname{span}\{v\}}, \\ u_t^{[3]} &\equiv e_{t,3}H^2v - \tfrac{1}{2}\sum_{i<t}\alpha_i s_i^2 T[v,v] \pmod{\operatorname{span}\{v,Hv\}}, \\ g_t^{[3]} &= Hu_t^{[3]} + T[u_t^{[1]}, u_t^{[2]}] + \tfrac{1}{6}U[u_t^{[1]}, u_t^{[1]}, u_t^{[1]}]. \end{aligned} \tag{27}$$

Define

$$\begin{aligned} a_t &= (-e_{t,2}, s_t^2/2), \\ f_t &= \left(e_{t,3}, -\tfrac{1}{2}\sum_{i<t}\alpha_i s_i^2, -s_t e_{t,2}, s_t^3/6\right). \end{aligned} \tag{28}$$

Projection and return to F yield

$$\begin{aligned} PQv = 0: &\quad [B^2]P\nabla F(\bar{u}_t) = a_t \cdot (X, Y), \\ E_3 \subseteq \operatorname{ran} W: &\quad [B^3]P\nabla F(\bar{u}_t) = f_t \cdot w. \end{aligned}$$

Every term containing a higher gain derivative in the preceding state formulas multiplies v or Hv; its projection is zero.

For the rescaled loss, put $g_t = \nabla\widehat{F}(\bar{u}_t)$ and

$$\lambda_N = g_N, \qquad K_t = \phi(\|g_t\|^2)\lambda_{t+1} + 2\phi'(\|g_t\|^2)g_t\langle g_t, \lambda_{t+1}\rangle, \qquad \lambda_t = \lambda_{t+1} - B\alpha_t D g_t K_t.$$

The first backward coefficients are generated by

$$\begin{aligned} b_N &= 1, & b_t - b_{t+1} &= (1-\vartheta)\alpha_t, \\ b_t &= 1 + (1-\vartheta)(1-s_t), & p_t &= b_{t+1} - \vartheta s_t, \\ U_N &= -e_{N,2}, & U_t - U_{t+1} &= -\alpha_t p_t, \\ V_N &= \tfrac{1}{2}, & V_t - V_{t+1} &= (1-\vartheta)\alpha_t s_t. \end{aligned} \tag{29}$$

The coefficients satisfy

$$\begin{aligned} K_t^{[0]} &= -(1-\vartheta)v, \\ \lambda_t^{[1]} &\equiv b_t Hv \pmod{\operatorname{span}\{v\}}, \\ K_t^{[1]} &\equiv p_t Hv \pmod{\operatorname{span}\{v\}}, \\ [B]Dg_t &= s_t T[v,\cdot], \\ \lambda_t^{[2]} - \lambda_{t+1}^{[2]} &= -\alpha_t\{HK_t^{[1]} + s_t T[v, K_t^{[0]}]\} \\ &\equiv -\alpha_t p_t H^2 v + (1-\vartheta)\alpha_t s_t T[v,v] \pmod{\operatorname{span}\{v,Hv\}}, \\ K_t^{[2]} &\equiv (U_{t+1} + \vartheta e_{t,2})H^2 v + (V_{t+1} - \vartheta s_t^2/2)T[v,v] \pmod{\operatorname{span}\{v,Hv\}}, \\ [B^2]Dg_t &\equiv -e_{t,2}T[Hv,\cdot] + \tfrac{1}{2}s_t^2 U[v,v,\cdot] \pmod{\operatorname{span}\{T[v,\cdot]\}}. \end{aligned}$$

When E_3 is retained, the normal third-degree increment is

$$\begin{aligned} P(\lambda_t^{[3]} - \lambda_{t+1}^{[3]}) &= -\alpha_t P\{HK_t^{[2]} + s_t T[v, K_t^{[1]}] + [B^2]Dg_t K_t^{[0]}\} \\ &= -\alpha_t P\big[(U_{t+1} + \vartheta e_{t,2})H^3 v + (V_{t+1} - \vartheta s_t^2/2)HT[v,v] \\ &\quad + (s_t p_t + (1-\vartheta)e_{t,2})T[v,Hv] - (1-\vartheta)s_t^2 U[v,v,v]/2\big]. \end{aligned}$$

Thus $\ell_{t+1}^{(2)} = (U_{t+1}, V_{t+1})$, and

$$\begin{aligned} \ell_N^{(3)} &= f_N, \\ \ell_t^{(3)} &= \ell_{t+1}^{(3)} - \alpha_t\big(U_{t+1} + \vartheta e_{t,2},\ V_{t+1} - \vartheta s_t^2/2,\ s_t p_t + (1-\vartheta)e_{t,2},\ -(1-\vartheta)s_t^2/2\big). \end{aligned} \tag{30}$$

B.2 THE SIGNED RESPONSE AND THE LOW-REGULARITY REMAINDER

At a retained state write $g = p + r$, $p = Wg$, $r = Pg$. Orthogonality and the fundamental theorem of calculus give

$$\begin{aligned} f(g) - f(p) &= -\phi(\|p\|^2 + \|r\|^2)(p + r) + \phi(\|p\|^2)p \\ &= -\phi(\|g\|^2)r - p\int_{\|p\|^2}^{\|p\|^2+\|r\|^2} \phi'(s)\, ds \\ &= -\phi(\|g\|^2)r - \|r\|^2 p \int_0^1 \phi'(\|p\|^2 + u\|r\|^2)\, du. \end{aligned} \tag{31}$$

Let K_1, K_2 bound the first two state derivatives of the full field. Let L_F, G_F bound the Hessian and gradient of $\widehat{F}$. Define

$$S = \sum_t B\alpha_t \|f(g_t) - f(p_t)\|,$$

$$\mathcal{E}_\alpha = \left| \Delta_\alpha^{\widehat{F}} + \sum_t B\alpha_t \langle \lambda_{t+1}, f(g_t) - f(p_t)\rangle \right|.$$

Then (18) and (21) give

$$\max_t \|u_t - \bar{u}_t\| \le e^{K_1 B} S, \qquad \max_t \|\lambda_t\| \le G_F e^{K_1 B},$$

$$\begin{aligned} \mathcal{E}_\alpha &\le \tfrac{1}{2} L_F e^{2K_1 B} S^2 + \tfrac{1}{2} G_F K_2 e^{3K_1 B} S^2 \sum_t B\alpha_t \\ &\le \tfrac{1}{2} e^{2K_1 B_0} (L_F + G_F K_2 B_0 e^{K_1 B_0}) S^2. \end{aligned}$$

These constants can be bounded from the gain. If $\|g\| \le G$ and $a_i \ge \sup |\phi^{(i)}|$, then

$$\|D(\phi(\|g\|^2)g)\| \le a_0 + 2a_1 G^2, \qquad \|D^2(\phi(\|g\|^2)g)\| \le 6a_1 G + 4a_2 G^3.$$

Composition with g gives K_1, K_2 from $\|Dg\|, \|D^2 g\|$.

Suppose the first normal degree is j. In physical units write $r_t^F = B^j f_t \cdot w + o(B^j)$ and $P\lambda_{t+1}^F = B^j \ell_{t+1} \cdot w + o(B^j)$. Since $\langle \lambda_{t+1}^{\widehat{F}}, p_t^{\widehat{F}} \rangle = 1 + O(B)$,

$$\begin{aligned} \Delta_\alpha^F &= \mu \sum_t B\alpha_t \left[\phi(\|g_t\|^2)\langle P\lambda_{t+1}, r_t\rangle + \|r_t\|^2 \langle \lambda_{t+1}, p_t\rangle \int_0^1 \phi'(\|p_t\|^2 + u\|r_t\|^2)\, du \right] + O(\mu S^2) \\ &= \frac{B^{2j+1}}{\mu} \sum_t \alpha_t \left[\langle f_t \cdot w, \ell_{t+1} \cdot w\rangle - \frac{\vartheta}{2}\|f_t \cdot w\|^2 \right] + o(B^{2j+1}). \end{aligned}$$

Thus

$$\Delta_\alpha^F = \frac{B^{2j+1}}{\mu} \sum_t \alpha_t \left[\langle f_t \cdot w, \ell_{t+1} \cdot w\rangle - \frac{\vartheta}{2}\|f_t \cdot w\|^2 \right] + o(B^{2j+1}). \tag{32}$$

For $j = 2$ use $a_t, \ell_{t+1}^{(2)}, (X, Y)$. For $j = 3$ use $f_t, \ell_{t+1}^{(3)}, w$. Symmetrizing the bilinear terms gives

$$\begin{aligned} A_\alpha(\vartheta) &= \tfrac{1}{2} \sum_t \alpha_t (a_t^\top \ell_{t+1}^{(2)} + \ell_{t+1}^{(2)\top} a_t - \vartheta a_t^\top a_t), \\ C_\alpha(\vartheta) &= \tfrac{1}{2} \sum_t \alpha_t (f_t^\top \ell_{t+1}^{(3)} + \ell_{t+1}^{(3)\top} f_t - \vartheta f_t^\top f_t). \end{aligned} \tag{33}$$

We next justify the expansions for $F \in C^{j+1}$. For this paragraph use the physical loss and $F^0 = T_{j+1}F$. Let common speed, Hessian and gradient bounds be c, L, G. Let U_1, U_2 bound the first two gradient-space derivatives of the gain map. On the radius-cB ball, Taylor integration gives

$$\varepsilon_g(B) := \sup \|\nabla F - \nabla F^0\| \le \frac{(cB)^j}{j!}\omega_{j+1}(cB),$$

$$\varepsilon_H(B) := \sup \|D^2 F - D^2 F^0\| \le \frac{(cB)^{j-1}}{(j-1)!}\omega_{j+1}(cB),$$

$$\|f - f^0\| \le U_1 \varepsilon_g, \qquad \|Df - Df^0\| \le U_1 \varepsilon_H + U_2 L \varepsilon_g.$$

Let $e_t = \|\bar{u}_t - \bar{u}_t^0\|$ and $d_t = \|\lambda_t - \lambda_t^0\|$. With $\Lambda = Ge^{K_1 B_0}$,

$$\begin{aligned} e_{t+1} &\le (1 + K_1 B\alpha_t)e_t + B\alpha_t U_1 \varepsilon_g, \\ e_* := \max_t e_t &\le U_1 \varepsilon_g B \sum_i \alpha_i \prod_{k>i}(1 + K_1 B\alpha_k) \le BU_1 e^{K_1 B_0}\varepsilon_g, \\ d_N &\le Le_* + \varepsilon_g, \\ d_t &\le (1 + K_1 B\alpha_t)d_{t+1} + B\alpha_t \Lambda(K_2 e_* + U_1\varepsilon_H + U_2 L\varepsilon_g), \\ \max_t d_t &\le e^{K_1 B_0}\{Le_* + \varepsilon_g + B\Lambda(K_2 e_* + U_1\varepsilon_H + U_2 L\varepsilon_g)\}. \end{aligned}$$

Apply the shared coefficient bounds with the retained forward field, the full Jacobian field, and terminal function F^0. Both fields use common derivative bounds. With $p = j$ and polynomial derivative bounds M_k^0, its constants are

$$\begin{aligned} C_g^0 &= M_2^0 C_Z + \mathcal{T}_{j+1}\left(\sum_{\ell=0}^{j} \frac{M_{\ell+1}^0}{\ell!} A(z)^\ell\right), \\ C_\lambda^0 &= e^{D_1 B_0}(C_T + C_A + B_0 D_2 A_\lambda C_Z). \end{aligned}$$

The gradient polynomial has degree j, so its Taylor remainder is zero. Adding the transported errors gives

$$\begin{aligned} \|P\nabla F(\bar{u}_t) - B^j f_t \cdot w\| &\le C_g^0 B^{j+1} + Le_* + \varepsilon_g \le C_g B^j m_j(B), \\ \|P\lambda_{t+1} - B^j \ell_{t+1} \cdot w\| &\le C_\lambda^0 B^{j+1} + \max_t d_t \le C_\lambda B^j m_j(B), \qquad m_j(B) = B + \omega_{j+1}(cB). \end{aligned}$$

Write these two errors as ϵ_t, ζ_t. Their product error expands as

$$\langle B^j f_t \cdot w + \epsilon_t, B^j \ell_{t+1} \cdot w + \zeta_t\rangle - B^{2j}\langle f_t \cdot w, \ell_{t+1} \cdot w\rangle = B^j\langle f_t \cdot w, \zeta_t\rangle + B^j\langle \epsilon_t, \ell_{t+1} \cdot w\rangle + \langle \epsilon_t, \zeta_t\rangle.$$

Each gain multiplier changes by $O(B)$. The signed identity and $\sum_t B\alpha_t = B$ therefore yield

$$\left|\Delta_\alpha - \frac{B^{2j+1}}{\mu}\operatorname{tr}(C_{\alpha,j}G^{[j]})\right| \le CB^{2j+1}m_j(B).$$

If E_{j+1} is retained, both leading rows vanish. In that case

$$\begin{aligned} \sup_{N,\alpha} |\Delta_\alpha| &\le CB \max_t\{\|\epsilon_t\|\|\zeta_t\| + \|\epsilon_t\|^2\} + CB^2 \max_t \|\epsilon_t\|^2 \\ &\le CB^{2j+1}m_j(B)^2 = o(B^{2j+1}). \end{aligned} \tag{34}$$

B.3 The gain-affine law at every finite order

Use the physical loss. Set $g = \nabla F$, $g_0 = -\mu v$, $\kappa(s) = \phi(s/\mu^2)/\mu$, and $\mathcal{U}(g) = \kappa(\|g\|^2)g$. Thus

$$\kappa(\mu^2) = \mu^{-1}, \qquad 2\mu^2\kappa'(\mu^2) = -\vartheta/\mu, \qquad h_t = B\alpha_t.$$

Fix $j \ge 2$ with $E_j \subseteq \operatorname{ran} W$. All coefficients below are computed for $F^0 = T_{j+1}F$. Write $\bar{x}_t = \bar{u}_t$, $g_t = \nabla F^0(\bar{x}_t)$, and $\lambda_N = g_N$. Let $E_0 = \{0\}$ and denote equality modulo E_k by $\equiv_k$.

For $z_i \in E_{k_i}$, the word definition gives

$$D^\ell g(0)[z_1, \dots, z_\ell] \in E_{1+k_1+\cdots+k_\ell}. \tag{35}$$

The effect of a lower-order argument is explicit. If $x_i, y_i \in E_{k_i}$ and $x_i - y_i \in E_{k_i-1}$, multilinearity gives

$$\begin{aligned} D^\ell g(0)[x_1, \dots, x_\ell] - D^\ell g(0)[y_1, \dots, y_\ell] &= \sum_{m=1}^{\ell} D^\ell g(0)[y_1, \dots, y_{m-1}, x_m - y_m, x_{m+1}, \dots, x_\ell] \\ &\in E_{k_1+\cdots+k_\ell}. \end{aligned}$$

This identity is used for each coefficient difference.

Put $x_{t,k} = [B^k]\bar{x}_t$ and $g_{t,k} = [B^k]g_t$. For $k \geq 1$, Taylor multiplication gives

$$g_{t,k} = \sum_{\ell=1}^{k} \frac{1}{\ell!} \sum_{\substack{i_1+\cdots+i_\ell=k \\ i_m \geq 1}} D^\ell g(0)[x_{t,i_1}, \ldots, x_{t,i_\ell}]. \tag{36}$$

Let $c_{t,\ell} = [B^\ell]\kappa(\|Wg_t\|^2)$. For $1 \leq k \leq j$,

$$\begin{aligned} x_{t+1,k} - x_{t,k} &= -\alpha_t \sum_{\ell=0}^{k-1} c_{t,\ell} W g_{t,k-1-\ell}, \\ x_{t,1} = s_t v \in E_1, \qquad g_{t,k-1-\ell} \in E_{k-\ell}, \qquad W|_{E_k} &= I, \\ x_{t+1,k} - x_{t,k} &\equiv_{k-1} -\alpha_t g_{t,k-1}/\mu. \end{aligned}$$

Use a star for the same coefficients with $\kappa \equiv 1/\mu$. Starting from $x_{0,k} = x^*_{0,k} = 0$, the induction is

$$\begin{aligned} x_{t,i} - x^*_{t,i} \in E_{i-1}\ (i<k) &\implies g_{t,k-1} - g^*_{t,k-1} \in E_{k-1}, \\ (x_{t+1,k} - x^*_{t+1,k}) - (x_{t,k} - x^*_{t,k}) &\equiv_{k-1} -\alpha_t (g_{t,k-1} - g^*_{t,k-1})/\mu \equiv_{k-1} 0, \\ x_{t,k} - x^*_{t,k} &= \sum_{i<t} \{(x_{i+1,k} - x^*_{i+1,k}) - (x_{i,k} - x^*_{i,k})\} \in E_{k-1}, \\ g_{t,k} - g^*_{t,k} &\in E_k. \end{aligned}$$

Hence

$$x_{t,k} \equiv_{k-1} x^*_{t,k}, \qquad g_{t,k} \equiv_k g^*_{t,k} \quad (k \leq j). \tag{37}$$

The physical adjoint and its intermediate vector are

$$\eta_t = D\mathcal{U}(g_t)\lambda_{t+1}, \qquad \lambda_t = \lambda_{t+1} - B\alpha_t Dg(\bar{x}_t)\eta_t.$$

Let $\lambda_{t,k} = [B^k]\lambda_t$. Differentiation yields

$$\eta_t = \kappa(\|g_t\|^2)\lambda_{t+1} + 2\kappa'(\|g_t\|^2)\langle g_t, \lambda_{t+1}\rangle g_t.$$

If $\widetilde{c}_{t,s}$ and $d_{t,s}$ are the coefficients of the two scalar multipliers, then

$$\begin{aligned} \widetilde{c}_{t,0} &= 1/\mu, \\ d_{t,0} &= 2\kappa'(\mu^2)\langle g_0, g_0\rangle = -\vartheta/\mu, \\ [B^k]\eta_t &= \mu^{-1}(\lambda_{t+1,k} - \vartheta g_{t,k}) + \underbrace{\sum_{s=1}^{k} (\widetilde{c}_{t,s}\lambda_{t+1,k-s} + d_{t,s} g_{t,k-s})}_{\in E_k}. \end{aligned}$$

Thus

$$[B^k]\eta_t \equiv_k \mu^{-1}(\lambda_{t+1,k} - \vartheta g_{t,k}). \tag{38}$$

To write the backward quotient recurrence, define

$$\mathcal{H}_{t,r} z = \sum_{\ell=0}^{r} \frac{1}{\ell!} \sum_{\substack{i_1+\cdots+i_\ell=r \\ i_m \geq 1}} D^{\ell+1} g(0)[x^*_{t,i_1}, \ldots, x^*_{t,i_\ell}, z].$$

The empty sum of indices is allowed only for $r = \ell = 0$. Then $\mathcal{H}_{t,0} = Dg(0)$. For $z \in E_{\ell+1}$ and $r + \ell = k - 1$,

$$\begin{aligned} ([B^r]Dg(\bar{x}_t) - \mathcal{H}_{t,r})z \in E_k, \\ \mathcal{H}_{t,r} E_\ell \subseteq E_k, \\ \lambda_{t,k} - \lambda_{t+1,k} &= -\alpha_t \sum_{r+\ell=k-1} [B^r]Dg(\bar{x}_t)[B^\ell]\eta_t \\ &\equiv_k -\frac{\alpha_t}{\mu} \sum_{r+\ell=k-1} \mathcal{H}_{t,r}(\lambda_{t+1,\ell} - \vartheta g^*_{t,\ell}). \end{aligned}$$

Define representatives $\Lambda^{\vartheta}_{t,k}$ by the exact recurrence

$$\Lambda^{\vartheta}_{t,0} = g_0, \qquad \Lambda^{\vartheta}_{N,k} = g^*_{N,k},$$
$$\Lambda^{\vartheta}_{t,k} = \Lambda^{\vartheta}_{t+1,k} - \frac{\alpha_t}{\mu} \sum_{r+\ell=k-1} \mathcal{H}_{t,r}(\Lambda^{\vartheta}_{t+1,\ell} - \vartheta g^*_{t,\ell}).$$

The preceding differences give $\lambda_{t,k} - \Lambda^{\vartheta}_{t,k} \in E_k$. For the affine defect $D_{t,k} = \Lambda^{\vartheta}_{t,k} - (1-\vartheta)\Lambda^0_{t,k} - \vartheta\Lambda^1_{t,k}$,

$$D_{t,0} = 0, \qquad D_{N,k} = 0,$$
$$D_{t,k} - D_{t+1,k} = -\frac{\alpha_t}{\mu} \sum_{r+\ell=k-1} \mathcal{H}_{t,r} D_{t+1,\ell},$$
$$D_{t,k} = -\frac{1}{\mu} \sum_{i=t}^{N-1} \alpha_i \sum_{r+\ell=k-1} \mathcal{H}_{i,r} \underbrace{D_{i+1,\ell}}_{0 \text{ by induction in } k} = 0.$$

Therefore

$$\lambda_{t,k} \equiv_k (1-\vartheta)\lambda^{\mathrm{GD}}_{t,k} + \vartheta\lambda^{\mathrm{NGD}}_{t,k}. \tag{39}$$

At $k = j$, applying P removes E_j. With $w = (\mu P u : u \in \mathcal{W}_{j+1})$, write

$$Pg_{t,j} = a_t \cdot w, \qquad P\lambda_{t+1,j} = b_{t+1}(\vartheta) \cdot w, \qquad b_{t+1}(\vartheta) = (1-\vartheta)b_{t+1}(0) + \vartheta b_{t+1}(1).$$

Every row is generated by the finite recurrences above. The signed identity (32) now becomes

$$[B^{2j+1}]\Delta^{\phi}_{\alpha} = \frac{1}{\mu}\sum_t \alpha_t \left\{ \sum_{k,\ell} a_{t,k} b_{t+1,\ell}(\vartheta)\langle w_k, w_\ell\rangle - \frac{\vartheta}{2}\sum_{k,\ell} a_{t,k}a_{t,\ell}\langle w_k, w_\ell\rangle \right\}$$
$$= \mu^{-1}\operatorname{tr}(C_{\alpha,j}(\vartheta)G^{[j]}),$$

where

$$C_{\alpha,j}(\vartheta) = \frac{1}{2}\sum_t \alpha_t \{a_t^\top b_{t+1}(\vartheta) + b_{t+1}(\vartheta)^\top a_t - \vartheta a_t^\top a_t\}. \tag{40}$$

In particular,

$$C_{\alpha,j}(\vartheta) = (1-\vartheta)C_{\alpha,j}(0) + \vartheta C_{\alpha,j}(1).$$

The normal-factor remainder estimate in (34) supplies the stated remainder.

For fixed design weights γ_i and $c > 0$, let $H_\theta = \sum_i \gamma_i C_{\alpha_i,j}(\theta)$. For every vector z and $0 \le \theta \le 1$,

$$z^\top(H_\theta - cI)z = (1-\theta)z^\top(H_0 - cI)z + \theta z^\top(H_1 - cI)z.$$

It follows that

$$H_0, H_1 \succeq cI \iff H_\theta \succeq cI \quad (0 \le \theta \le 1).$$

The forward direction uses the displayed nonnegative sum. The reverse direction takes $\theta = 0, 1$. This matrix design uses standard semidefinite optimization (Vandenberghe & Boyd, 1996). Necessity for a fixed geometry additionally requires realizability of its Gram matrices; the normalized cases below supply those constructions.

B.4 Successive elimination of the omitted layers

The cubic coefficient is

$$\Delta_\alpha(B) = \frac{B^3}{\mu}\sum_t \alpha_t(s_t b_{t+1} - \vartheta s_t^2/2)\|PQv\|^2 + o(B^3). \tag{41}$$

For $0 \le \vartheta \le 1$ and at least two positive steps,

$$b_{t+1} \ge 1, \quad s_t b_{t+1} - \vartheta s_t^2/2 \ge s_t(1 - s_t/2), \quad \sum_t \alpha_t s_t(1 - s_t/2) > 0.$$

Thus $\Delta_\alpha = o(B^5)$ implies $0 = \lim_{B\downarrow 0} \Delta_\alpha/B^3 \geq \mu^{-1} \sum_t \alpha_t s_t(1 - s_t/2)\|PQv\|^2$, so $PQv = 0$.

For the fifth-order tests, set $H_{25}(\theta) = 5^5 A_5(\theta) + 131\, 2^5 A_2(\theta)$. The recurrence (33) gives

$$H_{25}(0) = \begin{pmatrix} 126 & -413/2 \\ -413/2 & 355 \end{pmatrix}, \qquad H_{25}(1) = \begin{pmatrix} 83 & -581/4 \\ -581/4 & 2063/8 \end{pmatrix}.$$

Their Sylvester tests are

$$126 > 0, \quad 126 \cdot 355 - (413/2)^2 = 8351/4 > 0,$$
$$83 > 0, \quad 83 \cdot 2063/8 - (581/4)^2 = 4897/16 > 0.$$

Let $c = \min_{\theta=0,1} \lambda_{\min}(H_{25}(\theta)) > 0$. For $\theta \in [0, 1]$,

$$H_{25}(\theta) = (1 - \theta)H_{25}(0) + \theta H_{25}(1) \succeq cI,$$
$$0 = \mu \lim_{B\downarrow 0} \frac{5^5 \Delta_5(B) + 131\, 2^5 \Delta_2(B)}{B^5}$$
$$= \operatorname{tr}(H_{25}(\theta)G_3) \geq c \operatorname{tr} G_3 = c(\|X\|^2 + \|Y\|^2).$$

Hence $X = Y = 0$. Conversely, their vanishing and $PQv = 0$ retain E_3, and (34) gives the uniform $o(B^5)$ bound.

For the stated profiles a, b, use $H_5 = -40A_a + 123A_b$ and $H_7 = -3C_a + 14C_b$. The following rational intervals enclose the $LDL^\top$ pivots. Each pivot is d_i/d_{i-1}, where $d_i = \det(H_{1:i,1:i})$ and $d_0 = 1$.

Matrix	θ	pivot 1	pivot 2	pivot 3	pivot 4
H_5	0	$(179, 180)10^{-3}$	$(427, 428)10^{-5}$	–	–
H_5	1	$(151, 152)10^{-3}$	$(348, 349)10^{-5}$	–	–
H_7	0	$(579, 580)10^{-7}$	$(171, 172)10^{-7}$	$(237, 238)10^{-5}$	$(582, 583)10^{-8}$
H_7	1	$(381, 382)10^{-7}$	$(187, 188)10^{-7}$	$(113, 114)10^{-5}$	$(202, 203)10^{-7}$

All entries are obtained from (33) by rational arithmetic. The elimination is

$$\Delta_a, \Delta_b = o(B^7) \Longrightarrow PQv = 0,$$
$$0 = \operatorname{tr}(H_5(\theta)G_3) \geq \lambda_{\min}(H_5(\theta)) \operatorname{tr} G_3 \Longrightarrow G_3 = 0,$$
$$0 = \operatorname{tr}(H_7(\theta)G_4) \geq \lambda_{\min}(H_7(\theta)) \operatorname{tr} G_4 \Longrightarrow G_4 = 0,$$
$$E_4 \subseteq \operatorname{ran} W \Longrightarrow \sup_{N,\alpha} |\Delta_\alpha| \leq CB^7[B + \omega_4(cB)]^2 = o(B^7).$$

The ordinary-GD equal-step coefficients are

$$a_N^0 = N(N-1)(N-2)(2N-3)(2N-1)/30,$$
$$b_N^0 = -N(N-1)(12N^3 - 31N^2 + 23N - 6)/48,$$
$$c_N^0 = N(N-1)(28N^3 - 37N^2 + 23N - 2)/240.$$

Together with Appendix C, these give $a_N(\theta) = (1 - \theta)a_N^0 + \theta a_N^1$, and the corresponding identities for b_N, c_N.

B.5 Fixed metrics and linear coordinates

Let $M \succ 0$, $\Pi^2 = \Pi$, and $\Pi^\top M = M\Pi$. Set $z = M^{1/2}(x - x_0)$ and $\widetilde{F}(z) = F(x_0 + M^{-1/2}z)$. Then

$$\nabla \widetilde{F} = M^{-1/2} \nabla F,$$
$$W = M^{1/2} \Pi M^{-1/2},$$
$$W^\top = M^{-1/2} \Pi^\top M^{1/2} = M^{1/2} \Pi M^{-1/2} = W,$$
$$W^2 = M^{1/2} \Pi^2 M^{-1/2} = W,$$
$$M^{1/2} M^{-1} \nabla F = \nabla \widetilde{F},$$
$$M^{1/2} \Pi M^{-1} \nabla F = W \nabla \widetilde{F}.$$

The full and retained metric updates are therefore exactly the two radial recursions in z. Their terminal losses are unchanged.

For $x = x_0 + Cb$, assume C has full column rank. Let $P_C = C(C^\top C)^{-1}C^\top$ and $M^{-1} = CC^\top + I - P_C$. On $\operatorname{ran} C \oplus \operatorname{ran} C^\perp$, M^{-1} is positive definite. Moreover,

$$\begin{aligned}\nabla_b F(x_0 + Cb) &= C^\top \nabla F(x),\\ x^+ - x &= C(b^+ - b) = -\eta CC^\top \nabla F(x),\\ P_C M^{-1} &= P_C CC^\top + P_C - P_C^2 = CC^\top,\\ x^+ &= x - \eta P_C M^{-1} \nabla F(x).\end{aligned}$$

The corresponding full comparator uses M^{-1}. It equals Euclidean GD only when $M = I$.

C Sharp classification at the normalized endpoint

Throughout this appendix $\phi(s) = s^{-1/2}$ and $\vartheta = 1$. We write $\Delta_\alpha = \Delta_\alpha^\phi$, $\mathsf{A}_\alpha = A_\alpha(1)$ and $\mathsf{G}_j = G_j$. Here $Wv = v$, as in the converse statements of Theorem 1.

The gain is now fixed. The remaining question is which equal-step horizons rule out a nonzero omitted Gram matrix. Fix (W, v, μ). Let $\mathcal{F}_s(W, v, \mu)$ contain the C^s losses with $\nabla F(0) = -\mu v$ and defined local programs. A pair is universal when

$$\forall F \in \mathcal{F}_3(W, v, \mu) : \ [\Delta_m = o(B^5), \Delta_n = o(B^5)] \Longrightarrow E_3 \subseteq \operatorname{ran} W. \tag{42}$$

Corollary 1 (Sharp normalized endpoint)**.** *For $r, q \geq 2$, a pair $2 \leq m < n$ is universal exactly for $(2, n \geq 5)$, $(3, n \geq 12)$, $(4, n \geq 29)$, or $(5, n \geq 114)$. For $r \geq 2, q = 1$, every distinct pair suffices. No single equal horizon suffices in that case. For $r = 1, q \geq 1$, every $N \geq 2$ suffices. For $q = 0$, no observation is needed.*

At the next layer, all equal horizons characterize E_3 containment together with $w_2 = 4w_1$, $w_3 = -w_1$, $w_4 = 0$. A nonzero such tuple is realizable exactly when $r \geq 3, q \geq 1$. The family $(2, 8, \alpha_*)$, $\alpha_* = (1/6, 1/3, 1/2)$, is another sufficient E_4 design.

C.1 Coefficients and the integer-pair kernel

The general recurrence gives

$$\mathsf{A}_N = N^{-5} \begin{pmatrix} \mathsf{a}_N & \mathsf{b}_N/2 \\ \mathsf{b}_N/2 & \mathsf{c}_N \end{pmatrix}, \qquad \mathcal{B}_N(X, Y) = \mathsf{a}_N \|X\|^2 + \mathsf{b}_N \langle X, Y\rangle + \mathsf{c}_N \|Y\|^2, \tag{43}$$

where

$$\begin{aligned}\mathsf{a}_N &= N(N-1)(N-2)(4N^2 - 3N - 2)/60,\\ \mathsf{b}_N &= -N^2(N-1)(3N^2 - 4N - 1)/24,\\ \mathsf{c}_N &= N(N-1)(2N-1)(7N^2 + 3N + 1)/240.\end{aligned}$$

For an explicit derivation, multiply the fifth-order rows by N^2. They become $\widehat{f}_t = -\binom{t}{2}X + t^2 Y/2$ and $\widehat{\ell}_{t+1} = \widehat{f}_N - \binom{N-t}{2}X$. Then

$$\mathcal{B}_N = \sum_{t<N} [\langle \widehat{f}_t, \widehat{f}_N - \binom{N-t}{2} X\rangle - \|\widehat{f}_t\|^2/2].$$

For $S_j = \sum_{t=0}^{N-1} t^j$, the required power sums are

$$\begin{aligned}S_1 &= \frac{N(N-1)}{2}, \qquad S_2 = \frac{N(N-1)(2N-1)}{6},\\ S_3 &= \frac{N^2(N-1)^2}{4}, \quad S_4 = \frac{N(N-1)(2N-1)(3N^2-3N-1)}{30}.\end{aligned}$$

Expansion of the two products in the preceding sum gives

$$\begin{aligned}\mathsf{a}_N &= [S_4 + (2-4N)S_3 + (4N^2-3)S_2 - (4N^2-4N)S_1]/8,\\ \mathsf{b}_N &= [(2N-2)S_3 + (-3N^2+2N)S_2 + N^2 S_1]/4,\\ \mathsf{c}_N &= N^2 S_2/4 - S_4/8.\end{aligned}$$

These expressions give the three coefficient polynomials above. The scalar form is indefinite for every integer $N \geq 2$:

$$\mathsf{a}_N\mathsf{c}_N - \mathsf{b}_N^2/4 = -\frac{N^2(N-1)^2(N+1)}{57600}(N^5 + 31N^4 - 33N^3 - 3N^2 + 16N + 16) < 0. \quad (44)$$

The polynomial in parentheses is positive because

$$\begin{aligned}&N^5 + 31N^4 - 33N^3 - 3N^2 + 16N + 16\\ &\quad = N^5 + N^3(31N - 33) - 3N^2 + 16N + 16\\ &\quad \geq N^5 + N^2(29N - 3) + 16N + 16 > 0.\end{aligned}$$

We prove the pair classification in Corollary 1. We first prove the $r, q \geq 2$ branch for the fixed-geometry quantifiers in (42), then handle the rank restrictions when either dimension is one. This also establishes the dimension-free worst-case statement. For every fixed $N \geq 2$, the degree-one normal factors in (26) are $tPQv$ and $NPQv$. Hence

$$\Delta_N(B) = \frac{(N-1)(4N+1)}{12\mu N^2}\|PQv\|^2 B^3 + o(B^3),$$

since $\sum_{t<N}(Nt - t^2/2) = N(N-1)(4N+1)/12$. Consequently

$$\Delta_N = o(B^5) \Longrightarrow 0 = \lim_{B\downarrow 0}\frac{\Delta_N(B)}{B^3} = \frac{(N-1)(4N+1)}{12\mu N^2}\|PQv\|^2 \Longrightarrow PQv = 0.$$

Let $\mathsf{r}_N = (\mathsf{a}_N, \mathsf{b}_N, \mathsf{c}_N)/(N(N-1))$. The two observations act on (G_{11}, G_{12}, G_{22}) by these rows. For $2 \leq m < n$, write $k = \mathsf{r}_m \times \mathsf{r}_n$ and $K = \left(\begin{smallmatrix} k_1 & k_2 \\ k_2 & k_3 \end{smallmatrix}\right)$. The first coordinate is

$$k_1 = \frac{n-m}{5760}[53m^2n^2 + 11mn(m+n) - 3(m^2+n^2) + 4(m+n) + 1] > 0.$$

Indeed $m^{-2} + n^{-2} \leq 1/2$, so

$$53m^2n^2 - 3(m^2+n^2) \geq (53 - 3/2)m^2n^2 > 0.$$

The kernel and positivity tests are therefore

$$k_1 > 0 \Longrightarrow k \neq 0 \Longrightarrow \ker\begin{pmatrix}\mathsf{r}_m\\ \mathsf{r}_n\end{pmatrix} = \operatorname{span}\{k\},$$

$$K \succeq 0 \Longleftrightarrow k_3 - k_2^2/k_1 \geq 0 \Longleftrightarrow \det K \geq 0,$$

$$G \succeq 0, \quad \mathsf{r}_m(G_{11}, G_{12}, G_{22})^\top = \mathsf{r}_n(G_{11}, G_{12}, G_{22})^\top = 0 \Longleftrightarrow G = tK \succeq 0 \quad (t \geq 0).$$

If $\det K < 0$, the last condition forces $t = 0$.

Direct substitution of the three coefficient polynomials gives

$$\det K = \frac{(m-n)^2}{8294400}\,\mathsf{P}(m,n),$$

where the following symmetric polynomial is displayed by powers of n:

$$\begin{aligned}\mathsf{P}(m,n) = &(m^4 + 59m^3 - 287m^2 - 372m - 108)n^4\\ &+ (59m^4 + 695m^3 + 313m^2 - 77m - 24)n^3\\ &+ (-287m^4 + 313m^3 + 300m^2 + 118m + 31)n^2\\ &+ (-372m^4 - 77m^3 + 118m^2 + 125m + 32)n\\ &- 108m^4 - 24m^3 + 31m^2 + 32m + 4.\end{aligned}$$

For $m = 2, 3, 4, 5$, set $t = n - m - 1 \geq 0$. The coefficients of $\mathsf{P}(m, m+1+t)$, in descending powers $t^4, \ldots, t^0$, are

m	t^4	t^3	t^2	t	1
2	-1512	-10566	-14067	31770	57375
3	-2133	-8022	96793	582610	805375
4	-2156	21140	592363	3162390	4964319
5	-1143	103734	1981971	11354310	20593215

Let $Q_m(t) = \mathsf{P}(m, m+1+t)$. Descartes' rule states that its positive-root count, with multiplicity, is at most its coefficient sign-change count. Here

$$\#\{\text{positive roots of } Q_m\} \le 1, \qquad Q_m(0) > 0, \qquad \lim_{t\to\infty} Q_m(t) = -\infty.$$

The intermediate value theorem gives a positive root; the count makes it unique. Its location follows from these exact evaluations:

m	n_0	$\mathsf{P}(m, n_0-1)$	$\mathsf{P}(m, n_0)$
2	5	63000	-44073
3	12	1753623	-1183025
4	29	44932500	-1006929
5	114	1181716983	-464355081

For all remaining pairs write $m = 6+x$, $n = 7+x+y$, where $x, y \ge 0$. Then $\mathsf{P}(6+x, 7+x+y) = \sum_{j=0}^{4} p_j(x) y^j$, with

$$\begin{aligned}
p_4(x) &= x^4 + 83x^3 + 991x^2 + 3420x + 1368,\\
p_3(x) &= 4x^5 + 419x^4 + 8399x^3 + 66995x^2 + 230947x + 275670,\\
p_2(x) &= 6x^6 + 759x^5 + 20497x^4 + 242623x^3\\
&\quad + 1456836x^2 + 4346059x + 5094073,\\
p_1(x) &= 4x^7 + 593x^6 + 19761x^5 + 303596x^4\\
&\quad + 2549441x^3 + 12116677x^2 + 30674064x + 32176714,\\
p_0(x) &= x^8 + 170x^7 + 6672x^6 + 124548x^5 + 1328047x^4\\
&\quad + 8581818x^3 + 33386952x^2 + 72175744x + 66753823.
\end{aligned}$$

Every p_j has positive coefficients. Hence

$$x, y \ge 0 \Longrightarrow \mathsf{P}(6+x, 7+x+y) = \sum_{j=0}^{4} p_j(x) y^j \ge p_0(0) > 0.$$

For $m \le 5$, the unique root lies between $n_0 - 1$ and n_0. Thus

$$\mathsf{P}(m, n) < 0 \iff (m, n) \in \{(2, n \ge 5), (3, n \ge 12), (4, n \ge 29), (5, n \ge 114)\}.$$

The positive-semidefinite counterexamples are realizable, not merely feasible matrix relaxations. For any vectors $X, Y \in \mathbb{R}^2$, use retained coordinates (x, s) and omitted coordinates $y \in \mathbb{R}^2$:

$$F(x, s, y) = -x + \frac{\kappa}{2}(x^2 + s^2 + \|y\|^2) + xs + s\langle X, y\rangle + \frac{x^2}{2}\langle Y, y\rangle.$$

At zero,

$$\begin{aligned}
&\nabla F = -e_x, && Qe_x = \kappa e_x + e_s, && Qe_s = e_x + \kappa e_s + X,\\
&Q^2 e_x = (\kappa^2+1)e_x + 2\kappa e_s + X, && PQe_x = 0, && PQ^2 e_x = X,\\
&\mathsf{T}_3[e_x, e_x] = Y.
\end{aligned}$$

If $K \succeq 0$, a concrete Gram realization is

$$X = \sqrt{k_1}\, e_1, \qquad Y = \frac{k_2}{\sqrt{k_1}} e_1 + \sqrt{k_3 - k_2^2/k_1}\, e_2, \qquad \mathrm{Gram}(X, Y) = K.$$

For $F = -x + \kappa\|\cdot\|^2/2 + F_{\text{cross}}$ on a fixed ball,

$$\kappa > \sup\|D^2 F_{\text{cross}}\| \Longrightarrow D^2 F \succeq (\kappa - \sup\|D^2 F_{\text{cross}}\|) I \succ 0.$$

This permits strongly convex witnesses. Their two responses obey

$$\operatorname{tr}(A_m K) = \operatorname{tr}(A_n K) = 0, \quad K \neq 0 \Longrightarrow \Delta_m, \Delta_n = o(B^5), \quad E_3 \nsubseteq \operatorname{ran} W.$$

If $\det K < 0$, the kernel test gives $G_3 = 0$. Together with $PQv = 0$,

$$E_3 \subseteq \operatorname{ran} W \Longrightarrow \sup_{N,\alpha} |\Delta_\alpha| \leq C B^5 [B + \omega_3(cB)]^2 = o(B^5).$$

Fixed dimensions and the rank constraint. We prove the dimension-dependent clauses of Corollary 1. Fix the geometry (W, v, μ), and write $r = \operatorname{rank} W$ and $q = \operatorname{rank} P$. Universality in this paragraph quantifies over $F \in \mathcal{F}_3(W, v, \mu)$, not over one observed loss. Once the positive cubic coefficient has forced $PQv = 0$, the feasible Gram matrices satisfy

$$\mathsf{G}_3 \succeq 0, \qquad \operatorname{rank} \mathsf{G}_3 \leq \min\{2, q\}. \tag{45}$$

The polynomial sign calculation above has the additional consequence

$$\mathsf{P}(m, n) \neq 0 \qquad (2 \leq m < n, \quad m, n \in \mathbb{Z}). \tag{46}$$

For $m = 2, 3, 4, 5$ its unique root in $n > m$ lies strictly between $n_0 - 1$ and n_0 in the displayed threshold table. For $m \geq 6$ the positive-coefficient expansion gives $\mathsf{P}(m, n) > 0$. Thus every nonzero common-kernel matrix tK has rank two. If $q = 1$,

$$G_3 = tK \neq 0 \Longrightarrow \operatorname{rank} G_3 = 2, \qquad \operatorname{rank} G_3 \leq q = 1.$$

The contradiction forces $t = 0$, hence $X = Y = 0$.

If $r = 1$, the earlier cubic implication is stronger. Symmetry of Q gives

$$Qv = \langle v, Qv\rangle v, \qquad Q^2 v = \langle v, Qv\rangle^2 v, \qquad X = 0.$$

Consequently, for each $N \geq 2$,

$$\Delta_N(B) = \frac{\mathsf{c}_N}{\mu N^5}\|Y\|^2 B^5 + o(B^5), \qquad \mathsf{c}_N > 0.$$

Therefore

$$\Delta_N = o(B^5) \Longrightarrow 0 = \mathsf{c}_N \|Y\|^2/(\mu N^5) \Longrightarrow Y = 0.$$

If $q = 0$, then $W = I$, $u_t = \bar{u}_t$ at every step, and $\Delta_\alpha = 0$.

For $r \geq 2$ and $q \geq 1$, no single equal horizon is universal. The negative determinant in (44) gives a nonzero real scalar pair (x_N, y_N) with

$$\mathsf{a}_N x_N^2 + \mathsf{b}_N x_N y_N + \mathsf{c}_N y_N^2 = 0.$$

For $N = 2$, one may take $(x_N, y_N) = (1, 0)$; for $N > 2$ the quadratic in x_N/y_N has positive discriminant. The displayed scalar-potential realization with one omitted coordinate realizes $X = x_N e_y, Y = y_N e_y$, $PQv = 0$, and a nonzero omitted derivative. It has $o(B^5)$ bias for that horizon but fails containment. For $r \geq 2, q \geq 2$, every excluded pair has $K \succ 0$, so two omitted coordinates are both sufficient and necessary to realize its nonzero common Gram kernel. This proves the threshold classification for every fixed geometry of those dimensions. To embed any witness in a prescribed (W, v), choose one retained unit vector perpendicular to v and the required omitted unit vectors, apply the corresponding orthogonal coordinate map, and add decoupled quadratics on unused coordinates. Multiplying the loss by the prescribed $\mu > 0$ leaves normalized trajectories unchanged and scales their biases by μ. These operations preserve each zero coefficient and each failed containment condition. For $r = 1, q \geq 1$, a one-step profile carries no information: both first updates equal Bv, whereas the loss $-\mu x + x^2 y/2$ can have $Y \neq 0$.

C.2 THE EQUAL-STEP KERNEL AND A THIRTEEN-STEP ALTERNATIVE

For the unit-normalized endpoint, define $C_N = N^7 C_{(1/N,\dots,1/N)}(1)$. The general rows in (28) and (30) reduce, after rescaling by N^3, to

$$f_t = \left(\binom{t}{3}, -\frac{t(t-1)(2t-1)}{12}, -\frac{t^2(t-1)}{2}, \frac{t^3}{6}\right),$$
$$\ell_{t+1,N} = f_N + \sum_{s=t+1}^{N-1} ((N-1)(N-s), -(N^2-s^2)/2, -s(N-s), 0),$$
$$C_N = \tfrac{1}{2}\sum_{t<N} (f_t^\top \ell_{t+1,N} + \ell_{t+1,N}^\top f_t - f_t^\top f_t).$$

Every entry is a finite sum of rational polynomials. Define

$$R_4 = \begin{pmatrix} -4 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad S_4 = \begin{pmatrix} 151/8 & 155/4 & -193/24 \\ 155/4 & 163/2 & -223/12 \\ -193/24 & -223/12 & 161/24 \end{pmatrix}.$$

Substitution in those sums gives the exact matrix identity

$$868C_2 - 625C_3 + 297C_4 - 82C_5 + 10C_6 = R_4^\top S_4 R_4. \tag{47}$$

The leading principal minors of $S_4 - I/8$ are $75/4$, $775/32$, $16561/4608$. Sylvester's criterion therefore gives $S_4 \succeq I/8$. For $\gamma = (868, -625, 297, -82, 10)$,

$$\sum_{N=2}^{6} \gamma_N \operatorname{tr}(C_N G_4) = \operatorname{tr}(R_4^\top S_4 R_4 G_4) = \sum_{\ell} (R_4 w^{(\ell)})^\top S_4 (R_4 w^{(\ell)})$$
$$\geq \tfrac{1}{8}\sum_{\ell} \|R_4 w^{(\ell)}\|^2 = \tfrac{1}{8}\{\|w_2 - 4w_1\|^2 + \|w_3 + w_1\|^2 + \|w_4\|^2\}.$$

Conversely,

$$f_t(1, 4, -1, 0)^\top = \tfrac{1}{6} t(t-1)\{(t-2) - 2(2t-1) + 3t\} = 0.$$

Thus

$$\Delta_N = o(B^7) \text{ for every } N \Longrightarrow \Delta_2, \Delta_5 = o(B^5) \Longrightarrow E_3 \subseteq \operatorname{ran} W,$$
$$\operatorname{tr}(C_N G_4) = 0 \text{ for every } N \Longrightarrow R_4 w = 0 \Longleftrightarrow (w_2, w_3, w_4) = (4w_1, -w_1, 0),$$
$$R_4 w = 0 \Longrightarrow f_t \cdot w = 0 \Longrightarrow \operatorname{tr}(C_N G_4) = 0 \quad \text{for every } N,$$
$$\Delta_N(B) = \frac{B^7}{\mu N^7} \operatorname{tr}(C_N G_4) + O(B^7[B + \omega_4(cB)]) = o(B^7).$$

For the dimension threshold, first

$$Qv = av \Longrightarrow Q^3 v = a^3 v \Longrightarrow w_1 = 0.$$

Otherwise v, Qv are independent. If $r \leq 2$ and E_3 is retained,

$$\operatorname{span}\{v, Qv\} \subseteq \operatorname{ran} W, \quad r \leq 2 \Longrightarrow \operatorname{ran} W = \operatorname{span}\{v, Qv\},$$
$$Q \operatorname{ran} W = \operatorname{span}\{Qv, Q^2 v\} \subseteq \operatorname{ran} W \Longrightarrow PQ^3 v = 0 \Longrightarrow w_1 = 0,$$
$$w_1 = 0, \quad R_4 w = 0 \Longrightarrow w_1 = w_2 = w_3 = w_4 = 0.$$

For $r \geq 3, q \geq 1$, the fixed polynomial

$$F(x, s, z, y) = 20 - x + 10(x^2 + s^2 + z^2 + y^2) + xs + sz + zy + 2x^2 z - xsy \tag{48}$$

has the following derivatives for $v = e_x$:

$$Qv = 20e_x + e_s, \qquad Q^2 v = 401e_x + 40e_s + e_z, \qquad PQ^3 v = e_y,$$
$$\mathsf{T}_3[v, v] = 4e_z, \qquad PQ\mathsf{T}_3[v, v] = 4e_y,$$
$$P\mathsf{T}_3[v, Qv] = P(80e_z - e_y) = -e_y, \qquad \mathsf{T}_4[v, v, v] = 0.$$

The (x, s, z) projection retains E_3, and $w = (e_y, 4e_y, -e_y, 0)$. On the unit ball,

$$\|D^2F - 20I\| \le 2 + 4\sqrt{z^2 + 2x^2} + \sqrt{2(x^2 + s^2 + y^2)} \le 2 + \sqrt{32} + \sqrt{2} < 10, \qquad D^2F \succ 10I.$$

Orthogonal embedding and uncoupled quadratics extend this witness. For $r = 1$, write $Qv = av$ and $\mathsf{T}_3[v, v] = bv$ after lower containment. Then

$$w_1 = P(a^3v)/\mu^2 = 0, \quad w_2 = P(abv)/\mu = 0, \quad w_3 = P(abv)/\mu = 0.$$

For a multistep positive profile, some $s_t \in (0, 1)$, and

$$\sum_t \alpha_t \frac{s_t^2(2 - s_t^2)}{8} > 0, \qquad \sum_t \alpha_t \frac{s_t^3(2 - s_t^3)}{72} > 0.$$

These are the coefficients of $\|Y\|^2$ and $\|w_4\|^2$. Successive zero responses force $Y = 0$ and $w_4 = 0$.

For $\alpha_* = (1/6, 1/3, 1/2)$, its normal coefficient on the blind tuple is $D_t w_1$, where $D_t = \sum_{i<j<t} \alpha_i\alpha_j(\alpha_j - \alpha_i)$. Here $D_2 = 1/108$, $D_3 = 7/108$. Only the last step has a nonzero forward normal coefficient, so

$$\begin{aligned}\Delta_{\alpha_*}(B) &= \frac{B^7}{\mu}\alpha_2(D_2D_3 - D_2^2/2)\|w_1\|^2 + o(B^7)\\ &= \frac{B^7}{2\mu}\left(\frac{7}{108^2} - \frac{1}{2\cdot 108^2}\right)\|w_1\|^2 + o(B^7) = \frac{13}{46656\mu}\|w_1\|^2B^7 + o(B^7).\end{aligned}$$

The sufficient alternative using only $(2, 8, \alpha_*)$ follows without first imposing all five equal tests. Its fifth-order conditions already force E_3 by the pair classification. The matrix

$$H_* = -6C_{(1/2,1/2)}(1) + 21C_{(1/8,\ldots,1/8)}(1) + 42C_{\alpha_*}(1) \tag{49}$$

has positive leading principal minors

$$d_1 = \frac{76377}{2097152}, \qquad d_2 = \frac{1520910159503}{28855583159353344},$$
$$d_3 = \frac{356090562117619}{1680959553716782891008}, \qquad d_4 = \frac{19242324885235395872117}{27754822648878543787753074799411 2}.$$

They are the determinants of the matrices defined in (33). The final implication is

$$d_i > 0\ (1 \le i \le 4) \Longrightarrow H_* \succ 0, \qquad 0 = \operatorname{tr}(H_*G_4) \ge \lambda_{\min}(H_*)\sum_i \|w_i\|^2 \Longrightarrow w = 0.$$

Then (34) gives the uniform $o(B^7)$ bound. This design uses $2 + 8 + 3 = 13$ updates per route; the two-profile design uses $5 + 6 = 11$.

C.3 Realization by a scalar potential

This construction is shared by the negative statements above and the checkpoint-rank necessity proof. Take retained coordinates (x, s, z, t), omitted coordinates $y \in \mathbb{R}^q$, and arbitrary $w_i \in \mathbb{R}^q$:

$$\begin{aligned}F(x, s, z, t, y) = &-x + \frac{\kappa}{2}(x^2 + s^2 + z^2 + t^2 + \|y\|^2) + xs + sz\\ &+ \langle zw_1 + tw_2, y\rangle + \frac{x^2t}{2} + xs\langle w_3, y\rangle + \frac{x^3}{6}\langle w_4, y\rangle.\end{aligned} \tag{50}$$

At zero $\mu = 1$, $v = e_x$. The derivatives are

$$\begin{aligned}Qe_x &= \kappa e_x + e_s, & Qe_s &= e_x + \kappa e_s + e_z,\\ Qe_z &= e_s + \kappa e_z + w_1, & Qe_t &= \kappa e_t + w_2,\\ Q^2e_x &= (\kappa^2 + 1)e_x + 2\kappa e_s + e_z, & PQ^2e_x &= 0,\\ PQ^3e_x &= P[(\kappa^2 + 1)Qe_x + 2\kappa Qe_s + Qe_z] = w_1,\\ \mathsf{T}_3[e_x, e_x] &= e_t, & PQ\mathsf{T}_3[e_x, e_x] &= w_2,\\ P\mathsf{T}_3[e_x, Qe_x] &= P(\kappa e_t + w_3) = w_3, & P\mathsf{T}_4[e_x, e_x, e_x] &= w_4.\end{aligned}$$

The lower vectors lie in the retained coordinates. For $F_{\text{cross}} = F + x - \kappa\|\cdot\|^2/2$, choose $\kappa > \sup\|D^2F_{\text{cross}}\|$ on a fixed ball. Then $D^2F \succeq (\kappa - \sup\|D^2F_{\text{cross}}\|)I \succ 0$. Thus every prescribed tuple is realized by one scalar potential.

D CHECKPOINT RANK AND FINITE OBSERVATION TRANSFER

The scalar correction gives a bound on the normal gradient. The prefix coefficients then yield the rank criterion in Theorem 2 and the finite transfer bound.

D.1 CORRECTION ENERGY AND THE NORMAL GRADIENT

From a saved state y, form $r = P\nabla F(y)$ and $y^- = y - \sigma r$. Assume $\|D^2F\| \le L$ on the trial segment. Taylor integration gives

$$\begin{aligned} D(y) &= F(y) - F(y - \sigma r) \\ &= \sigma\langle \nabla F(y), r\rangle - \sigma^2 \int_0^1 (1-s) D^2F(y - s\sigma r)[r, r]\, ds, \\ \langle \nabla F(y), r\rangle &= \langle W\nabla F(y) + P\nabla F(y), P\nabla F(y)\rangle = \|r\|^2, \\ |D(y) - \sigma\|r\|^2| &\le \sigma^2 \int_0^1 (1-s) L\|r\|^2\, ds = L\sigma^2\|r\|^2/2. \end{aligned}$$

For $0 < \sigma < 2/L$ this is

$$\sigma(1 - L\sigma/2)\|r\|^2 \le D(y) \le \sigma(1 + L\sigma/2)\|r\|^2. \tag{51}$$

Both constants are positive. Dividing by B^{2j} proves

$$D(y(B)) = o(B^{2j}) \quad \Longleftrightarrow \quad \|P\nabla F(y(B))\| = o(B^j). \tag{52}$$

The executor pays for the correction direction and loss evaluations. Only the scalar decrease need be returned.

D.2 THE EXACT RANK CRITERION

Proof of Theorem 2. Write $r_t(B) = P\nabla F(\bar u_t(B))$. Initial containment follows, when it is not supplied, from

$$D(\bar u_1(B)) \to 0 \implies \|r_1(B)\| \to 0 \implies P\nabla F(0) = 0 \implies Wv = v.$$

The middle implication uses $\|\nabla F(\bar u_1(B)) - \nabla F(0)\| \to 0$.

For two first fractions $a, b > 0$, the successive coefficients give

$$\begin{aligned} r_1(B) &= BaPQv + O(B^2), \\ r_1 = o(B^2) &\implies 0 = \lim_{B\downarrow 0} r_1(B)/B = aPQv \implies PQv = 0, \\ \begin{pmatrix} r_1(B) \\ r_2(B) \end{pmatrix} &= B^2 \underbrace{\begin{pmatrix} 0 & a^2/2 \\ -ab & (a+b)^2/2 \end{pmatrix}}_{R_2} \begin{pmatrix} X \\ Y \end{pmatrix} + o(B^2). \end{aligned}$$

The two-vector map is invertible because

$$\det R_2 = 0 \cdot (a+b)^2/2 - (a^2/2)(-ab) = a^3 b/2 > 0. \tag{53}$$

Consequently

$$\begin{aligned} r_1, r_2 = o(B^2) &\Longrightarrow \begin{pmatrix} X \\ Y \end{pmatrix} = R_2^{-1} \lim_{B\downarrow 0} B^{-2} \begin{pmatrix} r_1(B) \\ r_2(B) \end{pmatrix} = 0, \\ E_3 \subseteq \operatorname{ran} W &\Longrightarrow r_t(B) = o(B^2) \Longrightarrow D(\bar u_t(B)) \le \sigma(1 + L\sigma/2) o(B^4) = o(B^4). \end{aligned}$$

This proves both directions at order three.

For four fractions $a, b, c, d > 0$, the $o(B^6)$ energy premise first gives the preceding E_3 containment. The next residual coefficients are

$$\begin{pmatrix} r_1(B) \\ r_2(B) \\ r_3(B) \\ r_4(B) \end{pmatrix} = B^3 R_3 w + o(B^3), \qquad (R_3)_{t,:} = \left(e_{t,3}, -\tfrac{1}{2}\sum_{i<t} \alpha_i s_i^2, -s_t e_{t,2}, s_t^3/6 \right). \tag{54}$$

Here and below a scalar matrix acts on a tuple of normal vectors. For the determinant calculation put $s = a + b + c$, $e = ab + ac + bc$ and $K = ba^2 + c(a + b)^2$. Expansion along the first row gives

$$\begin{aligned}\det R_3 &= -\frac{a^3}{6}\det\begin{pmatrix} 0 & -a^2b/2 & -ab(a+b) \\ abc & -K/2 & -se \\ abc+de & -(K+ds^2)/2 & -(s+d)(e+ds)\end{pmatrix} \\ &= -\frac{a^3}{6}\Big[\frac{a^2b}{2}\{-abc(s+d)(e+ds)+se(abc+de)\} \\ &\qquad -\frac{ab(a+b)}{2}\{-abc(K+ds^2)+K(abc+de)\}\Big] \\ &= \frac{a^4bcd}{12}\,p(a,b,c,d),\end{aligned}$$

where

$$\begin{aligned}p(a,b,c,d) = a^3bd - a^3c^2 + a^2b^2d - 2a^2bc^2 + a^2bcd \\ + ab^3c - 2ab^2c^2 + b^4c.\end{aligned} \tag{55}$$

If $p \neq 0$, then

$$\max_t D(\bar{u}_t) = o(B^6) \Longrightarrow R_3 w = 0 \Longrightarrow w = R_3^{-1}0 = 0 \Longrightarrow E_4 \subseteq \operatorname{ran} W.$$

Conversely, E_4 containment gives $r_t = o(B^3)$ and hence $D(\bar{u}_t) \leq \sigma(1 + L\sigma/2)o(B^6) = o(B^6)$.

For $(a, b, c, d) = (14, 4, 11, 1)/30$,

$$R_3 = \begin{pmatrix} 0 & 0 & 0 & 343/20250 \\ 0 & -49/3375 & -14/375 & 9/250 \\ 77/3375 & -1087/13500 & -3683/13500 & 24389/162000 \\ 29/900 & -5189/54000 & -283/900 & 1/6\end{pmatrix}. \tag{56}$$

Its determinant is $-594854953/4151882812500000$. For $H = R_3^\top R_3$ and its positive eigenvalues λ_i,

$$\begin{aligned}&\sum_i \lambda_i^{-1} \geq \lambda_{\min}(H)^{-1}, \\ &\|R_3 w\|^2 = \sum_\ell (w^{(\ell)})^\top H w^{(\ell)} \geq \lambda_{\min}(H)\sum_\ell \|w^{(\ell)}\|^2 \\ &\qquad \geq \underbrace{\frac{1}{\operatorname{tr}(H^{-1})}}\sum_i \|w_i\|^2.\end{aligned}$$

$$\kappa_3 = 5661638641738115344/863325553770595857515625$$

The index ℓ runs over orthonormal normal coordinates. For R_2 with $a = b = 1/2$, the corresponding margin is $1/336$.

If $p = 0$, choose $w \neq 0$ in $\ker R_3$. The potential (50) realizes that tuple with four retained coordinates and one omitted coordinate. It satisfies

$$E_3 \subseteq \operatorname{ran} W, \qquad r_t(B) = B^3(R_3 w)_t + o(B^3) = o(B^3), \qquad \sum_i \|w_i\|^2 > 0.$$

Thus all four energies vanish to the required order but E_4 is not retained. With $m < 4$ prescribed rows,

$$R \in \mathbb{R}^{m\times 4} \Longrightarrow \dim\ker R = 4 - \operatorname{rank} R \geq 4 - m > 0.$$

The same potential gives a counterexample. This proves necessity within the stated observation class. Finally, p is a nonzero polynomial. Its zero set has empty interior in the positive simplex. The displayed design has $p \neq 0$, whereas equal fractions and $(1, 2, 4, 8)/15$ have $p = 0$. □

D.3 A FINITE BOUND AT UNOBSERVED TARGET STEPS

Fix one loss, one projector and $E_j \subseteq \operatorname{ran} W$, where $j = 2$ or 3. Let $\mathcal{R}$ be the invertible two-row or four-row matrix above. Calibrate at budget $b > 0$ and denote its normal gradients by r_i^{cal}. At target budget $B > 0$, let f_t^{tar} be the corresponding forward coefficient row. The gains may differ between calibration and target. Both must satisfy the same local derivative bounds used below.

We first bound the Taylor remainder. Assume bounds $M_k \geq \sup \|D^k F\|$ through order $j+2$ and field bounds D_k through order $j+1$. Use the forward construction of Lemma 1 with $p = j$: it defines A, K, C_Z, B_0. Set

$$C_n = M_2 C_Z + \mathcal{T}_{j+1}\left(\sum_{\ell=0}^{j} \frac{M_{\ell+1}}{\ell!} A(z)^\ell\right) + \frac{M_{j+2} K^{j+1}}{(j+1)!}. \tag{57}$$

The vector Taylor calculation is

$$\begin{aligned}
\|P\nabla F(\bar{u}_t) - B^j f_t \cdot w\| &\leq \|\nabla F(\bar{u}_t) - \nabla F(Z_t)\| \\
&\quad + \|\nabla F(Z_t) - T_j(\nabla F)(Z_t)\| + \|T_j(\nabla F)(Z_t) - [T_j(\nabla F)(Z_t)]_{\leq j}\| \\
&\leq M_2 C_Z B^{j+1} + \frac{M_{j+2} K^{j+1}}{(j+1)!} B^{j+1} \\
&\quad + \mathcal{T}_{j+1}\left(\sum_{\ell=0}^{j} \frac{M_{\ell+1}}{\ell!} A(z)^\ell\right) B^{j+1} = C_n B^{j+1}.
\end{aligned}$$

Thus every remainder used in the finite transfer is computable from the declared regularity bounds.

Let $R_i^{\mathrm{cal}} \geq \|r_i^{\mathrm{cal}}\|$ come from scalar intervals through (62). For $0 < b, B \leq B_0$, put $\ell_t = f_t^{\mathrm{tar}} \mathcal{R}^{-1}$. Writing $\rho^{\mathrm{cal}}, \rho_t^{\mathrm{tar}}$ for the remainder vectors,

$$\begin{aligned}
r^{\mathrm{cal}} &= b^j \mathcal{R} w + \rho^{\mathrm{cal}}, & w &= b^{-j} \mathcal{R}^{-1}(r^{\mathrm{cal}} - \rho^{\mathrm{cal}}), \\
r_t^{\mathrm{tar}} &= B^j f_t^{\mathrm{tar}} w + \rho_t^{\mathrm{tar}} = (B/b)^j \ell_t (r^{\mathrm{cal}} - \rho^{\mathrm{cal}}) + \rho_t^{\mathrm{tar}}, \\
\|r_t^{\mathrm{tar}}\| &\leq (B/b)^j \sum_i |\ell_{t,i}| R_i^{\mathrm{cal}} + C_n B^{j+1} + C_n B^j b \sum_i |\ell_{t,i}| =: \widehat{R}_t.
\end{aligned} \tag{58}$$

For comparison, the coarser word-norm bound is

$$\sum_i \|w_i\|^2 \leq \frac{\sum_i (R_i^{\mathrm{cal}} + C_n b^{j+1})^2}{\kappa b^{2j}}. \tag{59}$$

Indeed $\kappa b^{2j} \|w\|^2 \leq b^{2j} \|\mathcal{R} w\|^2 = \|r^{\mathrm{cal}} - \rho^{\mathrm{cal}}\|^2 \leq \sum_i (R_i^{\mathrm{cal}} + C_n b^{j+1})^2$.

For an ordinary-GD target with $\eta_t = B\alpha_t/\mu$, Proposition 1 now uses $\widehat{R}_t$ at every unobserved step:

$$E_N = E_0 \prod_{k<N} q_k + \sum_{t<N} (\eta_t \widehat{R}_t + d_t) \prod_{t<k<N} q_k, \qquad |V(x_N) - V(y_N)| \leq G_V E_N + L_V E_N^2/2. \tag{60}$$

Four calibration energies can therefore serve any finite target horizon under the common-loss assumptions. Their conditioning and Taylor costs remain in ℓ_t and C_n. If calibration costs four reference steps, added audit work decreases only when

$$4c_{\mathrm{ref}} + 4c_{\mathrm{probe}} < N c_{\mathrm{probe}}.$$

This comparison charges calibration and assumes reusable stored states. It concerns added scalar-audit work; an available direct norm still avoids the trial losses.

The common-loss condition matters. At an unobserved stage s, let $F_s(x) = a n^\top x$ with $Pn = n$, $\|n\| = 1$, and let all other batch losses be zero. Then $y_t = 0$, all observed normal decreases are zero, but $x_N = -\eta_s a n$ and $V(x_N) = -\eta_s a$ for $V(x) = n^\top x$. The Hessian bounds remain zero as a varies. Thus arbitrary batch changes require separate residual or drift bounds at the missing stages.

E TARGET-BIAS BOUNDS FROM NORMAL RESIDUALS

Fix batch losses F_t, learning rates $\eta_t > 0$, and stored states y_t. The comparator and reference satisfy

$$x_{t+1} = \Phi_t(x_t) = x_t - \eta_t \nabla F_t(x_t), \qquad y_{t+1} = y_t - \eta_t \Pi_t \nabla F_t(y_t) + \xi_t.$$

Each Π_t is an orthogonal projector. It may depend on the reference history. Assume

$$\|\xi_t\| \le d_t, \qquad \|x_0 - y_0\| \le E_0, \qquad P_t = I - \Pi_t, \qquad r_t = P_t \nabla F_t(y_t). \tag{61}$$

Inexact optimization tracks gradient and proximal errors for a fixed convex objective (Schmidt et al., 2011). The bound below instead tracks update defects between two specified finite executions.

E.1 FROM SCALAR RESIDUALS TO TARGET BIAS

At a probed stage, take $y_t^- = y_t - \sigma_t r_t$. Assume $\|D^2 F_t\| \le L_t$ on the trial segment and comparison domain. For $0 < \sigma_t < 2/L_t$, let $[D_t^-, D_t^+]$ enclose $F_t(y_t) - F_t(y_t^-)$. Define

$$a_t = \sigma_t(1 - L_t\sigma_t/2), \qquad R_t = \sqrt{\max(0, D_t^+)/a_t}. \tag{62}$$

The interval must be consistent with $D_t \ge 0$. In particular, $D_t^+ < 0$ is inconsistent with the exact correction model. At an unprobed stage use any independently valid bound on $\|r_t\|$. Equation (58) provides one such bound under its common-loss assumptions.

Let q_t bound the Lipschitz constant of Φ_t. The algorithm computes

$$E_{t+1} = q_t E_t + \eta_t R_t + d_t, \tag{63}$$

$$U_V = G_V E_N + L_V E_N^2/2, \tag{64}$$

where $G_V \ge \|\nabla V(y_N)\|$ and $L_V \ge \sup \|D^2 V\|$ on the terminal comparison domain. Its output is

$$|V(x_N) - V(y_N)| \le U_V. \tag{65}$$

Proof of Proposition 1. The scalar report bounds the omitted gradient:

$$a_t\|r_t\|^2 \overset{(51)}{\le} D_t \le D_t^+ \Longrightarrow \|r_t\|^2 \le D_t^+/a_t \le R_t^2.$$

For $e_t = x_t - y_t$, subtraction gives

$$\begin{aligned}
e_{t+1} &= x_t - \eta_t \nabla F_t(x_t) - y_t + \eta_t \Pi_t \nabla F_t(y_t) - \xi_t \\
&= \Phi_t(x_t) - \Phi_t(y_t) - \eta_t P_t \nabla F_t(y_t) - \xi_t, \\
\|e_{t+1}\| &\le \|\Phi_t(x_t) - \Phi_t(y_t)\| + \eta_t\|r_t\| + \|\xi_t\| \\
&\le q_t\|e_t\| + \eta_t R_t + d_t, \\
\|e_t\| &\le \|e_0\| \prod_{k<t} q_k + \sum_{i<t} (\eta_i R_i + d_i) \prod_{i<k<t} q_k \\
&\le E_0 \prod_{k<t} q_k + \sum_{i<t} (\eta_i R_i + d_i) \prod_{i<k<t} q_k = E_t.
\end{aligned}$$

The domain check is part of this induction. At stage t, the ball $\{y_t + z : \|z\| \le E_t\}$ must be contained in the domain on which q_t was certified.

Taylor integration at the terminal state gives

$$\begin{aligned}
V(x_N) - V(y_N) &= \nabla V(y_N)^\top e_N + \int_0^1 (1-s) D^2 V(y_N + s e_N)[e_N, e_N]\, ds, \\
|V(x_N) - V(y_N)| &\le \|\nabla V(y_N)\|\|e_N\| + \int_0^1 (1-s) L_V \|e_N\|^2\, ds \\
&\le G_V E_N + L_V E_N^2/2 = U_V.
\end{aligned}$$

If $V(y_N)$ is replaced by a reported center $\widehat{V}_N$ with $|\widehat{V}_N - V(y_N)| \le \epsilon_N$, then

$$|\widehat{V}_N - V(x_N)| \le |\widehat{V}_N - V(y_N)| + |V(y_N) - V(x_N)| \le \epsilon_N + U_V.$$

This is the reported finite interval. □

The map constants can be computed without differentiating Π_t. For $\|D^2F_t\| \le L_t$,

$$\|D\Phi_t(x)\| = \|I - \eta_t D^2F_t(x)\| \le 1 + \eta_t L_t.$$

If $m_t I \preceq D^2F_t \preceq L_t I$ on a convex domain, set $\overline{H} = \int_0^1 D^2F_t(y + s(x-y))\,ds$. Then

$$\begin{aligned}\Phi_t(x) - \Phi_t(y) &= (I - \eta_t\overline{H})(x-y),\\ m_t I \preceq \overline{H} &\preceq L_t I,\\ \|I - \eta_t\overline{H}\| &= \max_{\lambda\in\mathrm{spec}(\overline{H})} |1-\eta_t\lambda| \le \max\{|1-\eta_t m_t|, |1-\eta_t L_t|\}.\end{aligned}$$

The terminal gradient bound has a scalar version. If D_V^+ encloses $V(y_N) - V(y_N - \sigma_V\nabla V(y_N))$ and $0 < \sigma_V < 2/L_V$, then

$$\sigma_V(1 - L_V\sigma_V/2)\|\nabla V(y_N)\|^2 \le D_V^+, \qquad G_V = \sqrt{\frac{D_V^+}{\sigma_V(1-L_V\sigma_V/2)}}.$$

The compared executions share the realized batches and learning rates. No expectation is used in the proof. For noisy reports, suppose each error ϵ_i, conditional on the observed past, satisfies $\Pr(|\epsilon_i| > a \mid \mathcal{H}_{i-1}) \le 2\exp[-na^2/(2\sigma^2)]$. With $a = \sigma\sqrt{2\log(2K/\delta)/n}$,

$$\Pr\{\exists i \le K : |\epsilon_i| > a\} \le \sum_{i=1}^{K} \mathbb{E}\Pr(|\epsilon_i| > a \mid \mathcal{H}_{i-1}) \le K \cdot 2e^{-\log(2K/\delta)} = \delta.$$

The deterministic proof applies on this joint coverage event.

E.2 Inexact correction directions

Suppose the stored trial is $\widehat{y}^- = y - \sigma\widehat{r}$, where $r = P\nabla F(y)$, $\|\widehat{r} - r\| \le \zeta$, and $\|\nabla F(y)\| \le G$. For $\rho = \|r\|$,

$$\begin{aligned}F(y) - F(\widehat{y}^-) &= \sigma\langle\nabla F(y), \widehat{r}\rangle - \sigma^2\int_0^1 (1-s)D^2F(y - s\sigma\widehat{r})[\widehat{r},\widehat{r}]\,ds\\ &\ge \sigma\{\|r\|^2 + \langle\nabla F(y), \widehat{r} - r\rangle\} - L\sigma^2\|\widehat{r}\|^2/2\\ &\ge \sigma\rho^2 - \sigma G\zeta - L\sigma^2(\rho+\zeta)^2/2 = a\rho^2 - b\rho - c,\end{aligned}$$

where

$$a = \sigma(1 - L\sigma/2) > 0, \quad b = L\sigma^2\zeta, \quad c = \sigma G\zeta + L\sigma^2\zeta^2/2.$$

For a valid upper endpoint D^+,

$$\begin{aligned}a\rho^2 - b\rho &\le D^+ + c \le \max(0, D^+) + c,\\ (2a\rho - b)^2 &\le b^2 + 4a\{\max(0, D^+) + c\},\\ \rho &\le \frac{b + \sqrt{b^2 + 4a\{\max(0, D^+) + c\}}}{2a} =: R.\end{aligned}$$

Thus

$$R = \frac{b + \sqrt{b^2 + 4a\{\max(0, D^+) + c\}}}{2a}. \tag{66}$$

If the trial point has additional error ϵ_y, replace ζ by $\zeta + \epsilon_y/\sigma$. Reference update error enters d_t. If $\widehat{r} - r$ is itself normal, then $\langle\nabla F(y), \widehat{r} - r\rangle = \langle r, \widehat{r} - r\rangle \ge -\rho\zeta$, so the term $G\zeta$ can be replaced by $\rho\zeta$. All interval endpoints are rounded outward.

E.3 The actual bilinear update

Use the bilinear parameterization $Y = Y_0 + AB$ of LoRA (Hu et al., 2022). Let $G = \nabla F(Y)$, $G_A = GB^\top$, and $G_B = A^\top G$. Simultaneous factor GD gives

$$\begin{aligned}Y^+ - Y &= (A - \eta G_A)(B - \eta G_B) - AB\\ &= AB - \eta AG_B - \eta G_A B + \eta^2 G_A G_B - AB\\ &= -\eta(AG_B + G_A B) + \eta^2 G_A G_B.\end{aligned} \tag{67}$$

Define $P_A = AA^\dagger$, $P_B = B^\dagger B$. On matrices let $\mathcal{L}Z = P_A Z$ and $\mathcal{R}Z = ZP_B$. They satisfy

$$\mathcal{L}^2 = \mathcal{L} = \mathcal{L}^*, \quad \mathcal{R}^2 = \mathcal{R} = \mathcal{R}^*, \quad \mathcal{L}\mathcal{R} = \mathcal{R}\mathcal{L}.$$

The image of $(\dot{A}, \dot{B}) \mapsto \dot{A}B + A\dot{B}$ is $\operatorname{ran}\mathcal{L} + \operatorname{ran}\mathcal{R}$:

$$\mathcal{L}Z = A(A^\dagger Z), \qquad \mathcal{R}Z = (ZB^\dagger)B, \qquad \dot{A}B = \mathcal{R}(\dot{A}B), \qquad A\dot{B} = \mathcal{L}(A\dot{B}).$$

Its orthogonal projector is

$$\Pi = \mathcal{L} + \mathcal{R} - \mathcal{L}\mathcal{R}, \qquad \Pi G = P_A G + G P_B - P_A G P_B. \tag{68}$$

Indeed $\Pi = I - (I - \mathcal{L})(I - \mathcal{R})$, so $\Pi^* = \Pi$ and $\Pi^2 = \Pi$. This proof does not assume full factor rank. The required action is recoverable from factor gradients:

$$P_A G = A(A^\top A)^\dagger G_B, \qquad G P_B = G_A (BB^\top)^\dagger B.$$

Subtracting the projected reference step from (67) gives

$$\xi = (Y^+ - Y) + \eta \Pi G = \eta(\Pi G - A G_B - G_A B) + \eta^2 G_A G_B. \tag{69}$$

Thus $d_t \geq \|\xi_t\|_F$ includes both the factor metric and cross term. Here we choose full weight-space GD as the comparator. A large d_t produces a wide certificate.

A two-factor step gives a nonzero update defect. Take $F(Y) = \|Y - T\|_F^2/2$, $T = \left(\begin{smallmatrix} 1 & 1/10 \\ 1/10 & 1/10 \end{smallmatrix}\right)$, $A = B^\top = e_1$, and $\eta = 1/10$. Both factors are updated. Their exact matrices are

$$Y^+ = \begin{pmatrix} 1 & 1/100 \\ 1/100 & 1/10000 \end{pmatrix}, \qquad X^+ = \begin{pmatrix} 1 & 1/100 \\ 1/100 & 1/100 \end{pmatrix}.$$

The normal probe with $\sigma = 1$ gives $D = 1/200$ and $R = 1/10$. Here the factor metric term vanishes, but the cross term is nonzero:

$$\begin{aligned} d &= \|\eta^2 G_A G_B\|_F = 1/10000, \\ E_1 &= \eta R + d = 101/10000, \\ \|\nabla F(Y^+)\|_F^2 &= 2(9/100)^2 + (999/10000)^2 < (17/100)^2, \\ |F(Y^+) - F(X^+)| &\leq \frac{17}{100}\frac{101}{10000} + \frac{1}{2}\left(\frac{101}{10000}\right)^2 = 0.001768005 < 0.002, \\ F(Y_0) - F(X^+) &= 0.00285 > 0.002. \end{aligned}$$

Thus this actual factor step is certified while zero update violates the tolerance.

E.4 Execution costs

A scalar report requires a gradient-directed probe at the executor. A forward-only loss service cannot produce that direction by itself. With white-box access, a direct normal norm avoids the trial losses. The costs include direction production, trial loss, saved-state storage, basis refresh, regularity bounds and interval arithmetic.

The passive designs use eleven and thirteen updates per route. The four-prefix design uses four reference steps and four corrections. Its terminal prefix needs a gradient unless already cached. The finite transfer (58) reuses these calibration reports under its stated fixed-loss assumptions. For changing batches, every missing residual needs a separate valid bound.

F A centered finite certificate

The norm tube in Proposition 1 discards the direction of each local error. Reference derivative actions allow a computed first-order correction. We bound its nonlinear remainder separately. This is a finite residual correction, within classical goal-oriented estimation (Becker & Rannacher, 2001; Janon et al., 2016).

F.1 Linear correction and the remaining error

Proposition 2 (Reference linearization and a centered interval)**.** *Let $x_{t+1} = \Phi_t(x_t)$ and let y_t be stored reference states. Put $A_t = D\Phi_t(y_t)$ and $r_t = \Phi_t(y_t) - y_{t+1}$. Suppose a computed sequence z_t satisfies*

$$\|z_{t+1} - A_t z_t - r_t\| \le \delta_t, \qquad \|x_0 - y_0 - z_0\| \le w_0.$$

Let $c_t = y_t + z_t$. On the comparison domains assume $\|D\Phi_t\| \le q_t$ and $\|D^2\Phi_t\| \le m_t$. The domains must contain $[y_t, c_t]$ and the radius-w_t ball about c_t. Define

$$w_{t+1} = q_t w_t + \tfrac{1}{2} m_t \|z_t\|^2 + \delta_t. \tag{70}$$

If $\|\nabla V(y_N)\| \le G_V$ and $\|D^2 V\| \le L_V$ on the terminal comparison domain, then

$$\begin{aligned} \|x_t - c_t\| &\le w_t, \\ V(y_N) - V(x_N) &\in [V(y_N) - V(c_N) - \epsilon_V,\ V(y_N) - V(c_N) + \epsilon_V], \\ \epsilon_V &= (G_V + L_V\|z_N\|)w_N + \tfrac{1}{2} L_V w_N^2. \end{aligned} \tag{71}$$

Interval evaluations of $V(y_N)$ and $V(c_N)$ replace the exact values by interval subtraction.

Proof. Write $\epsilon_t = z_{t+1} - A_t z_t - r_t$. The defect of the corrected reference is

$$\begin{aligned} \Phi_t(c_t) - c_{t+1} &= \Phi_t(y_t + z_t) - y_{t+1} - z_{t+1} \\ &= \Phi_t(y_t + z_t) - \Phi_t(y_t) - A_t z_t - \epsilon_t \\ &\overset{\text{Taylor}}{=} \int_0^1 (1-s) D^2\Phi_t(y_t + s z_t)[z_t, z_t]\, ds - \epsilon_t, \\ \|\Phi_t(c_t) - c_{t+1}\| &\le \int_0^1 (1-s) m_t \|z_t\|^2\, ds + \delta_t = m_t\|z_t\|^2/2 + \delta_t. \end{aligned}$$

For $d_t = x_t - c_t$, the error recursion and its expansion are

$$\begin{aligned} d_{t+1} &= \Phi_t(c_t + d_t) - \Phi_t(c_t) + \Phi_t(c_t) - c_{t+1}, \\ \|d_{t+1}\| &\le q_t\|d_t\| + m_t\|z_t\|^2/2 + \delta_t, \\ \|d_t\| &\le \|d_0\| \prod_{k<t} q_k + \sum_{i<t} (m_i\|z_i\|^2/2 + \delta_i) \prod_{i<k<t} q_k \\ &\le w_0 \prod_{k<t} q_k + \sum_{i<t} (m_i\|z_i\|^2/2 + \delta_i) \prod_{i<k<t} q_k = w_t. \end{aligned}$$

At the terminal state,

$$\begin{aligned} \nabla V(c_N) &= \nabla V(y_N) + \int_0^1 D^2 V(y_N + s z_N) z_N\, ds, \\ \|\nabla V(c_N)\| &\le G_V + \int_0^1 L_V\|z_N\|\, ds = G_V + L_V\|z_N\|, \\ V(x_N) - V(c_N) &= \nabla V(c_N)^\top d_N + \int_0^1 (1-s) D^2 V(c_N + s d_N)[d_N, d_N]\, ds, \\ |V(x_N) - V(c_N)| &\le (G_V + L_V\|z_N\|)\|d_N\| + L_V\|d_N\|^2/2 \le \epsilon_V, \\ V(y_N) - V(x_N) &= V(y_N) - V(c_N) - \{V(x_N) - V(c_N)\}. \end{aligned}$$

The last two lines give both endpoints in (71). □

F.2 Derivative inputs and the tolerance decision

For ordinary SGD, all correction inputs are at the reference states:

$$\Phi_t(x) = x - \eta_t \nabla F_t(x), \qquad A_t z = z - \eta_t D^2 F_t(y_t) z, \qquad r_t = y_t - \eta_t \nabla F_t(y_t) - y_{t+1}.$$

Thus a correction step costs one Hessian–vector action. It reuses the reference gradient. A bound $\|D^3 F_t\| \le M_{3,t}$ gives $m_t = \eta_t M_{3,t}$. This derivative input is additional to the second-order scalar

interface. One terminal loss at c_N gives the interval center. A separate terminal gradient is unnecessary because $G_V + L_V\|z_N\|$ is a valid bound.

For $[l, u]$ enclosing the signed bias and tolerance $\tau > 0$, the decision rule is

$$\begin{array}{ll} \max\{|l|, |u|\} \le \tau & \Longrightarrow \text{accept}, \\ l > \tau \text{ or } u < -\tau & \Longrightarrow \text{reject}, \\ \text{otherwise} & \Longrightarrow \text{undetermined}. \end{array}$$

The rule distinguishes a certified violation from an uninformative upper bound. An initial mismatch enters w_0. Stored addition and derivative-action errors enter δ_t and the terminal radius; they are not assumed exact.

G TARGET CERTIFICATES FOR THE FEATURE-HEAD EXPERIMENT

This section gives the derivative bounds, arithmetic error estimates and stored inputs used for Figure 2.

G.1 FEATURE MODEL AND DERIVATIVE BOUNDS

The inputs are archived binary64 RoBERTa-large features. They use attention-mask mean pooling, training-mean centering and a common RMS scale. The first 64 SST-2 training examples and 32 validation examples give rows $a_i, c_j \in \mathbb{R}^{1024}$ and labels $b_i, d_j \in \{0, 1\}$. The archive records model, data, preprocessing and array hashes. The backbone is fixed.

For a batch I_t of size eight, define

$$F_t(w) = \frac{1}{|I_t|}\sum_{i\in I_t}[\log(1 + e^{a_i^\top w}) - b_i a_i^\top w] + \frac{1}{4}\|w\|^2,$$
$$V(w) = \frac{1}{32}\sum_{j=1}^{32}[\log(1 + e^{c_j^\top w}) - d_j c_j^\top w].$$

For $p(z) = (1 + e^{-z})^{-1}$,

$$\ell''(z) = p(z)(1 - p(z)) \in [0, 1/4],$$
$$|\ell'''(z)| = p(z)(1 - p(z))|1 - 2p(z)| \le 1/(6\sqrt{3}).$$

The last maximum follows by differentiating $p(1-p)(1-2p)$ on $[0, 1/2]$: its derivative is $1-6p+6p^2$, with interior root $(3 - \sqrt{3})/6$. Symmetry gives the other half of $[0, 1]$. The stored arrays satisfy $\max_i \|a_i\|^2 < 2$ and $\sum_j \|c_j\|^2/128 < 0.26$. Therefore

$$\begin{aligned} \tfrac{1}{2}I \preceq D^2F_t(w) &= \tfrac{1}{2}I + \frac{1}{|I_t|}\sum_{i\in I_t} p_i(1 - p_i)a_i a_i^\top \preceq I, \\ 0 \preceq D^2V(w) &\preceq \frac{1}{128}\sum_j c_j c_j^\top \preceq 0.26I, \\ \|D^3F_t(w)\| &\le \frac{1}{6\sqrt{3}\,|I_t|}\sum_{i\in I_t}\|a_i\|^3 < \sqrt{2/27}. \end{aligned} \tag{72}$$

These bounds hold for all states and all batches. For $\eta = 1/8$, $q = 15/16$ and $m = \eta\sqrt{2/27}$.

The training matrix A has shape 64×1024. Its numerical rank is 63 under the default singular-value tolerance. The exact dimension bound used here is only $\operatorname{rank} A \le 64$. For the exact full program,

$$x_0 = 0, \quad x_t \in \operatorname{ran} A^\top \Longrightarrow \nabla F_t(x_t) \in \operatorname{ran} A^\top \Longrightarrow x_{t+1} \in \operatorname{ran} A^\top.$$

Thus the ambient dimension is not the effective dynamical dimension.

G.2 STORED PROGRAMS AND SCALAR INTERVALS

Seeds 19 and 41 each specify 64 minibatches. Sampling is uniform without replacement within each batch. All candidates on one seed share that batch sequence. The candidate dimensions are $8, 16, 32, 48$. The first basis uses the top training principal directions. Every eight steps, it is refreshed from the top $k-2$ fixed directions and the two most recent reference gradients. For the stored basis U_t, the exact projector is $\Pi_t = U_t(U_t^\top U_t)^{-1}U_t^\top$. The binary64 update uses $U_tU_t^\top$; the bound includes this discrepancy.

Set $d = 1024$, $u = 2^{-53}$ and $\gamma_n = nu/(1-nu)$. The arithmetic bounds below assume binary64 operations with rounding to nearest and no overflow or underflow. Each basic operation then has relative error at most u. Since $(2d+4)u < 1$, the dot-product bound gives

$$|\,\mathrm{fl}(a^\top w) - a^\top w| \underset{\text{Higham, 2002, §3.1}}{\leq} \gamma_d \sum_{i=1}^{d} |a_i w_i| \underset{\text{Cauchy–Schwarz}}{\leq} \gamma_d \|a\|\|w\| \leq \gamma_{2d+4}\|a\|\|w\|.$$

Norm bounds and the scalar exponential and logarithm are evaluated with outward 53-bit intervals. For $Q_w = \mathrm{fl}(w^\top w)$,

$$\frac{Q_w}{1+\gamma_{2d+4}} \leq \|w\|^2 \leq \frac{Q_w}{1-\gamma_{2d+4}}.$$

The same score intervals give sigmoid intervals. Gradient error includes their propagated error and vector-reduction error.

For $\kappa_U \geq \|U^\top U - I\| < 1$, the projector discrepancy is

$$\|U(U^\top U)^{-1}U^\top - UU^\top\| = \|U[(U^\top U)^{-1} - I]U^\top\| \leq \|U\|^2\|(U^\top U)^{-1}\|\|I - U^\top U\| \leq \frac{(1+\kappa_U)\kappa_U}{1-\kappa_U}.$$

An entrywise Gram-product enclosure bounds κ_U. For $v = \mathrm{fl}(U^\top g)$, two action-error bounds are

$$\epsilon_1 = \gamma_{2d+4}\sqrt{k(1+\kappa_U)}\,\|g\|, \qquad \epsilon_2 = \sqrt{1+\kappa_U}\,\epsilon_1 + \gamma_{2k+4}\sqrt{k(1+\kappa_U)}\,\|v\|.$$

Stored-state replay differences are also enclosed. This allows different BLAS reduction orders. The verified normal-direction errors are below $3\,10^{-11}$. The update errors are below $4\,10^{-12}$. Both use the declared allowance 10^{-8}. All 512 normal probes have positive lower endpoints. Their largest interval width is below $1.82\,10^{-13}$.

The scalar and direct-norm certificates use the same stored trajectories. Figure 2B plots all of their upper endpoints. The archive retains the exact binary interval endpoints. At tolerance 10^{-3}, all of these norm-only intervals are undetermined. Direct norms avoid the extra trial losses when available.

G.3 VERIFIED REFERENCE LINEARIZATION

We next use Proposition 2 on the same reference states. Let $X = A_{I_t}$, $p = (1+e^{-Xy_t})^{-1}$, and $\omega = p \odot (1-p)$. The exact correction is

$$z_{t+1} = \frac{15}{16}z_t - \frac{1}{8|I_t|}X^\top \operatorname{diag}(\omega)Xz_t + (y_t - \tfrac{1}{8}\nabla F_t(y_t) - y_{t+1}), \qquad z_0 = 0. \tag{73}$$

Only gradients and Hessian actions at y_t are used.

The following calculation bounds the arithmetic error in that action. Let $\widehat{\omega}$ be the computed weight vector and let $\epsilon_\omega \geq \|\widehat{\omega} - \omega\|_\infty$ come from the sigmoid intervals. Set $\widetilde{\gamma}_d = \gamma_{2d+4}$ and $\widetilde{\gamma}_b = \gamma_{2|I_t|+12}$. Write $n = |I_t|$, $s_i = a_i^\top z_t$ and $\widehat{s}_i = s_i + e_i$. Let e_{red} collect multiplication and reduction errors after

the scores have been computed. Then

$$\begin{aligned}
|e_i| &\le \widetilde{\gamma}_d \|a_i\| \|z_t\|, \\
\widehat{h} - \frac{1}{n}\sum_i \omega_i a_i s_i &= \frac{1}{n}\sum_i (\widehat{\omega}_i - \omega_i) a_i s_i + \frac{1}{n}\sum_i \widehat{\omega}_i a_i e_i + e_{\mathrm{red}}, \\
\|e_{\mathrm{red}}\| &\le \frac{\widetilde{\gamma}_b}{n}\sum_i \|a_i\| \, |\widehat{\omega}_i| \, |\widehat{s}_i| \\
&\le \widetilde{\gamma}_b (1/4 + \epsilon_\omega)(1 + \widetilde{\gamma}_d) \left(\frac{1}{n}\sum_i \|a_i\|^2 \right) \|z_t\|.
\end{aligned}$$

The triangle inequality now gives

$$\begin{aligned}
\left\| \widehat{h} - \frac{1}{|I_t|} X^\top \operatorname{diag}(\omega) X z_t \right\| &\le \left[2\epsilon_\omega + 2(1/4 + \epsilon_\omega)\widetilde{\gamma}_d + 2\widetilde{\gamma}_b (1/4 + \epsilon_\omega)(1 + \widetilde{\gamma}_d) \right] \|z_t\| \\
&=: \epsilon_h.
\end{aligned}$$

The three terms respectively bound weight error, score products and the final vector reduction. They use $|I_t|^{-1} \sum_{i \in I_t} \|a_i\|^2 < 2$. Let ϵ_g enclose the stored gradient error. For the computed residual $\widehat{r}_t$, the entire update error is bounded by

$$\delta_t^{\mathrm{calc}} = \frac{1}{8}(\epsilon_g + \epsilon_h) + \gamma_{12} \left(\|y_t\| + \frac{1}{8}\|\widehat{g}_t\| + \|y_{t+1}\| + \frac{15}{16}\|z_t\| + \frac{1}{8}\|\widehat{h}_t\| + \|\widehat{r}_t\| \right).$$

Every such value is below $4\,10^{-15}$. The certificate uses the larger common allowance $\delta_t = 10^{-8}$. The verified recurrence is therefore

$$w_{t+1} = \frac{15}{16} w_t + \frac{1}{16}\sqrt{\frac{2}{27}} \|z_t\|^2 + 10^{-8}, \qquad w_0 = 0.$$

The norm and recurrence endpoints are rounded outward. The terminal addition $c_N = y_N + z_N$ has a separate error bound. One interval evaluation of $V(c_N)$ provides the center. The scalar terminal probe at y_N supplies G_V. Thus no gradient at a full-comparator state is used.

At tolerance 10^{-3}, the intervals in the main text certify dimensions $16, 32, 48$ on both streams. They certify that dimension 8 violates the tolerance. Since $V(0) = \log 2$, subtracting the same full-loss enclosure from $\log 2$ gives the zero-update intervals in Section 7. Their positive lower endpoints exceed the tolerance.

These are development calculations on archived trajectories. The tolerance was fixed before this correction, after historical target values were available. All eight stored point checks lie inside the intervals. Section 7 lists the original and added execution costs.

H FINITE BOUNDS FROM PAIRED TERMINAL LOSSES

The preceding finite bounds use normal-correction reports. Here the observations are the paired terminal losses from Section 3. We derive their finite bounds from the response matrices and Taylor remainders in Appendices A and B.

H.1 OBSERVATION INTERVALS AND THE FINITE GUARANTEE

Let $\kappa(s) = \phi(s/\mu^2)/\mu$, $\psi(s) = \phi(s/\mu^2)$ and $\bar{x}_t = \bar{u}_t$. Thus the same executed field is $-\kappa(\|g\|^2)g$. Initially assume $Wv = v$; Section H.6 removes that assumption. For a profile write $s_t = \sum_{i<t} \alpha_i$, $T_t = \sum_{i>t} \alpha_i$, $d_t = 1 + (1 - \vartheta)T_t - \vartheta s_t$, and

$$c_\alpha(\vartheta) = \sum_t \alpha_t \{ s_t[1 + (1 - \vartheta)T_t] - \vartheta s_t^2/2 \}.$$

With the physical adjoint λ_t, $p_t = W\nabla F(\bar{x}_t)$, $r_t = P\nabla F(\bar{x}_t)$, and $s_t^g = \|\nabla F(\bar{x}_t)\|^2$, the exact defect and telescope proved above have the form

$$\begin{aligned}\Delta_\alpha &= \sum_t B\alpha_t[\kappa(s_t^g)\langle P\lambda_{t+1}, r_t\rangle + \{\kappa(s_t^g) - \kappa(\|p_t\|^2)\}\langle\lambda_{t+1}, p_t\rangle] + R,\\ |R| &\le C_R\left(\sum_t B\alpha_t\|f(\bar{x}_t) - \bar{f}(\bar{x}_t)\|\right)^2.\end{aligned} \tag{74}$$

All constants used by the following algorithm are computed from the stated derivative bounds by the formulas immediately below.

Proposition 3 (Finite scalar input–output certificate)**.** *Assume $Wv = v$. Fix the loss, gain, projector and anchor of* (1)*. Supply a working radius R, scalar bounds $L \ge \|D^2F\|$, $M \ge \|D^3F\|$, and $\|D^3F(x) - D^3F(0)\| \le M_\beta\|x\|^\beta$, $0 < \beta \le 1$, on that ball. The gain and its derivative bounds are known. Choose B_0 by the derived speed bound $D_0B_0 \le R$. For probes i, observe $\widehat{\Delta}_i$ at $b_i \in (0, B_0]$ with $|\widehat{\Delta}_i - \Delta_{\alpha_i}(b_i)| \le \eta_i$. Assume $0 \le \vartheta \le 1$.*

Compute the constants a_2, a_1, a_0 in Appendix H. For any probe with $k_i = c_{\alpha_i}(\vartheta)/\mu - a_2b_i > 0$, form

$$Z_i = \max\{0, (\widehat{\Delta}_i + \eta_i)/b_i^3 + a_0b_i^2\}, \qquad \rho^+ = \min\left\{L, \min_{i:k_i>0}\frac{a_1b_i + \sqrt{(a_1b_i)^2 + 4k_iZ_i}}{2k_i}\right\}. \tag{75}$$

An empty inner minimum leaves the prior value L. Then $\|PQv\| \le \rho^+$. Compute the explicit envelope ϵ_{ρ^+} in (83) *and set*

$$\begin{aligned}\mathcal{K}_0 &= \{Z \succeq 0 : Z_{11} \le L^4/(4\mu^2),\ Z_{22} \le M^2\},\\ y_i &= \mu\widehat{\Delta}_i/b_i^5, \qquad \delta_i = \mu\{\eta_i + \epsilon_{\rho^+}(b_i)\}/b_i^5,\\ \mathcal{K} &= \mathcal{K}_0 \cap \bigcap_i\{Z : |\operatorname{tr}(A_{\alpha_i}(\vartheta)Z) - y_i| \le \delta_i\}.\end{aligned} \tag{76}$$

For a target (τ, B) with $0 < B \le B_0$, minimize and maximize $\operatorname{tr}(A_\tau(\vartheta)Z)$ over $\mathcal{K}$, obtaining l_τ, u_τ. The guaranteed output is

$$\Delta_\tau(B) \in [B^5l_\tau/\mu - \epsilon_{\rho^+}(B), B^5u_\tau/\mu + \epsilon_{\rho^+}(B)]. \tag{77}$$

It is contained in the interval obtained with the identical prior and no probe constraints. Known dimension restrictions can also be intersected; using the unrestricted PSD set is a sound outer relaxation. An empty $\mathcal{K}$ reports inconsistent inputs. Certified dual inequalities may replace exact optimization, with the resulting interval intersected with the prior interval.

H.2 Common derivative and remainder bounds

Suppose $\|D^2F\| \le L$, $\|D^3F\| \le M$ on the working ball of radius R. Let $G = \mu + LR$ and let K_i bound $|\kappa^{(i)}|$, $0 \le i \le 3$, on the squared-norm ranges of the full and projected gradients. For constant or regularized gains this interval can include zero. For $\kappa(s) = s^{-1/2}$ a positive lower endpoint is needed; when $\mu > LR$, $(\mu - LR)^2$ is sufficient and does not require another oracle. Define

$$\begin{aligned}&U_1 = K_0 + 2K_1G^2, \qquad U_2 = 6K_1G + 4K_2G^3, \qquad U_3 = 6K_1 + 24K_2G^2 + 8K_3G^4,\\ &D_0 = K_0G, \qquad D_1 = LU_1, \qquad D_2 = MU_1 + L^2U_2.\end{aligned} \tag{78}$$

For unit vectors a, b, c, write $\sum_{\mathrm{cyc}}$ for the three cyclic permutations of (a, b, c). Differentiation gives

$$\begin{aligned}DU(g)a &= \kappa a + 2\kappa'\langle g, a\rangle g,\\ D^2U(g)[a, b] &= 2\kappa'\{\langle g, a\rangle b + \langle g, b\rangle a + \langle a, b\rangle g\} + 4\kappa''\langle g, a\rangle\langle g, b\rangle g,\\ D^3U(g)[a, b, c] &= 2\kappa'\sum_{\mathrm{cyc}}\langle a, b\rangle c\\ &\quad + 4\kappa''\sum_{\mathrm{cyc}}\{\langle g, a\rangle\langle b, c\rangle g + \langle g, a\rangle\langle g, b\rangle c\} + 8\kappa'''\langle g, a\rangle\langle g, b\rangle\langle g, c\rangle g.\end{aligned}$$

Thus $\|DU\| \le U_1$, $\|D^2U\| \le U_2$, $\|D^3U\| \le U_3$. For $f = -U \circ \nabla F$,

$$\|f\| \le K_0 G = D_0, \quad \|Df\| \le U_1 L = D_1, \quad \|D^2 f\| \le U_1 M + U_2 L^2 = D_2.$$

Inserting W in the gradient and its derivatives preserves these bounds because $\|W\| = 1$ (or $W = 0$). Choose B_0 with $D_0 B_0 \le R$. Set

$$\begin{aligned}
\Lambda &= G e^{D_1 B_0}, \qquad K_d = K_0 + K_1 G^2, \\
C_R &= \tfrac{1}{2} e^{2D_1 B_0}(L + G D_2 B_0 e^{D_1 B_0}), \\
c_2 &= D_1 D_0/2, \qquad c_3 = D_1 c_2 + D_2 D_0^2/2, \qquad C_g = L c_2 + M D_0^2/2, \\
C_{\lambda,1} &= L D_0 + D_1 \Lambda, \qquad C_a = C_g + D_1 C_{\lambda,1} + \mu D_2 D_0, \\
K_v &= K_1 (G + \mu) L D_0, \qquad \chi_* = K_1 \Lambda G, \\
C_\chi &= K_2 (G + \mu) L D_0 \Lambda G + K_1 (G C_{\lambda,1} + \mu L D_0).
\end{aligned}$$

The first state remainder follows from

$$\begin{aligned}
\|\bar{x}_t\| &\le B D_0 s_t, \\
\|\bar{x}_t - B s_t v\| &\le \sum_{i<t} B\alpha_i \|\bar{f}(\bar{x}_i) - \bar{f}(0)\| \le D_1 D_0 B^2 \sum_{i<t} \alpha_i s_i \\
&= \frac{D_1 D_0 B^2}{2} \left(s_t^2 - \sum_{i<t} \alpha_i^2 \right) \le c_2 B^2, \\
\|P \nabla F(\bar{x}_t) - B s_t P Q v\| &\le L \|\bar{x}_t - B s_t v\| + \tfrac{1}{2} M \|\bar{x}_t\|^2 \le C_g B^2.
\end{aligned}$$

For the adjoint,

$$\begin{aligned}
\|\lambda_t\| &\le G \prod_{i=t}^{N-1} (1 + B\alpha_i D_1) \le \Lambda, \\
\|\lambda_t - g_0\| &\le \|\nabla F(\bar{x}_N) - g_0\| + \sum_{i=t}^{N-1} B\alpha_i \|Df(\bar{x}_i)^\top \lambda_{i+1}\| \\
&\le L D_0 B + D_1 \Lambda B = C_{\lambda,1} B, \\
\|Df(\bar{x}_i)^\top \lambda_{i+1} - Df(0)^\top g_0\| &\le D_1 \|\lambda_{i+1} - g_0\| + \mu D_2 \|\bar{x}_i\| \le (D_1 C_{\lambda,1} + \mu D_2 D_0) B.
\end{aligned}$$

Since $Df(0)^\top g_0 = (1 - \vartheta) Q v$, summing gives

$$\begin{aligned}
&\|\bar{x}_t - B s_t v\| \le c_2 B^2, && \|r_t - B s_t z\| \le C_g B^2, \\
&\|\lambda_t - g_0\| \le C_{\lambda,1} B, \qquad \|P\lambda_{t+1} - B\{1 + (1 - \vartheta) T_t\} z\| \le C_a B^2, \quad z = PQv. && (79)
\end{aligned}$$

The last remainder is at most $C_g B^2 + \sum_{i>t} B\alpha_i (D_1 C_{\lambda,1} + \mu D_2 D_0) B \le C_a B^2$.

The scalar multiplier in the second term of (74) is

$$\chi_t = \left[\int_0^1 \kappa'(\|p_t\|^2 + s\|r_t\|^2)\, ds \right] \langle \lambda_{t+1}, p_t \rangle.$$

The bounds already defined give

$$|\kappa(s_t^g) - 1/\mu| \le K_v B, \qquad |\chi_t| \le \chi_*, \qquad |\chi_t + \vartheta/(2\mu)| \le C_\chi B. \tag{80}$$

The multiplier estimates use the explicit differences

$$\begin{aligned}
|\|g_t\|^2 - \mu^2| &\le (\|g_t\| + \mu)\|g_t - g_0\| \le (G + \mu) L D_0 B, \\
|\langle \lambda, p \rangle - \mu^2| &= |\langle \lambda - g_0, p \rangle + \langle g_0, p - g_0 \rangle| \\
&\le G C_{\lambda,1} B + \mu L D_0 B, \\
|\chi_t - \kappa'(\mu^2)\mu^2| &\le K_2 (G + \mu) L D_0 B \Lambda G + K_1 (G C_{\lambda,1} + \mu L D_0) B = C_\chi B.
\end{aligned}$$

H.3 THE DEFECT INFERRED FROM A LOWER-ORDER RESPONSE

Combining (79)–(80) with (74), and writing $\rho = \|z\|$, gives

$$\left|\Delta_\alpha(B) - \frac{c_\alpha(\vartheta)}{\mu}\rho^2 B^3\right| \le a_2\rho^2 B^4 + a_1\rho B^4 + a_0 B^5, \tag{81}$$

where

$$\begin{aligned} a_2 &= 2K_v + C_\chi + C_R K_d^2, \\ a_1 &= K_0(C_a + 2C_g) + 2\chi_* C_g + 2C_R K_d^2 C_g B_0, \\ a_0 &= K_0 C_g C_a + \chi_* C_g^2 + C_R K_d^2 C_g^2 B_0. \end{aligned}$$

The normal product error before multiplication by h_t is at most

$$K_0\{(C_a + 2C_g)\rho B^3 + C_g C_a B^4\} + 2K_v\rho^2 B^3;$$

the tangential error is at most $\chi_*(2C_g\rho B^3 + C_g^2 B^4) + C_\chi\rho^2 B^3$. Finally the quadratic state remainder is bounded by $C_R K_d^2 B^2(\rho B + C_g B^2)^2$. After multiplication by the total step mass B, their sum is

$$\begin{aligned} |\Delta_\alpha - c_\alpha\rho^2 B^3/\mu| &\le (2K_v + C_\chi)\rho^2 B^4 + [K_0(C_a + 2C_g) + 2\chi_* C_g]\rho B^4 \\ &\quad + (K_0 C_g C_a + \chi_* C_g^2)B^5 + C_R K_d^2(\rho^2 B^4 + 2C_g\rho B^5 + C_g^2 B^6) \\ &\le a_2\rho^2 B^4 + a_1\rho B^4 + a_0 B^5. \end{aligned}$$

For a probe with $|\widehat{\Delta} - \Delta_\alpha(b)| \le \eta$, let

$$k_b = c_\alpha(\vartheta)/\mu - a_2 b > 0, \qquad Z_b = \max\{0, (\widehat{\Delta} + \eta)/b^3 + a_0 b^2\}.$$

The scalar bootstrap is the following quadratic calculation:

$$\begin{aligned} (c_\alpha/\mu - a_2 b)\rho^2 - a_1 b\rho \le \Delta_\alpha(b)/b^3 + a_0 b^2 &\le (\widehat{\Delta} + \eta)/b^3 + a_0 b^2 \le Z_b, \\ (2k_b\rho - a_1 b)^2 &\le (a_1 b)^2 + 4k_b Z_b, \\ \rho &\le \frac{a_1 b + \sqrt{(a_1 b)^2 + 4k_b Z_b}}{2k_b}. \end{aligned}$$

Intersecting with $\rho = \|PQv\| \le L$ gives

$$\rho \le \rho^+(b) := \min\left\{L, \frac{a_1 b + \sqrt{(a_1 b)^2 + 4k_b Z_b}}{2k_b}\right\}. \tag{82}$$

Multiple valid probes give the minimum of their upper bounds. If $k_b \le 0$, that probe supplies no bootstrap bound and the prior $\rho \le L$ is retained. The same finite observation may subsequently be reused in the Gram constraints: both inequalities are deterministic implications of its one reported interval, so independence is not assumed. A vanishing numerical report does not prove $\rho = 0$; the bound retains truncation and observation errors, with scale $O(b + \sqrt{\eta/b^3})$ near a zero-defect response.

H.4 THE HIGHER-ORDER RESPONSE REMAINDER

Assume the anchored bound $\|D^3F(x) - D^3F(0)\| \le M_\beta\|x\|^\beta$, $0 < \beta \le 1$. The following constants, used only in this appendix, are computed once:

$$\begin{aligned} A_f &= \{Lc_3 + Mc_2(D_0 + 1)/2\}B_0^{1-\beta} + \frac{M_\beta D_0^{2+\beta}}{(1+\beta)(2+\beta)}, \\ C_J &= \{3LU_2 M/2 + U_3 L^3/2\}R^{1-\beta} + U_1 M_\beta/(1+\beta), \\ J_R &= D_2 c_2 B_0^{1-\beta} + C_J D_0^{1+\beta}, \\ A_a &= A_f + \mu J_R + (D_1 C_a + 2LD_2 D_0)B_0^{1-\beta}, \qquad \sigma = 4\mu^5\kappa''(\mu^2), \\ c_f &= (1+\vartheta)L/(2\mu), \qquad c_a = \{(1+\vartheta)/2 + 2\vartheta + |\sigma - 2\vartheta|\}L/\mu. \end{aligned}$$

Let $Z_t = Bs_t v + B^2 u_t^{[2]}$. The state recursion and second-order Taylor remainder give

$$\|\bar{x}_t - Z_t\| \le \sum_{i<t} B\alpha_i (D_1 c_2 B^2 + D_2 D_0^2 B^2/2) \le c_3 B^3.$$

The gradient coefficient remainder is bounded by

$$\begin{aligned}\|g(\bar{x}_t) - g_0 - Bs_t Qv - B^2 g_t^{[2]}\| &\le Lc_3 B^3 + \tfrac{1}{2} M c_2 (D_0+1) B^3 + \frac{M_\beta (D_0 B)^{2+\beta}}{(1+\beta)(2+\beta)} \\ &\le A_f B^{2+\beta}.\end{aligned}$$

For the adjoint, write $g_x = \nabla F(x)$ and $H_x = D^2 F(x)$. The derivative remainder decomposes as

$$\begin{aligned}Df(x) - Df(0) - D^2 f(0)[x,\cdot] = &- DU(g_x)(H_x - Q - \mathsf{T}_3[x,\cdot]) \\ &- [DU(g_x) - DU(g_0) - D^2 U(g_0)[Qx,\cdot]]Q \\ &- [DU(g_x) - DU(g_0)]\mathsf{T}_3[x,\cdot].\end{aligned}$$

The three terms have norm bounds

$$\frac{U_1 M_\beta}{1+\beta}\|x\|^{1+\beta}, \qquad \frac{L}{2}(U_2 M + U_3 L^2)\|x\|^2, \qquad U_2 LM\|x\|^2.$$

Using $\|x\|^2 \le R^{1-\beta}\|x\|^{1+\beta}$ gives

$$\|Df(x) - Df(0) - D^2 f(0)[x,\cdot]\| \le C_J \|x\|^{1+\beta}.$$

At a reference state,

$$\|Df(\bar{x}_t) - Df(0) - BD^2 f(0)[s_t v,\cdot]\| \le D_2 c_2 B^2 + C_J D_0^{1+\beta} B^{1+\beta} \le J_R B^{1+\beta}.$$

Expanding $Df(\bar{x}_t)^\top \lambda_{t+1}$ through degree one leaves the bound $D_1 C_a B^2 + 2LD_2 D_0 B^2 + \mu J_R B^{1+\beta}$. The terminal and accumulated errors are therefore bounded by

$$A_f B^{2+\beta} + \sum_t B\alpha_t [D_1 C_a B^2 + 2LD_2 D_0 B^2 + \mu J_R B^{1+\beta}] \le A_a B^{2+\beta}.$$

Let $\kappa_v = \langle v, Qv\rangle$. In this subsection set $g_t^{[2]} = [B^2]\nabla F(\bar{x}_t)$ and $\lambda_t^{[2]} = [B^2]\lambda_t$ in physical units. The forward coefficient uses $\widetilde{X} = X - (PQz + \vartheta\kappa_v z)/\mu$ instead of X. The first adjoint factor has the full coefficient

$$[B]DU(g_t)\lambda_{t+1} = d_t Qv/\mu + \frac{\kappa_v}{\mu}\{-\vartheta[1 + (1-\vartheta)T_t] + (\sigma - 2\vartheta)s_t\}v.$$

Let $b_{i+1} = 1 + (1-\vartheta)T_i$ and $\chi_i^v = -\vartheta b_{i+1} + (\sigma - 2\vartheta)s_i$. The terminal coefficient and backward sum give the exact correction

$$P\lambda_t^{[2]} - (U_t X + V_t Y) = \frac{e_{N,2}}{\mu}(PQz + \vartheta\kappa_v z) - \frac{\kappa_v}{\mu}\sum_{i=t}^{N-1}\alpha_i \chi_i^v z,$$

$$\|P\lambda_t^{[2]} - (U_t X + V_t Y)\| \le \frac{(1+\vartheta)L}{2\mu}\rho + \frac{L}{\mu}\sum_i \alpha_i(2\vartheta + |\sigma - 2\vartheta|)\rho \le c_a \rho,$$

$$\|g_t^{[2]} - a_t \cdot (X,Y)\|_P = e_{t,2}\|PQz + \vartheta\kappa_v z\|/\mu \le c_f \rho.$$

Here $\|u\|_P = \|Pu\|$. Including the lower degree and Taylor remainders gives

$$R_g = \rho B + c_f \rho B^2 + A_f B^{2+\beta}, \qquad R_a = 2\rho B + c_a \rho B^2 + A_a B^{2+\beta}.$$

Use $R_X = L^2/(2\mu)$, $R_Y = M$, $f_* = (R_X + R_Y)/2$ and $b_* = f_* + 2R_X + R_Y/2$. Then $\|a_t \cdot (X,Y)\| \le f_*$ and $\|b_{t+1} \cdot (X,Y)\| \le b_*$ for every positive profile. One common finite envelope is

$$\begin{aligned}\epsilon_\rho(B) = &BK_0\{B^2 f_* R_a + B^2 b_* R_g + R_g R_a\} \\ &+ B\chi_*\{2B^2 f_* R_g + R_g^2\} + (K_v f_* b_* + C_\chi f_*^2)B^6 \\ &+ C_R K_d^2 B^2 (B^2 f_* + R_g)^2.\end{aligned} \tag{83}$$

Write $r_t = B^2 f_t + e_t^g$ and $P\lambda_{t+1} = B^2 b_t + e_t^a$, with $\|e_t^g\| \le R_g$, $\|e_t^a\| \le R_a$. Then

$$\begin{aligned}
|\langle r_t, P\lambda_{t+1}\rangle - B^4\langle f_t, b_t\rangle| &\le B^2 f_* R_a + B^2 b_* R_g + R_g R_a,\\
|\|r_t\|^2 - B^4\|f_t\|^2| &\le 2B^2 f_* R_g + R_g^2,\\
|\kappa(s_t^g) - 1/\mu| B^4 |\langle f_t, b_t\rangle| &\le K_v f_* b_* B^5,\\
|\chi_t + \vartheta/(2\mu)| B^4 \|f_t\|^2 &\le C_\chi f_*^2 B^5,\\
|R| &\le C_R K_d^2 B^2 (B^2 f_* + R_g)^2.
\end{aligned}$$

Multiplying the first four rows by B and adding the fifth yields

$$\left|\Delta_\alpha(B) - \frac{B^5}{\mu}\operatorname{tr}(A_\alpha(\vartheta) G_3)\right| \le \epsilon_\rho(B). \tag{84}$$

The four lines in the envelope respectively bound the normal product, the normalizer square, the scalar multiplier changes, and the state remainder. It is increasing in ρ; therefore the observed upper bound (82) may replace ρ without circularity. Its order is $B^{5+\beta} + \rho B^4 + \rho^2 B^3$ for fixed local inputs.

H.5 Proof of the finite input–output guarantee

The prior cap follows from $0 \preceq Q^2 \preceq L^2 I$ and $Pv = 0$:

$$\|PQ^2 v\| = \|P(Q^2 - L^2 I/2)v\| \le L^2/2.$$

The true Gram matrix satisfies the prior bounds and, for every probe,

$$\begin{aligned}
|\operatorname{tr}(A_i G_3) - y_i| &\le \frac{\mu}{b_i^5}(|\Delta_{\alpha_i}(b_i) - \widehat{\Delta}_i| + |\Delta_{\alpha_i}(b_i) - b_i^5 \operatorname{tr}(A_i G_3)/\mu|)\\
&\le \frac{\mu}{b_i^5}(\eta_i + \epsilon_{\rho^+}(b_i)) = \delta_i.
\end{aligned}$$

Thus $G_3 \in \mathcal{K}$. The target endpoints follow from

$$\begin{aligned}
l_\tau \le \operatorname{tr}(A_\tau G_3) &\le u_\tau,\\
\frac{B^5 l_\tau}{\mu} - \epsilon_{\rho^+}(B) \le \frac{B^5}{\mu}\operatorname{tr}(A_\tau G_3) - \epsilon_{\rho^+}(B) &\le \Delta_\tau(B),\\
\Delta_\tau(B) \le \frac{B^5}{\mu}\operatorname{tr}(A_\tau G_3) + \epsilon_{\rho^+}(B) &\le \frac{B^5 u_\tau}{\mu} + \epsilon_{\rho^+}(B).
\end{aligned}$$

Since $\mathcal{K} \subseteq \mathcal{K}_0$, minimization can only increase the lower endpoint and maximization can only decrease the upper endpoint when the same envelope is used. If exact optimization is unavailable, weights satisfying

$$\operatorname{diag}(a, b) + \sum_i \gamma_i A_i - \zeta A_{\text{target}} \succeq 0, \quad a, b \ge 0, \quad \zeta \in \{-1, 1\},$$

give the following upper endpoint:

$$\begin{aligned}
\zeta \operatorname{tr}(A_{\text{target}} G_3) &\le a(G_3)_{11} + b(G_3)_{22} + \sum_i \gamma_i \operatorname{tr}(A_i G_3)\\
&\le a R_X^2 + b R_Y^2 + \sum_i \gamma_i y_i + \sum_i |\gamma_i| \delta_i.
\end{aligned}$$

The matrix inequality is checked independently. The resulting interval is intersected with the prior interval. This is standard set-membership and semidefinite inference, used here with a derived rather than externally measured lower-order defect. An empty feasible set means that the observations and supplied bounds are inconsistent. All endpoint arithmetic is outward-rounded.

H.6 Initial containment from scalar observations

Let $\kappa(s) = \phi(s/\mu^2)/\mu$ and $\psi(s) = \mu\kappa(s)$. Assume

$$0 \le -2s\psi'(s)/\psi(s) \le 1 \quad (q_0^2 \le s \le \mu^2), \qquad g_0 = \nabla F(0), \quad q_0 = \|Wg_0\|, \quad \rho_0 = \|Pg_0\|.$$

All executed directions must be defined. Constant, power and regularized normalized gains satisfy this condition.

The first state coefficients and terminal Taylor expansion are

$$\begin{aligned} u_N &= -B\kappa(\mu^2)g_0 + O(B^2), & \bar{u}_N &= -B\kappa(q_0^2)Wg_0 + O(B^2), \\ \Delta_\alpha(B) &= g_0^\top(\bar{u}_N - u_N) + O(B^2) \\ &= B\{\kappa(\mu^2)\|g_0\|^2 - \kappa(q_0^2)\|Wg_0\|^2\} + O(B^2) \\ &= B\{\mu - q_0^2\kappa(q_0^2)\} + O(B^2). \end{aligned} \tag{85}$$

For $d(s) = s\kappa(s)$,

$$d'(s) = \kappa(s) + s\kappa'(s) = \kappa(s)\left(1 + \frac{s\psi'(s)}{\psi(s)}\right) \ge \kappa(s)/2,$$
$$\psi' \le 0, \quad s \le \mu^2 \Longrightarrow \kappa(s) \ge \kappa(\mu^2) = 1/\mu.$$

Integration therefore gives

$$\mu - q_0^2\kappa(q_0^2) = \int_{q_0^2}^{\mu^2} d'(s)\,ds \ge \int_{q_0^2}^{\mu^2} \frac{ds}{2\mu} = \frac{\mu^2 - q_0^2}{2\mu} = \frac{\rho_0^2}{2\mu}. \tag{86}$$

Hence $\Delta_\alpha = o(B^5) \Longrightarrow 0 = \lim_{B\downarrow 0} \Delta_\alpha/B \ge \rho_0^2/(2\mu) \Longrightarrow Wv = v$. The same argument applies to $o(B^7)$. For unit normalization, $q_0 > 0$ is required by the executed update.

There is also a finite one-step bound. For $h(q) = q\kappa(q^2)$,

$$h'(q) = \kappa(q^2)\left(1 + \frac{2q^2\psi'(q^2)}{\psi(q^2)}\right) \ge 0, \qquad h(q_0) \le h(\mu) = 1.$$

The full step has length b, and the retained step has length at most b. If L bounds the Hessian on both segments, then

$$\begin{aligned} \left|\Delta_1(b) - b\{\mu - q_0^2\kappa(q_0^2)\}\right| &\le \tfrac{1}{2}L\|u_1\|^2 + \tfrac{1}{2}L\|\bar{u}_1\|^2 \le Lb^2, \\ \frac{\rho_0^2}{2\mu} &\le \mu - q_0^2\kappa(q_0^2) \le \Delta_1(b)/b + Lb \le (\widehat{\Delta}_1 + \eta_1)/b + Lb. \end{aligned}$$

Intersecting with $\rho_0 \le \mu$ gives

$$\rho_0^2 \le \min\{\mu^2,\, 2\mu\max(0, (\widehat{\Delta}_1 + \eta_1)/b + Lb)\}. \tag{87}$$

If $\rho_0 \le \rho_0^+$, then $q_0^2 = \mu^2 - \rho_0^2 \ge \max(0, \mu^2 - (\rho_0^+)^2)$.

For a nonzero initial defect, define $F^\flat(x) = F(x) - \langle Pg_0, x\rangle$. When $q_0 > 0$, set $v_0 = -Wg_0/q_0$. For $y \in \operatorname{ran} W$,

$$W\nabla F^\flat(y) = W\nabla F(y), \qquad F^\flat(y) = F(y).$$

Thus the retained states and losses agree. Let common bounds D_0, D_1, U_1, G control the full fields, gain map and losses on their comparison domain. For the two full states,

$$\begin{aligned} \|f_F(x) - f_{F^\flat}(x)\| &\le U_1\|\nabla F(x) - \nabla F^\flat(x)\| = U_1\rho_0, \\ \|x_N - x_N^\flat\| &\le \sum_t B\alpha_t U_1\rho_0 \prod_{i>t}(1 + B\alpha_i D_1) \le BU_1\rho_0 e^{D_1 B}, \\ |\Delta_\alpha^F - \Delta_\alpha^{F^\flat}| &= |F^\flat(x_N^\flat) - F(x_N)| \\ &\le |F(x_N^\flat) - F(x_N)| + |\langle Pg_0, x_N^\flat\rangle| \\ &\le GBU_1\rho_0 e^{D_1 B} + \rho_0 D_0 B =: D_{\mathrm{init}}(B). \end{aligned}$$

Normalize the auxiliary speed by

$$c_0 = q_0\kappa(q_0^2), \qquad \kappa^\flat = \kappa/c_0, \qquad B^\flat = c_0 B.$$

Then $q_0\kappa^\flat(q_0^2) = 1$ and $B^\flat\kappa^\flat(s) = B\kappa(s)$. The auxiliary program therefore has exactly the original update sizes. Proposition 3 applies with anchor norm q_0. Its observation errors become $\eta_i + D_{\mathrm{init}}(b_i)$, and its target endpoints are enlarged by $D_{\mathrm{init}}(B)$. If q_0 is only interval-known, use one common nuisance value across all observations and take the union of its target intervals.

Finally, if $\|\widetilde{W} - W\| \le \rho_W$ and the same gain bounds hold on both gradient ranges,

$$\begin{aligned}\|U(\widetilde{W}g) - U(Wg)\| &\le U_1\|(\widetilde{W} - W)g\| \le U_1 G\rho_W,\\ \|\widetilde{y}_N - y_N\| &\le BU_1 G\rho_W e^{D_1 B},\\ |F(\widetilde{y}_N) - F(y_N)| &\le G\|\widetilde{y}_N - y_N\| \le G^2 U_1\rho_W B e^{D_1 B}.\end{aligned}$$

This error is included at each affected probe and target.

H.7 A RANK CERTIFICATE FROM 53-BIT INTERVAL OBSERVATIONS

This example uses only the paired-terminal interface. Calibration adds information unavailable from the same prior. Consider

$$F_{a,c}(x,y) = -x + \frac{a}{2}x^2 y + \frac{c}{6}y^3, \qquad a, c \ge 0, \quad \|D^3 F\| \le M = 13/100, \quad r_{\mathrm{dom}} = 1/2. \tag{88}$$

The prior gives this family, the coefficient signs, $D^2F(0) = 0$, and M. It does not give a, c separately. Since a, c are third-derivative contractions, $a, c \le M$. On the radius-$1/2$ ball,

$$\|D^2F(x)\| \le M\|x\| \le 13/200 =: L, \qquad \|\nabla F(x)\|,\ \|W_1\nabla F(x)\| \ge 1 - L/2 = 387/400 =: g_*.$$

The candidates are $W_1 = \mathrm{diag}(1, 0)$ and $W_2 = I$. The retained path is $(Bs_t, 0)$, so $\bar{J} = -B$. At the anchor, $PQv = 0$, $X = 0$, and $Y = ae_y$.

A homogeneous rescaling yields a much sharper remainder than a global derivative majorant. Write $x_t = B\xi_t$, $y_t = B\upsilon_t$, $\tau = aB^2$, and $\chi = cB^2$. The scaled recurrence is

$$\begin{aligned}A_t &= 1 - \tau\xi_t\upsilon_t, \quad D_t = (\tau\xi_t^2 + \chi\upsilon_t^2)/2, \quad q_t = D_t/A_t,\\ \xi_{t+1} &= \xi_t + \alpha_t(1 + q_t^2)^{-1/2},\\ \upsilon_{t+1} &= \upsilon_t - \alpha_t q_t(1 + q_t^2)^{-1/2}.\end{aligned} \tag{89}$$

For $\chi\tau \le 1$, the induction step is

$$\begin{aligned}0 \le \xi_t \le s_t, \quad -\tau s_t \le \upsilon_t \le 0 \Longrightarrow A_t \ge 1, \quad 0 \le D_t \le (\tau + \chi\tau^2)/2 \le \tau,\\ 0 \le q_t = D_t/A_t \le \tau,\\ 0 \le \xi_{t+1} \le \xi_t + \alpha_t \le s_{t+1},\\ -\tau s_{t+1} \le \upsilon_t - \alpha_t\tau \le \upsilon_{t+1} \le 0.\end{aligned}$$

Starting from $\xi_0 = \upsilon_0 = 0$, this proves

$$0 \le \xi_t \le s_t, \qquad -\tau s_t \le \upsilon_t \le 0, \qquad A_t \ge 1, \qquad 0 \le q_t \le \tau.$$

For the stated prior and $B \le 1/2$, $\chi\tau = acB^4 \le M^2/16 < 1$. The elementary bounds $0 \le 1 - (1 + q^2)^{-1/2} \le q^2/2$ and $|(1 + q^2)^{-1/2} - 1 + q^2/2| \le 3q^4/8$ give

$$\begin{aligned}|\xi_t - s_t| &\le \tau^2/2,\\ |q_t - \tau s_t^2/2| &\le \frac{\tau}{2}|\xi_t^2 - s_t^2| + \frac{\chi}{2}\upsilon_t^2 + D_t|A_t^{-1} - 1| \le \frac{3}{2}\tau^3 + \frac{1}{2}\chi\tau^2,\\ \left|\upsilon_N + \frac{\tau}{2}\sum_t \alpha_t s_t^2\right| &\le 2\tau^3 + \frac{1}{2}\chi\tau^2,\\ \left|\xi_N - 1 + \frac{\tau^2}{8}\sum_t \alpha_t s_t^4\right| &\le \frac{3}{2}\tau^4 + \frac{3}{8}\chi\tau^3.\end{aligned} \tag{90}$$

The last two estimates use the following full sums:

$$\left|\upsilon_N + \frac{\tau}{2}\sum_t \alpha_t s_t^2\right| = \left|\sum_t \alpha_t \left(\frac{\tau s_t^2}{2} - \frac{q_t}{\sqrt{1+q_t^2}}\right)\right|$$
$$\leq \sum_t \alpha_t \left(\left|q_t - \frac{\tau s_t^2}{2}\right| + q_t\left|1 - \frac{1}{\sqrt{1+q_t^2}}\right|\right) \leq 2\tau^3 + \chi\tau^2/2,$$
$$|q_t^2 - \tau^2 s_t^4/4| \leq (q_t + \tau s_t^2/2)|q_t - \tau s_t^2/2| \leq \frac{3\tau}{2}\left(\frac{3}{2}\tau^3 + \frac{1}{2}\chi\tau^2\right),$$
$$\left|\xi_N - 1 + \frac{\tau^2}{8}\sum_t \alpha_t s_t^4\right| \leq \frac{1}{2}\sum_t \alpha_t |q_t^2 - \tau^2 s_t^4/4| + \frac{3}{8}\sum_t \alpha_t q_t^4 \leq \frac{3}{2}\tau^4 + \frac{3}{8}\chi\tau^3.$$

The terminal scaled loss difference is exactly

$$\Delta_\alpha(B)/B = \xi_N - 1 - \tfrac{\tau}{2}\xi_N^2\upsilon_N - \tfrac{\chi}{6}\upsilon_N^3.$$

The product and output errors are

$$\left|\xi_N^2\upsilon_N + \frac{\tau}{2}\sum_t \alpha_t s_t^2\right| \leq |\xi_N^2 - 1||\upsilon_N| + \left|\upsilon_N + \frac{\tau}{2}\sum_t \alpha_t s_t^2\right| \leq 3\tau^3 + \chi\tau^2/2,$$
$$\left|\frac{\Delta_\alpha(B)}{B} - \tau^2\sum_t \alpha_t(s_t^2/4 - s_t^4/8)\right| \leq \left(\frac{3}{2}\tau^4 + \frac{3}{8}\chi\tau^3\right) + \frac{\tau}{2}\left(3\tau^3 + \frac{1}{2}\chi\tau^2\right) + \frac{\chi}{6}\tau^3$$
$$= 3\tau^4 + \frac{19}{24}\chi\tau^3.$$

Substituting $\tau = aB^2$, $\chi = cB^2$ gives

$$\begin{aligned}\Delta_\alpha(B) &= a^2B^5\sum_t \alpha_t(s_t^2/4 - s_t^4/8) + R_\alpha(B),\\ |R_\alpha(B)| &\leq [3a^4 + (19/24)ca^3]B^9.\end{aligned} \tag{91}$$

No constant depends on the profile length. For equal steps, the sum is $k_N := N^{-1}\sum_{t=0}^{N-1}\{(t/N)^2/4 - (t/N)^4/8\}$; $k_2 = 7/256$.

If $a \leq A$ and $c \leq M$, put $D(A) = 3A^4 + (19/24)MA^3$. For $\Delta_2(b) \in [l, u]$,

$$l - D(A)b^9 \leq k_2a^2b^5 \leq u + D(A)b^9.$$

The upper and lower updates are

$$\begin{aligned}A_0 &= M, \qquad A_{j+1} = \min\left(A_j, \sqrt{\frac{u + D(A_j)b^9}{k_2b^5}}\right),\\ a_- &= \sqrt{\max\left(0, \frac{l - D(A_j)b^9}{k_2b^5}\right)}.\end{aligned} \tag{92}$$

Upper square roots are rounded upward; lower roots are rounded downward. The validity induction is

$$\begin{aligned}a \leq A_0 &= M,\\ a \leq A_j &\Longrightarrow |R_2(b)| \leq D(A_j)b^9\\ &\Longrightarrow a \leq \sqrt{\frac{u + D(A_j)b^9}{k_2b^5}},\\ a &\leq \min\left(A_j, \sqrt{\frac{u + D(A_j)b^9}{k_2b^5}}\right) = A_{j+1}.\end{aligned}$$

For validation use $a = 1/100, c = 1/10$. The tensor norm obeys $\|D^3F\| \le 3a + c = 13/100$. A two-step observation at $b = 1/10$ uses directed interval arithmetic with a 53-bit binary significand. Its enclosure is

$$[l, u] \subseteq [2.73437106290685, 2.73437938957955] \times 10^{-11}.$$

The archive preserves the exact binary endpoints. Their half-width is $4.163336342344337 \times 10^{-17}$. Rational certification arithmetic gives

$$A_1 = 0.0101961080418, \quad A_2 = 0.01000003390354,$$
$$A_3 = 0.01000003233164, \quad A_4 = 0.01000003233163.$$

and $a \ge 0.00999996849567$. The coefficient-sensitive target remainder at $B = 1/2$ is then below 2.596056×10^{-10}, compared with the prior remainder 2.115114×10^{-6}. For $H = 20$, (91) gives the outward interval

$$\Delta_{20}(1/2) \in [1.72265430412552, 1.72786844230886] \times 10^{-7}. \tag{93}$$

The actual target is evaluated only after forming this certificate and lies near $1.725261492417945 \times 10^{-7}$.

The same prior permits $a' = 1/25, c' = 1/100$, since $3a' + c' = 13/100$. The prior information does not distinguish these coefficients from the actual coefficients. Their target obeys

$$\Delta'_{20}(1/2) \ge k_{20}(a')^2(1/2)^5 - D(a')(1/2)^9 > 2.73255 \times 10^{-6}.$$

Thus a prior-only certificate cannot accept W_1 at tolerance 10^{-6} for every compatible loss. The calibrated interval (93) accepts W_1 for the actual loss. The full projector W_2 always has zero bias. The least certified dimension therefore changes from two to one.

Here $Q = 0$ is part of the prior. Only a^2 remains in the leading response, so one two-step probe suffices. Direct access to $\mathsf{T}_3[v, v]$ would reveal a more cheaply. The information gain concerns the stated scalar interface. The precision claim concerns directed 53-bit intervals, not unmodified hardware loss subtraction.