WHEN DOES LEARNING-RATE RETUNING HELP? ATTAINABLE OUTPUT GEOMETRY UNDER A FIXED UPDATE BUDGET

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning-rate schedules are often tuned for one training setting and reused when models are adapted to new data under fixed update budgets. A finite-step optimization trajectory can itself act as a statistical regularizer. We study when retuning the schedule after a task change improves on preserving the nominal finite-step filter. We characterize the leading retuned-minus-preserved risk through the statistical distance between the oracle drift and the one-sided attainable output set. For K distinct interior steps and a simple spectrum in p dimensions, the attainable correction space has dimension $\min(K, p)$. Some spectral corrections are therefore inaccessible when $K < p$, whereas the unrestricted commuting optimum becomes locally attainable when $K \geq p$. At equal steps, $O(\sqrt{|r|})$ step splitting generates first-order output corrections missed by the parameter Jacobian, where r indexes the local task change. For two steps, this singular geometry is the endpoint of a near-equal transition governed by $\delta_r^2/|r|$, where δ_r is the nominal half-gap. These results explain why unrestricted spectral retuning can overestimate attainable corrections while local Jacobian sensitivity can underestimate them. Synthetic experiments support the theoretical predictions, while real-data diagnostics distinguish preservation, schedule reuse, and retuning.

1 INTRODUCTION

Learning-rate schedules are often reused when models are adapted to new data under fixed update budgets, as in continued pretraining and fine-tuning. Finite-step optimization can itself provide statistical regularization (Yao et al., 2007; Raskutti et al., 2014; Ali et al., 2019). In linear models, the schedule determines a filter that retains or suppresses components of the input. Preserving this filter or retuning after a task change therefore affects the estimator and its test risk.

Existing analyses of early stopping, spectral filtering, and schedule design explain the statistical effects of particular trajectories, filters, or learning rates (Bauer et al., 2007; Sonthalia et al., 2024; Velikanov et al., 2024; Tang et al., 2025). The remaining question is when fixed-budget schedule retuning improves on preserving the nominal regularizer after a task change. The difficulty is to identify the full set of filters that admissible schedules can produce after the task changes. Allowing each spectral component to change independently may include corrections that no schedule can realize. Conversely, the parameter Jacobian measures first-order sensitivity to learning rates and can miss attainable corrections at equal steps.

To this end, our work addresses the following central question:

When does fixed-budget retuning improve on preserving the nominal finite-step regularizer?

We answer this question by characterizing the full set of output changes that the fixed-budget schedule family can realize. The leading retuned-minus-preserved risk depends on how closely these attainable changes approximate the oracle drift, the first-order change in the optimal linear filter. This comparison explains why the spectral benchmark can overestimate attainable corrections, while the Jacobian can underestimate them. Equal steps arise in constant-learning-rate phases, including the stable phase of warmup–stable–decay schedules (Wen et al., 2025). Small perturbations of individual

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.

step sizes produce near-equal schedules. In a two-step linear model, we show how their attainable output geometry interpolates between the singular and regular regimes.

We distinguish filter preservation from schedule reuse on the changed task, and study population-optimal retuning around a nominally optimal linear filter (Section 3). This oracle comparison uses the task distribution to select the best schedule; the budget counts updates, excluding search and estimation costs.

Our Contributions. Our contributions are threefold.

1. **When retuning lowers risk.** We give a local risk criterion for deciding between retuning and preservation. It compares the oracle drift with the full one-sided set of attainable output changes, for both regular and singular schedules.
2. **Which corrections a fixed budget permits.** For K pairwise distinct interior steps and a simple spectrum in p dimensions, we prove that the correction space has dimension $\min(K, p)$. We quantify the risk cost of unavailable corrections when $K < p$ and show that the current commuting optimum becomes locally attainable when $K \geq p$. For equal steps, square-root splitting creates one-sided corrections and decision reversals missed by the Jacobian. For two steps, a half-gap-dependent risk formula connects this singular limit to regular geometry.
3. **Tests of the risk predictions.** Controlled experiments test the risk coefficients, both decision-reversal mechanisms, and the error rates at finite perturbations. We also examine near-equal schedules and noncommuting statistical weights. On real data, we compare preservation, schedule reuse, and validation-selected retuning.

2 Related Work

We build on studies of finite-computation regularization and spectral adaptation under task change. Work on restricted parameterizations provides the geometric tools for describing which filter changes a schedule can produce.

Finite computation and schedule design. Iterative regularization links finite computation to spectral filters (Landweber, 1951; Engl et al., 1996; Bauer et al., 2007; Lo Gerfo et al., 2008), and early-stopped least squares and boosting establish its statistical value (Yao et al., 2007; Bühlmann & Yu, 2003; Zhang & Yu, 2005; Raskutti et al., 2014; Wei et al., 2019). Discrepancy principles, cross-validation, and finite-iterate risk rules help choose where to stop along a trajectory (Blanchard et al., 2018; Celisse & Wahl, 2021; Patil et al., 2024b; Bellec & Tan, 2026). Gradient-flow and spectral-risk analyses characterize the resulting filters (Ali et al., 2019; 2020; Velikanov et al., 2024; Kanade et al., 2023; Wegel et al., 2026; Wu et al., 2026). Sonthalia et al. (2024) identify the regularizer of one finite schedule. Tang et al. (2025) design schedules under an iteration budget; Li et al. (2026) derive optimal schedules under functional scaling laws at a fixed horizon. Our comparison uses the nominal finite-step filter as a reference and measures the statistical value of retuning the full K-step family.

Spectral benchmarks and the fixed-budget restriction. Spectral and out-of-distribution analyses relate regularization to covariance change (Patil et al., 2024a; LeJeune et al., 2024; Fan et al., 2025). Rotation-equivariant shrinkage adjusts spectral coefficients while keeping the curvature eigenbasis (Ledoit & Wolf, 2021). These independently adjustable coefficients give a commuting benchmark for retuning. A K-step schedule, however, couples the coefficients through a single polynomial. Our rank and restriction-cost results identify which benchmark corrections remain attainable under this budget constraint.

Parameterizations and singular output geometry. Matrix perturbation theory describes filter motion under operator change (Kato, 1995; Higham, 2008; Greenbaum et al., 2020). Polynomial-factorization analyses link repeated factors to the function-space geometry of linear convolutional networks (Kohn et al., 2022; 2024). Smooth-parameterization theory distinguishes optimization in parameter space from optimization over its image (Levin et al., 2025). We study positive-step filters as the operator changes, identify their complete one-sided output limits over the step box, and compare their risk with preservation. The risk calculation uses classical quadratic projection (Chernoff, 1954; Geyer, 1994; Drton, 2009); the schedule analysis determines which directions this projection must include.

3 PRELIMINARIES

We fix the update budget at K and ask when retuning the learning-rate schedule after a local task change lowers prediction risk relative to preserving the nominal filter.

Notations. We use p for the state dimension and K for the fixed number of updates. A schedule $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K) \in \mathbb{R}^K$ assigns one learning rate to each update, and $\mathcal{A}$ is its admissible set. The scalar r indexes the task change, with $r = 0$ the nominal task. For $r \neq 0$, let $s = \operatorname{sign}(r) \in \{-1, +1\}$. The notation $r \to 0^s$ means convergence to zero through values with sign s; all such limits keep p and K fixed. We write I for the identity matrix and Id for the identity operator on matrix space. For a matrix X, we write $X^\top$ for its transpose and $\operatorname{tr}(X)$ for its trace; $X \succ 0$ means symmetric positive definite. We use $\|v\|_2$, $\|X\|_F$, and $\|X\|_{\mathrm{op}}$ for the Euclidean vector norm, Frobenius matrix norm, and induced Euclidean operator norm, respectively. The operators range and span denote a linear image and the span of listed directions. The subscript 0 on an inner product, norm, or distance refers to the nominal statistical metric defined below; projections use that metric unless stated otherwise. The symbols $O(\cdot), o(\cdot)$ denote deterministic local remainders, while $O_p(\cdot), o_p(\cdot)$ denote their probabilistic counterparts. A C^2 path has two continuous derivatives with respect to r.

Problem Setup. We estimate a square-integrable target $u \in \mathbb{R}^p$ from a task-dependent observation $z_r \in \mathbb{R}^p$ with finite second moments. Along this task path, we hold the marginal distribution of u fixed while allowing the observation mechanism to change. A linear filter $F \in \mathbb{R}^{p \times p}$ predicts u using Fz_r. We condition on any independent training information and hold F fixed when evaluating risk on (u, z_r). We call such filters response-independent and evaluate their population risk as $\mathcal{R}_r(F) = \mathbb{E}\|Fz_r - u\|_W^2$ for a fixed metric $W \succ 0$. Here, $\|v\|_W^2 = v^\top W v$, and the expectation is over (u, z_r). We write $\Sigma_r = \mathbb{E}[z_r z_r^\top] \succ 0$ for the observation's second moment. The unrestricted optimal linear filter is $F_r^\star = \mathbb{E}[uz_r^\top]\Sigma_r^{-1}$; it need not be symmetric.

We use $\mathcal{F}_r$ for the current admissible filter family and $F_0 \in \mathcal{F}_0$ for the nominal filter. Our comparison is between population-optimal retuning over $\mathcal{F}_r$ and preservation of a reference $\widehat{F}_{0,r}$. This reference is F_0 or an independently learned approximation, giving the risk difference

$$\Delta R_r = \inf_{F \in \mathcal{F}_r} \mathcal{R}_r(F) - \mathcal{R}_r(\widehat{F}_{0,r}). \tag{1}$$

Negative ΔR_r favors retuning; positive ΔR_r favors preservation. Centering at $F_0 = F_0^\star$ removes pre-existing nominal approximation error; Appendix E.1 gives Gaussian realizations with a fixed target distribution. We measure filter changes by their effect on prediction using $\langle X, Y \rangle_0 = \operatorname{tr}(X^\top W Y \Sigma_0)$, $\|X\|_0^2 = \langle X, X \rangle_0$, and $\operatorname{dist}_0(X, \mathcal{C}) = \inf_{Y \in \mathcal{C}} \|X - Y\|_0$. All projections use this nominal risk metric unless stated otherwise.

Finite-Step Spectral Regularization. Finite-step gradient descent provides a classical instance of an implicit spectral regularizer (Landweber, 1951; Engl et al., 1996; Bauer et al., 2007; Sonthalia et al., 2024). We write $S_r \in \mathbb{R}^{p \times p}$ for the task's curvature operator. Starting from $x_0 = 0$, we apply $x_{j+1} = x_j + \alpha_{j+1} S_r(z_r - x_j)$ for $j = 0, \dots, K-1$. We represent the resulting dependence on the schedule through the following filter.

Definition 1 (Finite-step response filter)**.** For a schedule $\boldsymbol{\alpha} \in \mathcal{A} \subset \mathbb{R}^K$, we define the finite-step filter by

$$\begin{aligned} F_{\boldsymbol{\alpha}}(S_r) &:= I - \prod_{k=1}^{K} (I - \alpha_k S_r), \\ x_K &= F_{\boldsymbol{\alpha}}(S_r) z_r. \end{aligned} \tag{2}$$

A matrix product with no factors is interpreted as I. The current retuning family is $\mathcal{F}_r = \{F_{\boldsymbol{\alpha}}(S_r) : \boldsymbol{\alpha} \in \mathcal{A}\}$, and a nominal schedule $\boldsymbol{\alpha}_0 \in \mathcal{A}$ defines $F_0 = F_{\boldsymbol{\alpha}_0}(S_0)$.

For symmetric S_r, an eigendirection with eigenvalue λ is multiplied by $1 - \prod_{k=1}^{K}(1 - \alpha_k \lambda)$. Thus one schedule couples regularization across spectral directions. Preservation keeps F_0, schedule reuse applies $F_{\boldsymbol{\alpha}_0}(S_r)$, and population-optimal retuning minimizes $\mathcal{R}_r(F_{\boldsymbol{\alpha}}(S_r))$ over $\boldsymbol{\alpha} \in \mathcal{A}$. The compact step boxes below ensure a minimizer and allow all K step coordinates to vary independently. Reuse is feasible for retuning. Preservation can lie outside the current family, so its risk can be

lower than the retuned optimum. The general comparison in (1) also permits independently learned references and other response-independent filter families.

Standing statistical conditions. We keep p, K fixed and center at $F_0 = F_0^\star$, with finite second moments, $W, \Sigma_0 \succ 0$, $\Sigma_r = \Sigma_0 + O(r)$, and $F_r^\star = F_0 + rD + O(r^2)$. Here D is the oracle drift. For the leading comparison, we assume a reference error of $o(|r|)$, or $o_p(|r|)$ for an independently learned reference. Each theorem states its additional schedule conditions. Appendix A explains their roles and scope.

4 Main Results

We first express retuning risk through attainable output velocities, then identify these velocities for distinct and equal schedules.

4.1 Risk from Attainable Output Velocities

We compare the oracle drift D with changes that admissible schedules can produce. Under $F_r^\star = F_0 + rD + O(r^2)$, we consider each one-sided path $r \to 0^s$ and measure filter changes per unit task change. This includes changes generated by nonlinear schedule paths.

Definition 2 (Attainable output velocities)**.** For $r \neq 0$, let

$$\mathcal{Z}_r = \left\{ \frac{F - F_0}{r} : F \in \mathcal{F}_r \right\}. \tag{3}$$

Fix a side $s \in \{-1, +1\}$. We say that $\mathcal{Z}_r$ converges to V_s as $r \to 0^s$ when the following inner and outer conditions hold. For every $Z \in V_s$ and every sequence $r_n \to 0^s$, there exist $Z_n \in \mathcal{Z}_{r_n}$ such that $Z_n \to Z$. Conversely, whenever $r_n \to 0^s$ and $Z_n \in \mathcal{Z}_{r_n}$ converge to a matrix Z, that matrix belongs to V_s. The elements of V_s are the attainable output velocities on side s.

These are the conditions of Painlevé–Kuratowski set convergence (Rockafellar & Wets, 1998). We use this limit to express the leading risk difference as a distance from the oracle drift.

Theorem 4.1 (Attainable-output risk principle)**.** *Consider the comparison in* (1)*, with finite second moments and fixed $W \succ 0$. Suppose that $\Sigma_0 \succ 0$, $\Sigma_r = \Sigma_0 + O(r)$, and $F_r^\star = F_0 + rD + O(r^2)$. Let the reference satisfy $\|\widehat{F}_{0,r} - F_0\|_0 = o(|r|)$. For each side s, suppose that $\mathcal{Z}_r$ converges to a nonempty closed set V_s in the inner–outer sense described above. Then, as $r \to 0^s$,*

$$\begin{aligned} \Delta R_r &= r^2 \Gamma_s(D) + o(r^2), \\ \Gamma_s(D) &:= \operatorname{dist}_0^2(D, V_s) - \|D\|_0^2. \end{aligned} \tag{4}$$

The limiting set V_s may be nonconvex. For an independently learned reference with $o_p(|r|)$ error, the expansion holds in probability.

Discussion of Theorem 4.1. Preservation applies no first-order correction, whereas retuning selects the attainable velocity closest to D. Consequently, retuning helps locally when $\operatorname{dist}_0(D, V_s) < \|D\|_0$ and hurts when the inequality is reversed. If $0 \in V_s$, retuning can reproduce preservation to first order, so the leading coefficient is nonpositive. A positive coefficient is possible only when this zero correction is unattainable. At a zero coefficient, higher-order terms decide.

4.2 Budget-Limited Corrections at Distinct Steps

At pairwise distinct interior steps and a simple spectrum, the schedule map is locally regular: its derivative captures all local first-order schedule corrections. For the gradient filter in (2), we separate these corrections from the filter motion induced by changing curvature. Let $E = S_0'$, and let $\boldsymbol{\alpha}_0 = (\alpha_{0,1}, \ldots, \alpha_{0,K})$ be the nominal schedule. The step directions are $A_j = S_0 \prod_{\ell \neq j} (I - \alpha_{0,\ell} S_0)$, and the fixed-schedule velocity is

$$G_K = \sum_{j=1}^{K} \alpha_{0,j} \left\{ \prod_{\ell < j} (I - \alpha_{0,\ell} S_0) \right\} E \left\{ \prod_{\ell > j} (I - \alpha_{0,\ell} S_0) \right\}. \tag{5}$$

The order around E matters when the curvature change does not commute with S_0. Each learning rate may vary independently within the same interval; their admissible product set is the step box. The next theorem identifies all first-order corrections attainable over this box.

Theorem 4.2 (Distinct-step attainable geometry)**.** *Let S_r be a symmetric C^2 path with $S_r = S_0 + rE + O(r^2)$, where $S_0 \succ 0$ has p distinct eigenvalues, that is, a simple spectrum. Consider the filters generated by* (2) *with independently varying steps in $[\underline{\alpha}, \overline{\alpha}]^K$, where $0 < \underline{\alpha} < \overline{\alpha} < 1/\lambda_{\max}(S_0)$. Suppose that the nominal steps are pairwise distinct and interior to this box. Then the full current family has $V_+ = V_- = G_K + T_K$, where*

$$T_K := \operatorname{span}\{A_j : 1 \le j \le K\} = \{S_0 q(S_0) : \deg q \le K-1\}, \qquad \dim(T_K) = \min(K, p). \tag{6}$$

Here, q ranges over real polynomials of degree at most $K-1$.

Under the statistical-path and reference-accuracy conditions of Theorem 4.1, its risk coefficient is $\Gamma_K = \operatorname{dist}_0^2(D, G_K + T_K) - \|D\|_0^2$.

Discussion of Theorem 4.2. Relative to the same nominal filter, $K < p$ leaves some spectral corrections unavailable; at $K \ge p$, the restriction disappears. This threshold concerns the current eigenbasis, so eigenvector changes can still affect risk.

Commuting benchmark and restriction cost. The commuting benchmark allows any symmetric filter satisfying $FS_r = S_r F$. To measure the restriction, let $\mathcal{C}_0 = \{F = F^\top : FS_0 = S_0 F\}$ and let Π be its weighted orthogonal projector. Write $S_0 = U \operatorname{diag}(\lambda_1, \dots, \lambda_p) U^\top$ and $F_0 = U \operatorname{diag}(f_1, \dots, f_p) U^\top$. The curvature-induced spectral transport is defined by

$$(U^\top G_{\mathrm{comm}} U)_{ij} = \begin{cases} \dfrac{f_i - f_j}{\lambda_i - \lambda_j} (U^\top E U)_{ij}, & i \ne j, \\ 0, & i = j. \end{cases}$$

The corresponding commuting coefficient is $\Gamma_{\mathrm{comm}} = \operatorname{dist}_0^2(D, G_{\mathrm{comm}} + \mathcal{C}_0) - \|D\|_0^2$. Here, Π projects onto all commuting corrections, whereas P_T below keeps only those accessible to the schedule. Their difference measures the restriction cost.

Proposition 4.1 (Cost of the fixed update budget)**.** *Under the geometric and statistical conditions of Theorem 4.2, let P_T be the orthogonal projector onto T_K in the nominal risk metric. Then*

$$\Gamma_K = \Gamma_{\mathrm{comm}} + \|(\Pi - P_T)(D - G_K)\|_0^2. \tag{7}$$

The two coefficients are equal if and only if $\Pi(D - G_K) \in T_K$. In particular, when $K \ge p$, the restriction term vanishes and the current population-optimal symmetric commuting filter is attained by the step family for all sufficiently small r.

The additional square is exactly the statistical cost of corrections allowed by the commuting benchmark but inaccessible to the K-step schedule family. It can reverse a favorable commuting prediction: the two-step construction in Appendix E.1 satisfies

$$\Gamma_{\mathrm{comm}} = -\frac{19}{1600} < 0 < \frac{19}{200} = \Gamma_2. \tag{8}$$

Thus independently tunable spectral coefficients can recommend retuning even when the full fixed-budget family favors preservation.

4.3 Hidden Corrections at Equal Steps

At an equal nominal schedule, all first-order step derivatives coincide. We therefore expand the filter to second order in the learning rates. Consider $K \ge 2$ and an equal nominal schedule $\boldsymbol{\alpha}_0 = a\mathbf{1}$, where $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^K$. Set $J = I - aS_0$, $A_K = S_0 J^{K-1}$, and $H_K = S_0^2 J^{K-2}$. We write $\mathbb{R}_+ = [0, \infty)$. Writing $\delta_j = \alpha_j - a$ and $e_1 = \sum_j \delta_j$, the filter expansion is

$$F_{a\mathbf{1}+\delta}(S_0) - F_0 = e_1 A_K + \tfrac{1}{2}(\|\delta\|_2^2 - e_1^2) H_K + O(\|\delta\|_2^3). \tag{9}$$

A zero-sum split increases some learning rates and decreases others while keeping their sum fixed. At size $\sqrt{|r|}$, its quadratic effect is of order $|r|$, although its linear effect vanishes. The next theorem shows that these movements exhaust the additional directions available to the full step box.

Let P project onto $\mathrm{span}\{A_K\}$ in the nominal risk metric, and evaluate G_K at $a\mathbf{1}$. Define the missing direction b, its alignment t, and the Jacobian risk coefficient by

$$b := (\mathrm{Id} - P)H_K, \qquad t := \langle b, D - G_K\rangle_0, \qquad t_\pm := \max(\pm t, 0),$$
$$\Gamma_{\mathrm{Jac}} := \mathrm{dist}_0^2(D, G_K + \mathrm{span}\{A_K\}) - \|D\|_0^2.$$

Theorem 4.3 (Equal-step attainable geometry)**.** *Let $K \geq 2$, and let S_r be a symmetric C^2 path with $S_r = S_0 + rE + O(r^2)$, where $S_0 \succ 0$ has at least two distinct eigenvalues. Consider the filters generated by* (2) *with independently varying steps in $[\underline{\alpha}, \overline{\alpha}]^K$, where $0 < \underline{\alpha} < a < \overline{\alpha} < 1/\lambda_{\max}(S_0)$, and let the nominal schedule be $a\mathbf{1}$. The full current family has the following translations of one-sided cones:*

$$V_+ = G_K + \mathrm{span}\{A_K\} + \mathbb{R}_+ H_K,$$
$$V_- = G_K + \mathrm{span}\{A_K\} - \mathbb{R}_+ H_K. \tag{10}$$

Suppose also that the statistical-path and reference-accuracy conditions of Theorem 4.1 hold. The matrices A_K and H_K are linearly independent under the spectral assumption, so $b \neq 0$. The one-sided risk coefficients satisfy

$$\Gamma_{K,\pm} = \Gamma_{\mathrm{Jac}} - \frac{t_\pm^2}{\|b\|_0^2}. \tag{11}$$

Discussion of Theorem 4.3. The additional direction H_K comes from the quadratic effect of splitting repeated steps. Its sign is one-sided because the nonnegative squared split is divided by r. If the steps must remain equal, or each step changes only by $O(r)$, this additional ray is absent.

A decision reversal missed by the Jacobian. The squared term in (11) is the gain from the additional ray. It is positive when the ray points toward the remaining oracle drift and can reverse the decision. Appendix E.1 gives an equal two-step example:

$$\Gamma_{2,+} = -\frac{1}{1600} < 0 < \frac{3}{200} = \Gamma_{\mathrm{Jac}}, \tag{12}$$

while the left-side coefficient is $3/200$. Thus the Jacobian can recommend preservation when independent fixed-budget retuning is beneficial.

We next ask how the geometry changes when two nominal steps are close but unequal. The relevant scale is the squared nominal half-gap relative to the task perturbation magnitude.

Proposition 4.2 (Two-step near-equal transition)**.** *Use the curvature and step-box conditions of Theorem 4.3 with $K = 2$. Let $\delta_r \geq 0$, $\delta_r \to 0$, $\bar{F}_r = F_{(a-\delta_r, a+\delta_r)}(S_0)$, and $\delta_r^2/|r| \to \rho \in [0, \infty]$ as $r \to 0^s$. Write $A = A_2$, $H = H_2$, and $G = G_2$ at (a, a). The complete limits of $(F - \bar{F}_r)/r$ over the step family are*

$$V_{s,\rho} = G + \mathrm{span}\{A\} + \{qH : sq \geq -\rho\},$$

where $\rho = \infty$ removes the constraint on q. Along this joint limit (r, δ_r), suppose $\Sigma_{r,\delta_r} \to \Sigma_0 \succ 0$ and $F^\star_{r,\delta_r} = \bar{F}_r + rD + o(r)$, with fixed $W \succ 0$ and finite second moments. Evaluate $b, t, \Gamma_{\mathrm{Jac}}$ in the limiting metric. For exact preservation versus population-optimal retuning, $\Delta R_{r,\delta_r} = r^2\Gamma_{s,\rho} + o(r^2)$, where

$$\Gamma_{s,\rho} = \Gamma_{\mathrm{Jac}} - \frac{t^2}{\|b\|_0^2} + \|b\|_0^2 \left(-\frac{st}{\|b\|_0^2} - \rho\right)_+^2.$$

The last term is zero at $\rho = \infty$.

At $\rho = 0$, we recover the equal-step cone. For $0 < \rho < \infty$, a one-sided constraint truncates the regular two-direction affine space. At $\rho = \infty$, this constraint disappears. The coefficient is continuous and nonincreasing in ρ, reaching the regular two-direction value once $\rho \geq (-st/\|b\|_0^2)_+$. Thus exact equality is an endpoint of a transition at half-gap scale $\sqrt{|r|}$. Appendix D.4 proves the result and gives a sign reversal at $\rho = 1$ using the existing example in (12).

4.4 FINITE-PERTURBATION ACCURACY

We next quantify how quickly the leading coefficients approximate finite-radius risk in the regular and singular regimes.

Corollary 4.1 (Finite-perturbation accuracy)**.** *Suppose that the conditions of Theorem 4.2 or Theorem 4.3 hold with the Taylor, chart, and nondegeneracy bounds on a fixed neighborhood specified in Appendix A, and use the exact reference F_0. Then*

$$\Delta R_r = \begin{cases} r^2\Gamma_K + O(|r|^3), & \textit{distinct steps}, \\ r^2\Gamma_{K,s} + O(|r|^{5/2}), & \textit{equal steps}. \end{cases} \tag{13}$$

For equal steps with $K = 2$, the remainder improves to $O(|r|^3)$. For a reference error $\epsilon_r = \|\widehat{F}_{0,r} - F_0\|_0$, add $O(|r|\epsilon_r + \epsilon_r^2)$ to the corresponding bound.

Constants depend on the fixed instance and the bounds in Appendix A. At two equal steps, algebraic cancellation improves the rate while leaving the one-sided direction intact; three equal steps attain the general singular order (Appendix E.2). The reference term quantifies the accuracy needed for these rates; $o(|r|)$ suffices for the leading coefficient.

4.5 PROOF STRATEGY

We first localize near-optimal filters and express their risk using distance from the limiting output set. We then identify this set for distinct and equal schedules.

Statistical localization. Completing the square gives the exact excess-risk identity:

$$\mathcal{R}_r(F) - \mathcal{R}_r(F_r^\star) = \operatorname{tr}((F - F_r^\star)^\top W (F - F_r^\star)\Sigma_r).$$

Lemma 4.1 (Localization of near-optimal filters)**.** *Under the conditions of Theorem 4.1, let $F_r \in \mathcal{F}_r$ satisfy $\mathcal{R}_r(F_r) \le \inf_{F\in\mathcal{F}_r} \mathcal{R}_r(F) + o(r^2)$ as $r \to 0^s$. Then $\|F_r - F_0\|_0 = O(|r|)$, so its scaled displacement $(F_r - F_0)/r$ remains bounded.*

An attainable candidate has $O(r^2)$ excess risk. Positive definiteness therefore keeps near-optimal outputs within $O(|r|)$ of the oracle and hence of F_0. On this bounded velocity scale, risk converges uniformly to squared distance from D. Outer completeness gives a lower bound; attainable paths give the matching upper bound. Subtracting preservation's risk proves Theorem 4.1 (Appendix B.1).

Algorithmic reachability. For distinct steps, polynomial identification localizes the schedules when $K \le p$; a full-row-rank chart attains every nearby commuting filter when $K \ge p$. Together these identify all velocities of the family (Appendix C). The derivative polynomials span the degree-$K-1$ correction class in (6). Evaluating this class on the p spectral nodes gives exactly $\min(K, p)$ independent corrections, which explains the threshold at $K = p$.

For equal steps, global isolation rules out other nominal representations. The next lemma then bounds every schedule producing an $O(|r|)$ output change.

Lemma 4.2 (Square-root scale at equal steps)**.** *Under the gradient-family conditions of Theorem 4.3, suppose $\boldsymbol{\alpha}_r \in [\underline{\alpha}, \overline{\alpha}]^K$ and $\|F_{\boldsymbol{\alpha}_r}(S_r) - F_0\|_0 = O(|r|)$. Writing $\delta_r = \boldsymbol{\alpha}_r - a\mathbf{1}$, we have $|\sum_j \delta_{r,j}| + \|\delta_r\|_2^2 = O(|r|)$.*

The common $O(r)$ shift gives a free direction; dividing the nonnegative squared split by r gives a ray with sign s. The Jacobian retains only the common-shift direction A_K. Removing its component from H_K gives b, while t measures how this missing direction aligns with the oracle drift left after the Jacobian projection. Only the sign-compatible part $t_\pm$ can be used on each side of the task path. Once these attainable sets are identified, both risk formulas reduce to weighted projections of D; unavailable directions remain as squared residual terms. Localization excludes other velocities, while explicit split paths attain the cone. Projection then gives Theorem 4.3 (Appendix D.1). Retaining the nominal squared half-gap gives the shifted constraint in Proposition 4.2. Quantitative output approximations give Corollary 4.1; the exact two-step expansion improves the singular rate (Appendices C.3 and D.3).

5 EXPERIMENTS

We first check whether the coefficients predict the size and sign of the risk gap, including the benchmark reversals (Figure 1). We then test finite-radius error rates (Figure 2) and examine near-equal schedules alongside the three adaptation choices (Figure 3). Appendix I gives the complete protocols and results.

Setup. We use 162 Gaussian population instances: 108 distinct-step and 54 equal-step, with $p \in \{2, 3, 4, 5\}$ and $K \in \{2, 3\}$ in the applicable cells. Seven perturbation magnitudes from 0.1 to 0.0015625, each with both signs, yield 2,268 comparisons. All methods share exact response moments and the risk metric; full step-box retuning uses globally checked polynomial optimization, with search costs outside the update budget. Appendix I.1 gives the prespecified protocol and solver verification.

Local risk coefficients and benchmark reversals. At the smallest radius, the maximum coefficient error is 4.80×10^{-7} and all 324 signs agree (Figure 1(a,b)). The commuting benchmark is too optimistic at distinct steps and the Jacobian too pessimistic at equal steps in panel (c); attainable coefficients resolve both mismatches. Two $p = 3$ controls also verify noncommuting weighted projections (Appendix I.7). Three larger-radius sign disagreements remain in one distinct-step instance (Appendix I.2).

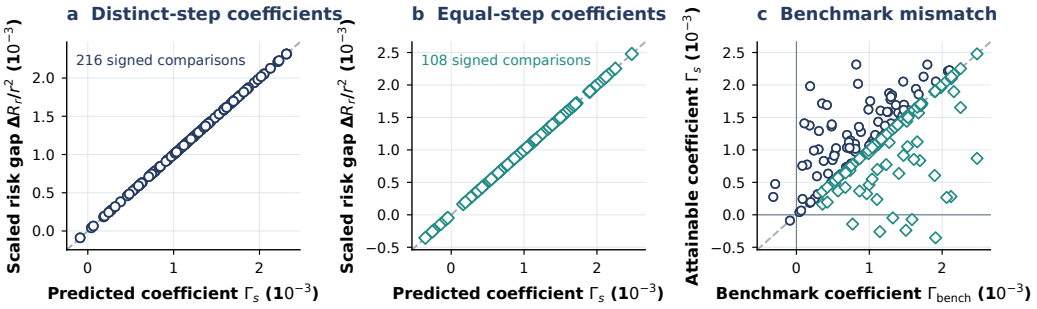


Figure 1: Local risk geometry at $|r| = 0.0015625$. (a) Distinct-step coefficients, 216 signed comparisons. (b) Equal-step coefficients, 108 comparisons. (c) Commuting (navy circles) and Jacobian (teal diamonds) benchmarks versus attainable coefficients. Here $\Gamma_{\text{bench}} = \Gamma_{\text{comm}}$ at distinct steps and Γ_{Jac} at equal steps. Dashed lines mark equality; zero axes in (c) separate sign agreement from reversal.

Finite-perturbation accuracy. We test finite-radius accuracy using three fixed controls down to $r = 0.1 \times 4^{-7}$ (Figure 2). Distinct and equal two-step schedules have $O(r)$ scaled error. At equal steps, a vanishing cubic split term gives this faster rate while the geometry remains singular. For three equal steps, error divided by $\sqrt{r}$ approaches $\sqrt{15}/3825$, attaining the general singular order. Dashed guides use the theoretical exponents and pass through the smallest-radius error.

Approaching the singular boundary. For two-step schedules $\boldsymbol{\alpha}_{0,\delta} = (a - \delta, a + \delta)$, Figure 3(a) plots the scaled risk gap $\Delta R_{r,\delta}/r^2$ against $\delta^2/|r|$. Each curve fixes the half-gap δ and varies r. As this ratio changes, negative-side optima move between the equal-step boundary and the interior; positive-side optima remain interior (Appendix I.6). Proposition 4.2 describes the corresponding transition along joint paths with $\delta_r \to 0$.

Preservation, reuse, and retuning. Since reuse belongs to the current family, population-optimal retuning attains an equal or lower risk. Figure 3(b) shows eight smallest-radius comparisons where reuse is worse than preservation but retuning is better.

Bike Sharing (Fanaee-T, 2013) compares preservation, reuse, and validation-selected retuning after a year shift. All predictors use the same target-domain input, and preservation applies the nominal filter to that input. Source and target validation select the nominal and retuned schedules; held-out target mean squared error (MSE) measures risk. With training-set centering, a shared target intercept, and reuse included among target candidates, retuning has lower test MSE than preservation in 8 months and higher MSE in 4 (panel (c)). Retuning and reuse coincide in 9 months. Appendix I.5 compares

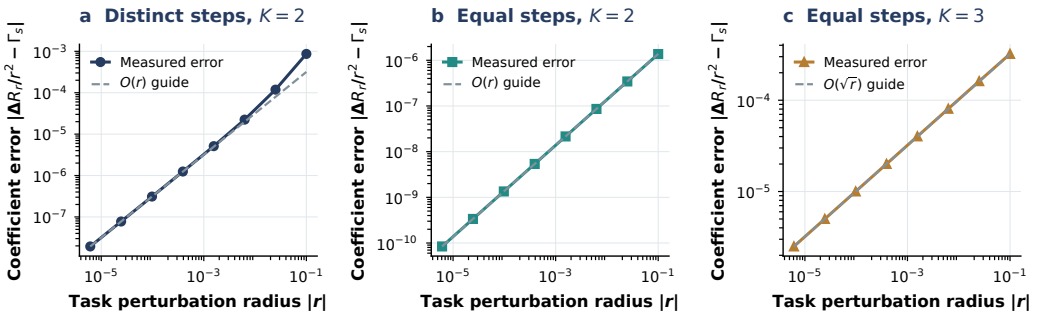


Figure 2: Finite-radius error in three fixed controls. (a) Distinct steps, $K = 2$. (b) Equal steps, $K = 2$. (c) Equal steps, $K = 3$. Solid curves show absolute coefficient errors; dashed guides show $O(r)$, $O(r)$, and $O(\sqrt{r})$, respectively. All points use exact population moments and verified global optima.

the original, candidate-only, and centered versions; centering substantially reduces the relative gaps. Confidence intervals (CIs) describe held-out-loss variability conditional on fitted predictors.

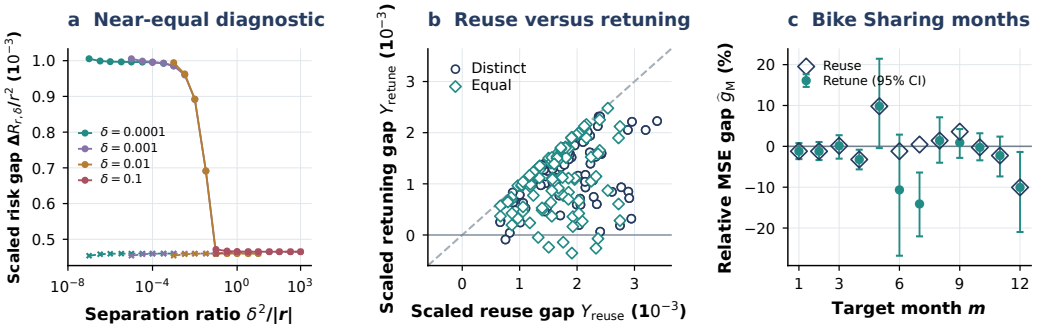


Figure 3: Near-equal geometry (a) and adaptation choices (b,c). (a) Scaled risk gaps for four positive half-gaps, each with its own target distribution; solid circles use $r < 0$, dashed crosses $r > 0$. (b) Scaled retuning gap versus scaled reuse gap, each relative to preservation and divided by r^2, in 324 smallest-radius comparisons. (c) Centered Bike Sharing sensitivity: risk gaps relative to preservation, comparing reuse (navy diamonds) and validation-selected retuning (teal circles). Nominal 95% paired block-bootstrap intervals use the centered version's original-sample preservation MSE to normalize absolute MSE gaps, with this denominator fixed across draws (Appendix I.5).

6 Discussion and Conclusion

We show that the leading risk difference between retuning and preservation is determined by how closely attainable filter changes follow the oracle drift. Two familiar approximations can give the wrong decision: unrestricted commuting filters may allow corrections that the schedule cannot realize, while the Jacobian can miss corrections at equal steps. Working with the full schedule family quantifies both errors and the statistical cost of the fixed update budget. In particular, even an exactly optimized schedule can have higher risk than preservation when the required correction is unavailable. For two steps, the near-equal result shows how the equal-step cone transitions continuously to regular geometry as $\delta_r^2/|r|$ increases. This transition explains why the nominal half-gap must be considered together with the size and direction of the task change.

Controlled experiments support the predicted coefficients, reversals, and rates. The real-data diagnostic compares preservation, schedule reuse, and validation-selected retuning, showing that these choices can lead to different test risks. Our theory concerns local task changes in fixed-dimensional, response-independent linear filters. The comparison uses nominal oracle centering and population-optimal retuning. A natural next step is to estimate the attainable directions and oracle drift from data. Inference for singular schedules and comparisons centered at a source-optimal finite-step filter are further questions.

ARTIFICIAL INTELLIGENCE USE STATEMENT

In this work, we used generative AI tools to aid manuscript writing, including organization and language editing, and to retrieve and identify related work. We also consulted these tools while refining mathematical claims, proofs, and the experimental design. The authors reviewed all AI-assisted material, and take responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

The appendix supplies complete proofs, statistical realizations, the prespecified 162-instance population study, all radius-level results and finite-radius controls. The anonymous supplement contains sampled matrices, all 2,268 signed comparisons, figure programs and global step-box minimization routines. Two-step minimization exhausts its coefficient-region boundary. Three-step minimization either attains an unrestricted quadratic lower bound or enumerates step-box faces and repeated-root critical points with rational polynomial root isolation. The separate 1,920-replication commuting-benchmark study includes its adverse cells; its inference claims remain separate from the population validation of budget-matched coefficients. The supplement also provides the fixed three-strategy diagnostics, preprocessing code and complete public-data records used in the Bike Sharing analysis.

ETHICS STATEMENT

The theorem-directed numerical studies use synthetic Gaussian constructions. The practical diagnostic uses the publicly available, aggregate Bike Sharing data from the University of California, Irvine (UCI) Machine Learning Repository and contains no personal identifiers. No human participants were recruited and no real-world intervention was performed.

REFERENCES

Alnur Ali, J. Zico Kolter, and Ryan J. Tibshirani. A continuous-time view of early stopping for least squares regression. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1370–1378, 2019.

Alnur Ali, Edgar Dobriban, and Ryan J. Tibshirani. The implicit regularization of stochastic gradient flow for least squares. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 233–244, 2020.

Frank Bauer, Sergei Pereverzev, and Lorenzo Rosasco. On regularization algorithms in learning theory. *Journal of Complexity*, 23(1):52–72, 2007. doi: 10.1016/j.jco.2006.07.001.

Pierre C. Bellec and Kai Tan. Uncertainty quantification for iterative algorithms in linear models with application to early stopping. *The Annals of Statistics*, 54(3):1315–1344, 2026. doi: 10.1214/25-aos2606.

Rajendra Bhatia and Peter Rosenthal. How and why to solve the operator equation $AX - XB = Y$. *Bulletin of the London Mathematical Society*, 29(1):1–21, 1997. doi: 10.1112/S0024609396001828.

Gilles Blanchard, Marc Hoffmann, and Markus Reiß. Early stopping for statistical inverse problems via truncated svd estimation. *Electronic Journal of Statistics*, 12(2):3204–3231, 2018. doi: 10.1214/18-ejs1482.

Peter Bühlmann and Bin Yu. Boosting with the L_2 loss: Regression and classification. *Journal of the American Statistical Association*, 98(462):324–339, 2003. doi: 10.1198/016214503000125.

Alain Celisse and Martin Wahl. Analyzing the discrepancy principle for kernelized spectral filter learning algorithms. *Journal of Machine Learning Research*, 22(76):1–59, 2021.

Herman Chernoff. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25(3):573–578, 1954. doi: 10.1214/aoms/1177728725.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001.

Aymeric Dieuleveut and Francis Bach. Nonparametric stochastic approximation with large step-sizes. *The Annals of Statistics*, 44(4):1363–1399, 2016. doi: 10.1214/15-aos1391.

Mathias Drton. Likelihood ratio tests and singularities. *The Annals of Statistics*, 37(2):979–1012, 2009. doi: 10.1214/07-AOS571.

Lutz Dümbgen. On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, 95(1):125–140, 1993. doi: 10.1007/bf01197342.

Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996. doi: 10.1007/978-94-009-1740-8.

Jun Fan, Zheng-Chu Guo, and Lei Shi. Spectral algorithms under covariate shift, 2025. URL `https://arxiv.org/abs/2504.12625v2`. Version 2, revised 22 July 2025.

Hadi Fanaee-T. Bike sharing. UCI Machine Learning Repository, 2013.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412, 2019. doi: 10.1093/restud/rdy049.

Rong Ge, Sham M. Kakade, Rahul Kidambi, and Praneeth Netrapalli. The step decay schedule: A near optimal, geometrically decaying learning rate procedure for least squares. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Charles J. Geyer. On the asymptotics of constrained m-estimation. *The Annals of Statistics*, 22(4): 1993–2010, 1994. doi: 10.1214/aos/1176325768.

Anne Greenbaum, Ren-Cang Li, and Michael L. Overton. First-order perturbation theory for eigenvalues and eigenvectors. *SIAM Review*, 62(2):463–482, 2020. doi: 10.1137/19M124784X.

Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, 2008. doi: 10.1137/1.9780898717778.

Nicholas J. Higham and Samuel D. Relton. Higher order fréchet derivatives of matrix functions and the level-2 condition number. *SIAM Journal on Matrix Analysis and Applications*, 35(3): 1019–1037, 2014. doi: 10.1137/130945259.

Varun Kanade, Patrick Rebeschini, and Tomas Vaškevičius. The statistical complexity of early-stopped mirror descent. *Information and Inference: A Journal of the IMA*, 12(4):3010–3041, 2023. doi: 10.1093/imaiai/iaad047.

Tosio Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer, second edition, 1995. doi: 10.1007/978-3-642-66282-9.

Kathlén Kohn, Thomas Merkh, Guido Montúfar, and Matthew Trager. Geometry of linear convolutional networks. *SIAM Journal on Applied Algebra and Geometry*, 6(3):368–406, 2022. doi: 10.1137/21M1441183.

Kathlén Kohn, Guido Montúfar, Vahid Shahverdi, and Matthew Trager. Function space and critical points of linear convolutional networks. *SIAM Journal on Applied Algebra and Geometry*, 8(2): 333–362, 2024. doi: 10.1137/23M1565504.

Vladimir Koltchinskii and Karim Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133, 2017. doi: 10.3150/15-bej730.

L. Landweber. An iteration formula for fredholm integral equations of the first kind. *American Journal of Mathematics*, 73(3):615–624, 1951. doi: 10.2307/2372313.

Olivier Ledoit and Michael Wolf. Shrinkage estimation of large covariance matrices: Keep it simple, statistician? *Journal of Multivariate Analysis*, 186:104796, 2021. doi: 10.1016/j.jmva.2021.104796.

Daniel LeJeune, Jiayu Liu, and Reinhard Heckel. Monotonic risk relationships under distribution shifts for regularized risk minimization. *Journal of Machine Learning Research*, 25(54):1–37, 2024.

Eitan Levin, Joe Kileel, and Nicolas Boumal. The effect of smooth parametrizations on nonconvex optimization landscapes. *Mathematical Programming*, 209:63–111, 2025. doi: 10.1007/s10107-024-02058-3.

Binghui Li, Zilin Wang, Fengling Chen, Shiyang Zhao, Ruiheng Zheng, and Lei Wu. Optimal learning rate schedules under functional scaling laws: Power decay and warmup–stable–decay (extended abstract). In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 4722–4723, 2026.

L. Lo Gerfo, Lorenzo Rosasco, Francesca Odone, Ernesto De Vito, and Alessandro Verri. Spectral algorithms for supervised learning. *Neural Computation*, 20(7):1873–1897, 2008. doi: 10.1162/neco.2008.05-07-517.

Roy Mathias. A chain rule for matrix functions and applications. *SIAM Journal on Matrix Analysis and Applications*, 17(3):610–620, 1996. doi: 10.1137/s0895479895283409.

Pratik Patil, Jin-Hong Du, and Ryan J. Tibshirani. Optimal ridge regularization for out-of-distribution prediction. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 39908–39954. PMLR, 2024a.

Pratik Patil, Yuchen Wu, and Ryan J. Tibshirani. Failures and successes of cross-validation for early-stopped gradient descent. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2260–2268, 2024b.

Loucas Pillaud-Vivien, Alessandro Rudi, and Francis Bach. Statistical optimality of stochastic gradient descent on hard learning problems through multiple passes. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Garvesh Raskutti, Martin J. Wainwright, and Bin Yu. Early stopping and non-parametric regression: An optimal data-dependent stopping rule. *Journal of Machine Learning Research*, 15(11): 335–366, 2014.

R. Tyrrell Rockafellar and Roger J. B. Wets. *Variational Analysis*. Springer, 1998. doi: 10.1007/978-3-642-02431-3.

Lorenzo Rosasco and Silvia Villa. Learning with incremental iterative regularization. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Rishi Sonthalia, Jackie Lok, and Elizaveta Rebrova. On regularization via early stopping for least squares regression, 2024. URL `https://arxiv.org/abs/2406.04425v2`. Version 2, revised 3 July 2026.

Anda Tang, Yiming Dong, Yutao Zeng, Xun Zhou, and Zhouchen Lin. Stepsize anything: A unified learning rate schedule for budgeted-iteration training. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-5022.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.

Maksim Velikanov, Maxim Panov, and Dmitry Yarotsky. Generalization error of spectral algorithms. In *The Twelfth International Conference on Learning Representations*, 2024.

Tobias Wegel, Gil Kur, and Patrick Rebeschini. Sharp risk bounds for early-stopping in gaussian linear regression. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 1792–1800, 2026.

Yuting Wei, Fanny Yang, and Martin J. Wainwright. Early stopping for kernel boosting algorithms: A general analysis with localized complexities. *IEEE Transactions on Information Theory*, 65(10): 6685–6703, 2019. doi: 10.1109/TIT.2019.2927563.

Kaiyue Wen, Zhiyuan Li, Jason Wang, David Hall, Percy Liang, and Tengyu Ma. Understanding warmup-stable-decay learning rates: A river valley loss landscape view. In *International Conference on Learning Representations*, 2025.

Jingfeng Wu, Peter L. Bartlett, Sham M. Kakade, Jason D. Lee, and Bin Yu. Risk comparisons in linear regression: Implicit regularization dominates explicit regularization (extended abstract). In *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 6849–6851. PMLR, 2026. Related full manuscript: arXiv:2509.17251.

Yuan Yao, Lorenzo Rosasco, and Andrea Caponnetto. On early stopping in gradient descent learning. *Constructive Approximation*, 26(2):289–315, 2007. doi: 10.1007/s00365-006-0663-2.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008.

Tong Zhang and Bin Yu. Boosting with early stopping: Convergence and consistency. *The Annals of Statistics*, 33(4):1538–1579, 2005. doi: 10.1214/009053605000000255.

APPENDIX OVERVIEW

The appendix contains the proofs, exact examples, extensions, and complete experimental results. The overview below links to each section and subsection.

Contents	Page
I. Core risk and attainable geometry	
A. Assumptions, Scope, and Proof Roadmap	15
B. Proof of the Attainable-Output Risk Result	16
B.1 Set limits and local risk	16
B.2 Regular parameter charts	17
B.3 Nested families	18
B.4 Signed commuting risk and the drift-ball bound	19
B.5 Reference error and quantitative distance comparison	21
C. Fixed-Budget Gradient Retuning	22
C.1 Filter derivatives	22
C.2 Global localization and attainment of the commuting optimum	23
C.3 Finite-radius accuracy at distinct steps	23
D. Singular Equal-Step Geometry	24
D.1 Equal-step isolation and square-root localization	24
D.2 Cone attainability and finite-radius accuracy	25
D.3 Why equal two-step schedules can converge faster	25
D.4 Two-step transition as the nominal half-gap vanishes	26
II. Exact examples and extensions	
E. Exact Reversals and Sharpness	27
E.1 Statistical realizability and decision reversals	27
E.2 An attained square-root error at three equal steps	28
E.3 Continuity of strict reversals	30
F. Separated Spectral Clusters	30
F.1 Separated groups and smooth projector transport	30
F.2 Attainable velocities and the comparison-class boundary	31
G. Finite-Update Derivative and Filter Motion	32
III. Estimation and numerical evidence	
H. Estimation of the Commuting-Benchmark Coefficient	32
H.1 An aligned commuting-filter estimator	32
H.2 Aligned estimands and weighted projections	33
H.3 Perturbation bounds and error envelopes	34
H.4 Consistency, central differences, and sign recovery	36
I. Complete Numerical Results and Reproducibility	37
I.1 Prespecified population protocol	37
I.2 Complete budget-matched sweep	38
I.3 Finite-perturbation rate controls and complete values	43
I.4 Commuting-benchmark estimation study	44
I.5 Three adaptation choices	46
I.6 Near-equal finite-radius diagnostic	50
I.7 Noncommuting weighted-metric controls	52
I.8 Bike Sharing sensitivity, reproducibility, and scope	53

A Assumptions, Scope, and Proof Roadmap

We collect the assumptions used in the results and explain where each enters. We start with the common statistical conditions, then give the conditions for distinct and equal schedules and for the finite-radius error bounds.

Standing statistical conditions. Theorem 4.1 works in fixed dimension with finite second moments and response-independent filters. The prediction metric and nominal observation moment satisfy $W, \Sigma_0 \succ 0$, with $\Sigma_r = \Sigma_0 + O(r)$ and $F_r^\star = F_0 + rD + O(r^2)$. The nominal reference is oracle-centered, $F_0 = F_0^\star$, and its error is $o(|r|)$, or $o_p(|r|)$ when independently estimated. The scaled output family must converge from each side to a nonempty closed set in the inner–outer sense of Definition 2. This set limit describes the local output changes available to the algorithmic family. It is not an additional assumption on the data distribution, and the limiting set need not be convex.

Distinct-step regime. Theorem 4.2 uses a symmetric C^2 path $S_r = S_0 + rE + O(r^2)$, $S_0 \succ 0$ with a simple spectrum, and a compact box $[\underline{\alpha}, \overline{\alpha}]^K$ with $0 < \underline{\alpha} < \overline{\alpha} < 1/\lambda_{\max}(S_0)$. The nominal steps are pairwise distinct and interior, and all K coordinates can be retuned independently. Distinctness gives the regular derivative rank $\min(K, p)$; interiority allows local parameter adjustments. Compactness and polynomial identification provide global localization when $K \le p$, while full row rank gives a local chart onto the commuting filters when $K \ge p$ (Appendix C).

Equal-step regime. Theorem 4.3 keeps the symmetric C^2 curvature path and positive compact step box, but takes $K \ge 2$ and an equal interior nominal schedule $a\mathbf{1}$, with $\underline{\alpha} < a < \overline{\alpha} < 1/\lambda_{\max}(S_0)$. At least two distinct nominal eigenvalues make A_K and H_K linearly independent. Independent retuning of the coordinates allows square-root splitting; permanently tied coordinates define a different family. The equal-step result describes retuning from a nominal constant-step schedule.

Finite-radius regularity. Corollary 4.1 assumes a fixed neighborhood where the expansions $S_r = S_0 + rE + O(r^2)$, $F_r^\star = F_0 + rD + O(r^2)$, and $\Sigma_r = \Sigma_0 + O(r)$ have uniformly bounded remainders. The required derivatives are bounded, the risk metric remains positive definite, and the nominal schedule remains interior. The relevant chart singular values, spectral and step gaps, and equal-step Gram eigenvalue stay bounded away from zero. These are local constants for a fixed instance. Exact preservation gives the stated rates; reference error adds $O(|r|\epsilon_r + \epsilon_r^2)$. Appendices C.3, D.2, and D.3 use these conditions.

Role and strength: nominal centering and reference accuracy. Nominal oracle centering is substantive: it removes pre-existing approximation error so that the comparison starts at order r^2. If $Q = F_0 - F_0^\star \neq 0$, Appendix B.5 shows that the risk change can contain $2r\langle Q, Z\rangle_0$, which can dominate the quadratic attainable-geometry effect. This bias differs from learning error in the preserved reference. The requirement $\|\widehat{F}_{0,r} - F_0\|_0 = o(|r|)$ makes that learning error smaller than the task perturbation. It isolates adaptation geometry but is not indispensable: for $\widehat{F}_{0,r} = F_0 + rH + o(r)$, Proposition B.2 gives $\mathrm{dist}_0^2(D, V_s) - \|D - H\|_0^2$. Only the preservation term changes; the attainable-family distance still uses D.

Role and strength: spectral and schedule conditions. A simple spectrum is the distinct-step specialization that exposes the $K = p$ threshold; it is not required by the general attainable-output risk principle. Appendix F treats separated spectral groups with within-group multiplicity. Pairwise distinct interior steps define the regular regime, while equal interior steps define the singular regime. Proposition 4.2 connects these regimes for $K = 2$ along triangular paths with $\delta_r \to 0$ and $\delta_r^2/|r| \to \rho$. The statistical expansions are required along the joint path and center the oracle at the moving reference $\bar{F}_r$. The resulting coefficient describes the leading risk on each such path; uniform finite-radius bounds would additionally require uniform remainder control.

Role and strength: oracle comparison and asymptotic scope. Population-optimal retuning isolates the statistical value of the schedule family's geometry from validation and schedule-selection error. The Bike Sharing diagnostic compares validation-selected schedules on finite data, complementing the population-oracle comparison. The state dimension p and update budget K are fixed.

Growing dimension, growing budget, and vanishing eigengaps require uniform control beyond the present results. The fixed marginal distribution of u makes the task-change constructions and their realizability easier to interpret. The general risk principle itself uses the induced moment and oracle expansions, rather than requiring that fixed-marginal convention. For the triangular comparison, each fixed-δ task path may have its own target marginal, as in Appendix I.6.

Proof roadmap. The main comparison separates a statistical calculation from an algorithmic reachability calculation. Appendix B first proves that, once the complete one-sided output limits are known, population risk reduces to weighted distance from the oracle drift. Appendices C and D then identify those limits for distinct and equal learning-rate schedules. The exact decision reversals in Appendix E use the resulting coefficients.

General risk result (Theorem 4.1). The exact oracle risk identity and positive-definite metric localize every near-optimal filter to an $O(|r|)$ output neighborhood of F_0. Inner and outer convergence of the scaled family then transfer that local quadratic risk to distance from D. The proof in Appendix B.1 follows

$$\text{risk localization} \longrightarrow \text{one-sided output limits} \longrightarrow \text{Theorem 4.1.}$$

Regular charts and multiple nominal parameter representations are treated in Appendix B.2. Appendix B.5 states exactly how reference error at scale r changes the coefficient.

Distinct steps (Theorem 4.2 and Proposition 4.1). The derivative polynomials span only $\min(K, p)$ independent spectral corrections. For a global step-box optimum, polynomial identification also rules out an unexamined nominal schedule branch when $K \leq p$; a full-row-rank spectral chart attains the commuting optimum when $K \geq p$. Weighted orthogonal projection then separates the commuting residual from the exact finite-budget restriction cost. Appendix C gives these arguments and the regular finite-radius bound.

Equal steps (Theorem 4.3). The key difficulty is that $O(r)$ output displacement permits $O(\sqrt{|r|})$ schedule splitting. Appendix D.1 isolates the equal nominal schedule, bounds every relevant split, and identifies its possible one-sided limits. Appendix D.2 then constructs the attainable cones and proves the finite-radius bound. Cone projection gives the risk coefficient, and Appendix D.2 controls the finite-radius error.

Appendix F extends the comparison to separated spectral blocks. Appendix G connects filter derivatives to finite-update dynamics. Appendix H develops estimation bounds for the commuting-benchmark coefficient. Appendix I describes the numerical protocols and reproducibility procedures.

B Proof of the Attainable-Output Risk Result

This appendix proves Theorem 4.1 in the finite-dimensional Hilbert space of matrix filters. We first reduce excess risk to a quadratic distance from the oracle drift, then use inner and outer convergence of the scaled output family to identify the limiting feasible set. The later subsections explain when regular parameter charts suffice, how multiple nominal representations enter a global comparison, and how reference error changes the leading coefficient.

B.1 Set limits and local risk

Let $\mathcal{H} = \mathbb{R}^{p\times p}$ with inner product $\langle X, Y\rangle_0 = \operatorname{tr}(X^\top W Y \Sigma_0)$, where $W, \Sigma_0 \succ 0$. For a nonempty family $\mathcal{F}_r$, define $\mathcal{Z}_r = (\mathcal{F}_r - F_0)/r$ for $r \neq 0$.

Step 1: identify the complete output limits. The one-sided limit conditions used in Theorem 4.1 mean the following. For every $Z \in V_s$ and every sequence $r_n \to 0$ with sign s, there exist $Z_n \in \mathcal{Z}_{r_n}$ with $Z_n \to Z$ (inner attainability). Conversely, if $r_n \to 0$ with sign s and $Z_n \in \mathcal{Z}_{r_n}$ converges, its limit belongs to V_s (outer completeness). In fixed dimension this also describes the limit points of every bounded sequence. The nonempty set V_s is closed. No convexity is required.

Step 2: reduce risk to oracle distance. Write $\Sigma_r = \mathbb{E}[z_r z_r^\top]$ and $C_r = \mathbb{E}[u z_r^\top]$. Completing the square, using $F_r^\star = C_r \Sigma_r^{-1}$, gives exactly

$$\mathcal{R}_r(F) - \mathcal{R}_r(F_r^\star) = \operatorname{tr}((F - F_r^\star)^\top W (F - F_r^\star)\Sigma_r). \tag{B.1}$$

The trace of the cross term vanishes by $C_r - F_r^\star \Sigma_r = 0$. The oracle need not be symmetric. For bounded Z and $F = F_0 + rZ$,

$$r^{-2}\{\mathcal{R}_r(F) - \mathcal{R}_r(F_r^\star)\} = \|Z - D\|_0^2 + o(1)$$

uniformly on bounded sets, because $\Sigma_r = \Sigma_0 + O(r)$ and $F_r^\star = F_0 + rD + O(r^2)$.

Step 3: localize the infimum and match both bounds. The localization argument below also proves Lemma 4.1. An attainable fixed $Z \in V_s$ gives a bounded upper comparison. Choose filters within $o(r^2)$ of the infimum. Positive definiteness of the metric implies $\|F - F_r^\star\|_F = O(|r|)$, so their scaled outputs are bounded. Every subsequence attaining the lower limit has a convergent further subsequence. Outer completeness puts its limit in V_s, giving the lower bound $\operatorname{dist}_0^2(D, V_s)$. Inner attainability, applied to a nearest point of the nonempty closed set V_s, gives the matching upper bound. Finally $\|\widehat{F}_{0,r} - F_0\|_0 = o(|r|)$ gives $\mathcal{R}_r(\widehat{F}_{0,r}) - \mathcal{R}_r(F_r^\star) = r^2\|D\|_0^2 + o(r^2)$. Subtraction proves Theorem 4.1. The same argument holds in probability for an independently trained random reference with $o_p(|r|)$ error.

The role of localization. If parameter choices have several nominal representations or can escape to infinity, differentiating at one representation need not describe $\mathcal{Z}_r$. The limit assumptions concern the image of the entire comparison family. They are a concrete finite-dimensional set-convergence condition, in the same variational tradition as tangent-cone risk and likelihood expansions (Drton, 2009); the quadratic projection argument is classical.

B.2 Regular parameter charts

The following corollary records the regular case used in the discussion of Theorem 4.1.

Corollary B.1 (Regular affine velocity space)**.** *Let $F(\theta, S)$ be C^2 near an interior nominal representation $F(\theta_0, S_0) = F_0$, and suppose $S_r = S_0 + rE + O(r^2)$. Here, $D_\theta F$ and $D_S F$ denote Fréchet derivatives with respect to the parameter and matrix arguments, respectively; they are distinct from the oracle drift D. Write $A = D_\theta F(\theta_0, S_0)$, $T = \operatorname{range} A$, and $G_{\mathrm{alg}} = D_S F(\theta_0, S_0)[E]$. If A has full column rank and Θ is a sufficiently small closed neighborhood of θ_0, then the complete output family over Θ has $V_+ = V_- = G_{\mathrm{alg}} + T$. Under the statistical-path and reference-accuracy conditions of Theorem 4.1, let P_T be the weighted projector onto T and $g_T = (\mathrm{Id} - P_T)G_{\mathrm{alg}}$. Its risk coefficient is*

$$\Gamma_{\mathrm{alg}}(D) = \|g_T\|_0^2 - 2\langle g_T, D\rangle_0 - \|P_T D\|_0^2. \tag{B.2}$$

For a compact global parameter set with finitely many interior regular nominal preimages, use the union of their affine velocity spaces and the minimum of their coefficients.

The projection is taken in the statistical metric, not by unweighted diagonal extraction. The local injectivity argument below establishes the required parameter localization; the final paragraph treats multiple nominal preimages.

For a C^2 map $F(\theta, S)$, set $A = D_\theta F(\theta_0, S_0)$ and $G_{\mathrm{alg}} = D_S F(\theta_0, S_0)[E]$. Suppose A is injective. Taylor expansion yields, for sufficiently small δ,

$$\|F(\theta_0 + \delta, S_0) - F_0\|_F \geq \sigma_{\min}(A)\|\delta\|_2 - C\|\delta\|_2^2 \geq c\|\delta\|_2.$$

Choose a small closed parameter ball centered at θ_0 on which this bound holds. For parameters in this ball with $F(\theta, S_r) - F_0 = O(r)$, uniform smoothness gives $F(\theta, S_0) - F_0 = O(r)$, hence $\delta = O(r)$. Consequently

$$\frac{F(\theta_0 + \delta, S_r) - F_0}{r} = G_{\mathrm{alg}} + A(\delta/r) + o(1).$$

This proves outer completeness for $V = G_{\mathrm{alg}} + \operatorname{range} A$. The paths $\delta = rv$ prove inner attainability, as θ_0 is interior.

Let $T = \operatorname{range} A$ and P_T be its weighted orthogonal projector. Writing $g_T = (\mathrm{Id} - P_T)G_{\mathrm{alg}}$ gives

$$\begin{aligned}\operatorname{dist}_0^2(D, G_{\mathrm{alg}} + T) - \|D\|_0^2 &= \|(\mathrm{Id} - P_T)(D - G_{\mathrm{alg}})\|_0^2 - \|D\|_0^2 \\ &= \|g_T\|_0^2 - 2\langle g_T, D\rangle_0 - \|P_T D\|_0^2.\end{aligned}$$

Multiple representations of the nominal filter. For the global version, let Θ be compact, F be C^2 on a neighborhood, and $\{\theta \in \Theta : F(\theta, S_0) = F_0\}$ consist of finitely many interior points $\theta_1, \dots, \theta_m$, each with injective derivative $A_j = D_\theta F(\theta_j, S_0)$. Let $G_j = D_S F(\theta_j, S_0)[E]$ be the curvature-induced velocity on branch j. Away from their small neighborhoods, compactness gives a positive lower bound on $\|F(\theta, S_0) - F_0\|_F$. Every near-optimal parameter therefore approaches one of these representations; the same argument gives

$$V = \bigcup_{j=1}^{m} \big(G_j + \operatorname{range} A_j\big).$$

The coefficient is the minimum of the branch coefficients. A single-branch formula for a global infimum needs this localization or another image argument.

B.3 Nested families

For the commuting class under a simple spectrum, let $S_0 = U \operatorname{diag}(\lambda_i) U^\top$ and $F_0 = U \operatorname{diag}(f_i) U^\top$. Its affine velocity class is $G_{\mathrm{comm}} + \mathcal{C}_0$, where

$$(U^\top G_{\mathrm{comm}} U)_{ij} = \frac{f_i - f_j}{\lambda_i - \lambda_j} (U^\top E U)_{ij} \quad (i \neq j), \qquad (U^\top G_{\mathrm{comm}} U)_{ii} = 0.$$

This follows either from differentiating the commutation identity or from smooth spectral projectors, as shown below. If $T \subseteq \mathcal{C}_0$ and $G_{\mathrm{alg}} - G_{\mathrm{comm}} \in \mathcal{C}_0$, the weighted projectors satisfy $P_T \Pi = \Pi P_T = P_T$. The two orthogonal components of the residual are $(\mathrm{Id} - \Pi)(D - G_{\mathrm{alg}})$ and $(\Pi - P_T)(D - G_{\mathrm{alg}})$. The first equals $(\mathrm{Id} - \Pi)(D - G_{\mathrm{comm}})$. Pythagoras proves

$$\Gamma_{\mathrm{alg}} - \Gamma_{\mathrm{comm}} = \|(\Pi - P_T)(D - G_{\mathrm{alg}})\|_0^2. \tag{B.3}$$

Equality holds precisely when $\Pi(D - G_{\mathrm{alg}}) \in T$. For any possibly nonlinear velocity set $V_s \subseteq G_{\mathrm{comm}} + \mathcal{C}_0$, distance monotonicity still gives $\Gamma_s \geq \Gamma_{\mathrm{comm}}$.

The commuting benchmark and its weighted projection. We use the commutator $[A, B] = AB - BA$. The task-dependent statistical inner product is $\langle X, Y\rangle_r = \operatorname{tr}(X^\top W Y \Sigma_r)$, with $\|X\|_r^2 = \langle X, X\rangle_r$ and $\operatorname{dist}_r(X, \mathcal{C}) = \inf_{Y \in \mathcal{C}} \|X - Y\|_r$. The commuting comparison is useful as an intermediate lower bound. The following calculation establishes its moving affine velocity space directly, including localization at the risk infimum. It does not impose a finite-step representation on arbitrary spectral coefficients.

Let $S_r = S_0 + rE + O(r^2)$ be symmetric, with $S_0 = U \operatorname{diag}(\lambda_i) U^\top$ having a fixed simple spectrum. Let the nominal optimal statistical filter be $F_0 = U \operatorname{diag}(f_i) U^\top$. Work in the nominal eigenbasis below. Define

$$\mathcal{C}_r = \{F = F^\top : F S_r = S_r F\}, \qquad \langle A, B\rangle_0 = \operatorname{tr}(A^\top W B \Sigma_0), \qquad \|A\|_0^2 = \langle A, A\rangle_0,$$

with fixed $W, \Sigma_0 \succ 0$. Let Π be the orthogonal projection, in this matrix inner product, onto $\mathcal{C}_0$. It is an operator on matrices, not a state-space projector. Set

$$\Omega_{ij} = \frac{E_{ij}}{\lambda_j - \lambda_i} \ (i \neq j), \quad \Omega_{ii} = 0, \qquad G_{\mathrm{comm}} = [\Omega, F_0], \quad g = (\mathrm{Id} - \Pi) G_{\mathrm{comm}}, \quad \kappa = \|g\|_0. \tag{B.4}$$

The matrix Ω is skew-symmetric and G_{comm} is symmetric, with $(G_{\mathrm{comm}})_{ij} = (f_i - f_j) E_{ij} / (\lambda_i - \lambda_j)$ and zero diagonal.

Lemma B.1 (Retuning tangent and local spectral obstruction)**.** *The possible derivatives at zero of differentiable paths $F(r) \in \mathcal{C}_r$ with $F(0) = F_0$ form exactly the affine space*

$$G_{\mathrm{comm}} + \mathcal{C}_0. \tag{B.5}$$

The same space is realized by scalar polynomials $p_r(S_r)$ of degree at most $p-1$, with coefficients fixed independently of the current response. If the risk metric uses $\Sigma_r = \Sigma_0 + O(r)$, then

$$\mathrm{dist}_r(F_0, \mathcal{C}_r)^2 = r^2\kappa^2 + O(|r|^3), \qquad \mathrm{dist}_r(F_0, \mathcal{C}_r) = |r|\kappa + O(r^2). \tag{B.6}$$

For the unweighted Frobenius metric,

$$\kappa^2 = \sum_{i \neq j} \frac{(f_i - f_j)^2 E_{ij}^2}{(\lambda_i - \lambda_j)^2}. \tag{B.7}$$

For arbitrary positive-definite weights, $\kappa = 0$ if and only if $[E, F_0] = 0$.

Proof. Simple-eigenvalue perturbation yields an eigenbasis $U_r = U(I + r\Omega) + O(r^2)$, where Ω is defined in (B.4) (Greenbaum et al., 2020). Every commuting symmetric matrix has the form $U_r \,\mathrm{diag}(a_i(r))U_r^\top$. Differentiating at $a_i(0) = f_i$ gives $G_{\mathrm{comm}} + U\,\mathrm{diag}(a_i'(0))U^\top$. Conversely, any chosen diagonal derivative gives such a path. Lagrange interpolation through the distinct current eigenvalues realizes its coefficients by a scalar polynomial; this statement imposes no positive-step or degree-K constraint.

For rigor at the infimum, use the smooth basis $B_i(r) = u_i(r)u_i(r)^\top$. Let $M(r)$ denote the current weighted Gram matrix, with entries $M_{ij}(r) = \mathrm{tr}(B_i(r)^\top W B_j(r)\Sigma_r)$. Since $M(r) = M(0) + O(r)$ and $M(0)$ is invertible, the weighted least-squares coefficients are well defined near zero. At zero they equal f_i, and the projection residual vanishes. Expanding the normal equations therefore gives the projection residual, defined as F_0 minus its current weighted projection, $-r(\mathrm{Id} - \Pi)G_{\mathrm{comm}} + O(r^2)$. The smooth basis expansion and $\Sigma_r = \Sigma_0 + O(r)$ then give (B.6). Equivalently, minimize $\|G_{\mathrm{comm}} + A\|_0$ over $A \in \mathcal{C}_0$. In the unweighted metric the diagonal and off-diagonal matrix spaces are orthogonal, so $\Pi G_{\mathrm{comm}} = 0$ and (B.7) follows. For general positive weights, $\kappa = 0$ means $G_{\mathrm{comm}} \in \mathcal{C}_0$. Since G_{comm} has zero diagonal in the simple nominal eigenbasis, this means $G_{\mathrm{comm}} = 0$, exactly $[E, F_0] = 0$. □

Computing the coefficient by weighted projection. For general W, Σ_0, diagonal and off-diagonal matrices need not be orthogonal. With $B_i = u_i u_i^\top$, let $M_{ij} = \langle B_i, B_j\rangle_0$ and $h_i = \langle B_i, G_{\mathrm{comm}}\rangle_0$. Then

$$\kappa^2 = \|G_{\mathrm{comm}}\|_0^2 - h^\top M^{-1} h. \tag{B.8}$$

When both weights are diagonal in the nominal eigenbasis, write $U^\top W U = \mathrm{diag}(w_1, \ldots, w_p)$ and $U^\top \Sigma_0 U = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$. Then this reduces to $\sum_{i \neq j} w_i \sigma_j (G_{\mathrm{comm}})_{ij}^2$. Changing W reweights target errors and changes both the obstruction and the norm of oracle drift in the comparison below.

Connection to the finite-update filter derivative. For a differentiable scalar step path $a(r)$ with $a(0) = a > 0$, let $J_r = I - a(r)S_r$ and $B = J_0' = -aE - a'(0)S_0$. Define $\mathcal{L}_K(B) = \sum_{j=0}^{K-1} J_0^{K-1-j} B J_0^j$. If $F_0 = I - J_0^K$, differentiating the filter gives

$$(\mathrm{Id} - \Pi)\{-\mathcal{L}_K(B)\} = g, \qquad \kappa = \|(\mathrm{Id} - \Pi)\mathcal{L}_K(B)\|_0 . \tag{B.9}$$

Indeed, every differentiable scalar filter path has derivative in (B.5); scalar step adjustment changes only its $\mathcal{C}_0$ component. The component g survives even allowing arbitrary independent scalar spectral coefficients. More generally this holds for the derivative of any differentiable nominal polynomial $p_0(S)$ with $p_0(S_0) = F_0$.

Thus the finite-update filter derivative has a component that commuting spectral adjustments cannot absorb. An eigenvalue-only perturbation can have $[E, F_0] = 0$ while changing the finite-step output. The condition $E = 0$ gives first-order curvature invariance; exact invariance requires constancy along the whole path. With changing step scale, $E = 0$ alone does not imply $B = 0$.

B.4 Signed commuting risk and the drift-ball bound

Condition on training information independent of the current evaluation pair (u, z_r). The current response z_r and target u remain random with finite second moments. All compared filters are fixed under this conditioning; they may not adapt their coefficients to the current z_r. Set

$$\mathcal{R}_r(F) = \mathbb{E}[(Fz_r - u)^\top W (Fz_r - u)], \quad \Sigma_r = \mathbb{E}[z_r z_r^\top] \succ 0, \quad C_r = \mathbb{E}[u z_r^\top].$$

The fixed-linear oracle is $F_r^\star = C_r \Sigma_r^{-1}$. Assume

$$\Sigma_r = \Sigma_0 + O(r), \quad F_r^\star = F_0 + rD + O(r^2), \quad F_0 \in \mathcal{C}_0, \quad d = \|D\|_0 .$$

The matrix D need not be symmetric. Completing the square gives $\mathcal{R}_r(F) - \mathcal{R}_r(F_r^\star) = \|F - F_r^\star\|_r^2$.

Proposition B.1 (Signed risk coefficient and sharp uniform preservation)**.** *Let $r_n \to 0$ be nonzero and let an independently learned reference satisfy $\left\|\widehat{F}_{0,n} - F_0\right\|_0 = o(|r_n|)$. Under the preceding assumptions,*

$$\inf_{F \in \mathcal{C}_{r_n}} \mathcal{R}_{r_n}(F) - \mathcal{R}_{r_n}(\widehat{F}_{0,n}) = r_n^2 \Gamma(D) + o(r_n^2), \tag{B.10}$$

where the exact signed coefficient is

$$\begin{aligned} \Gamma(D) &= \|(\mathrm{Id} - \Pi)(D - G_{\mathrm{comm}})\|_0^2 - \|D\|_0^2 \\ &= \kappa^2 - 2\langle g, D\rangle_0 - \|\Pi D\|_0^2 . \end{aligned} \tag{B.11}$$

Positive and negative Γ give, respectively, preservation benefit and failure for sufficiently small r_n. At $\Gamma = 0$, higher-order terms decide.

Over all local oracle directions with $\|D\|_0 \le d_{\max}$,

$$\inf_{\|D\|_0 \le d_{\max}} \Gamma(D) = (\kappa - d_{\max})_+^2 - d_{\max}^2. \tag{B.12}$$

For $\kappa > 0$, a strictly positive second-order coefficient uniformly over this ball is possible exactly when $\kappa > 2d_{\max}$. This is a local robustness statement over specified oracle directions, not a lower bound over all estimation algorithms. A common nonzero task radius over a family additionally requires uniform remainder bounds over that family.

Proof. The smooth weighted projection argument from Lemma B.1, now applied to $F_r^\star$, gives

$$\inf_{F \in \mathcal{C}_r} \|F - F_r^\star\|_r^2 = r^2 \operatorname{dist}_0(D, G_{\mathrm{comm}} + \mathcal{C}_0)^2 + O(|r|^3).$$

Equivalence of the nearby metrics and the learned-reference assumption give $\left\|\widehat{F}_{0,n} - F_{r_n}^\star\right\|_{r_n}^2 = r_n^2 \|D\|_0^2 + o(r_n^2)$. Subtracting proves the first line of (B.11). Decompose $D = \Pi D + (\mathrm{Id} - \Pi)D$ orthogonally and use $g \perp \mathcal{C}_0$ to obtain the second line.

For (B.12), the reverse triangle inequality yields $\|(\mathrm{Id} - \Pi)(D - G_{\mathrm{comm}})\|_0 \ge (\kappa - \|D\|_0)_+$, which gives the stated lower bound. If $0 < d_{\max} \le \kappa$, choose $D = (d_{\max}/\kappa)g$ to attain it. If $d_{\max} \ge \kappa$, choose $D = g + A$ with $A \in \mathcal{C}_0$ of norm $\sqrt{d_{\max}^2 - \kappa^2}$; its residual is zero and its total norm is $d_{\max}$. The class $\mathcal{C}_0$ is nonzero, so such an A exists. If $\kappa = 0$, choose any $D \in \mathcal{C}_0$ of norm $d_{\max}$. These also cover zero-radius limits. □

Statistical realizability and the limit of the sharpness claim. The optimizing directions in (B.12) are realizable in a broad, well-defined response-moment class with a task-invariant target distribution. Suppose F_0 is invertible. Let z_0 be zero-mean with covariance Σ_0, let $u = F_0 z_0 + \varepsilon$ with independent mean-zero finite-variance noise, and define

$$A_r = (F_0 + rD)^{-1} F_0, \qquad z_r = A_r z_0.$$

For small r, A_r is invertible, $\Sigma_r = A_r \Sigma_0 A_r^\top$, and $\mathbb{E}[u z_r^\top]\Sigma_r^{-1} = F_0 A_r^{-1} = F_0 + rD$. Thus the marginal distribution of u remains unchanged while the oracle has precisely direction D. This is not a claim that every direction is attainable by a particular data-generating model or by unbiased additive noise with fixed signal covariance. The controls in the commuting-benchmark estimation study in Appendix I.4 use the latter, more restrictive model; their exact signed coefficient, rather than an unattainable worst-case direction, determines their result.

Relation to a norm-only sufficient inequality. With $a_r = |r|\kappa + O(r^2)$, $\delta_r = |r|d + O(r^2)$ and reference error $o(|r|)$, the triangle-inequality risk bound gives $(\kappa - d)_+^2 - d^2$. When $\kappa > 2d$, this equals the positive coefficient $\kappa(\kappa - 2d)$. Equation (B.11) is stronger: it retains the signed alignment

$\langle g, D\rangle_0$ and the recoverable oracle movement ΠD. For example, $D = tg$ gives $\Gamma = \kappa^2(1-2t)$, whereas $D = -tg$ gives $\Gamma = \kappa^2(1+2t)$ for $t > 0$. Large drift can help or hurt preservation.

If reference mean-square estimation error is $O(n^{-1})$, Markov's inequality yields $o_p(|r_n|)$ when $nr_n^2 \to \infty$. Then (B.10) holds in probability, conditional on the model assumptions. It is not automatic at $r_n \asymp n^{-1/2}$ or for a fixed poorly learned reference.

B.5 Reference error and quantitative distance comparison

We next consider reference errors of the same order as the task perturbation.

Let $\widehat{F}_{0,r}$ denote the preserved filter, fixed independently of the current evaluation response. Unlike the retuning family, it need not be a member of $\mathcal{F}_r$. Its only role in the comparison is the second distance in the exact oracle identity.

Proposition B.2 (Scale-level reference correction)**.** *Under the statistical and set-limit assumptions of Theorem 4.1, replace its reference-accuracy condition by*

$$\widehat{F}_{0,r} = F_0 + rH + o(r)$$

in the nominal matrix norm. Define

$$\Gamma_s(D; H) := \operatorname{dist}_0^2(D, V_s) - \|D - H\|_0^2. \tag{B.13}$$

Then $\Delta R_r = r^2\Gamma_s(D; H) + o(r^2)$. The same result holds in probability when H is deterministic and the displayed reference expansion holds in probability. More generally, if $H_r = (\widehat{F}_{0,r} - F_0)/r = O_p(1)$, then

$$\frac{\Delta R_r}{r^2} = \operatorname{dist}_0^2(D, V_s) - \|D - H_r\|_0^2 + o_p(1).$$

Proof. The family infimum is independent of the choice of preserved reference. The first part of Theorem 4.1 therefore gives

$$r^{-2} \inf_{F \in \mathcal{F}_r} \{\mathcal{R}_r(F) - \mathcal{R}_r(F_r^\star)\} = \operatorname{dist}_0^2(D, V_s) + o(1).$$

For a deterministic reference expansion, $(\widehat{F}_{0,r} - F_r^\star)/r = H - D + o(1)$. Metric equivalence and $\Sigma_r = \Sigma_0 + O(r)$ imply

$$r^{-2}\{\mathcal{R}_r(\widehat{F}_{0,r}) - \mathcal{R}_r(F_r^\star)\} = \|D - H\|_0^2 + o(1).$$

Subtract the two displays. For $H_r = O_p(1)$, the expansion of the quadratic loss is uniform on each bounded set of H_r. Given any probability tolerance, tightness restricts H_r to such a set with arbitrarily high probability. The remainder on that set is $o(1)$, proving the last assertion. □

Expanding the correction gives

$$\Gamma_s(D; H) - \Gamma_s(D; 0) = 2\langle D, H\rangle_0 - \|H\|_0^2.$$

A reference displacement towards the oracle can increase the advantage of preservation. A displacement away from the oracle can reduce it. The sign of this correction is therefore not determined by the reference-error norm alone.

If $r_n \asymp n^{-1/2}$ and reference error is $O_p(n^{-1/2})$, the last expansion retains its random direction. Boundedness in probability does not imply convergence to a deterministic H. When the scaled reference error converges in distribution, the continuous mapping theorem gives the corresponding squared-distance limit for the risk coefficient (van der Vaart, 1998).

Nominal centering versus reference estimation. The condition $F_0 = F_0^\star$ differs from the accuracy of an estimated reference: $\widehat{F}_{0,r} - F_0$ is estimation error, whereas $Q = F_0 - F_0^\star$ is nominal model bias. For bounded Z, the oracle identity gives

$$\mathcal{R}_r(F_0 + rZ) - \mathcal{R}_r(F_0) = 2r\langle F_0 - F_r^\star, Z\rangle_r + r^2\|Z\|_r^2,$$

where the task-dependent inner product is defined in Appendix B.3. If $F_r^\star = F_0^\star + O(r)$ and $\Sigma_r = \Sigma_0 + O(r)$, then

$$\mathcal{R}_r(F_0 + rZ) - \mathcal{R}_r(F_0) = 2r\langle Q, Z\rangle_0 + O(r^2).$$

A nonzero Q can therefore introduce a first-order term, although it vanishes when $\langle Q, Z\rangle_0 = 0$. This pointwise calculation does not extend the global retuning theorem beyond nominal oracle centering: localization of the global infimum would need a separate argument.

Distance stability. The following elementary estimate is the common step in the finite-radius proofs. It also makes explicit why a bound on the filter approximation transfers to a bound on minimized risk.

Lemma B.2 (Distance stability on relevant bounded sets)**.** *Let $\eta_r \to 0$ and V be a nonempty closed set. Suppose the current scaled sets $\mathcal{Z}_r$ have uniformly bounded near-minimizers for distance to D. Suppose each such point has distance at most $C\eta_r$ from V, and one nearest point of V has an approximation in $\mathcal{Z}_r$ within $C\eta_r$. Then*

$$\inf_{Z\in\mathcal{Z}_r} \|Z - D\|_0^2 = \operatorname{dist}_0^2(D, V) + O(\eta_r).$$

If $F_r^\star = F_0 + rD + O(r^2)$ and $\Sigma_r = \Sigma_0 + O(r)$, the scaled excess-risk infimum differs from this distance by $O(|r|)$.

Proof. On a ball of radius M, the squared distance to D is Lipschitz:

$$\left|\|X - D\|_0^2 - \|Y - D\|_0^2\right| \le \|X - Y\|_0(\|X - D\|_0 + \|Y - D\|_0).$$

The two approximation assumptions give the upper and lower bounds by this inequality. Near-minimizers can be chosen with objective slack $o(\eta_r)$, so an attained infimum is unnecessary. For the risk assertion, write $(F - F_r^\star)/r = Z - D + O(r)$. Expand the quadratic norm and use the uniform boundedness of Z and the $O(r)$ metric perturbation. □

For an exact reference, a distance error $O(\eta_r)$ gives a risk-difference error $O(r^2\eta_r + |r|^3)$. For a reference displacement Q_r, the exact change in preserved excess risk is

$$2\langle Q_r, F_0 - F_r^\star\rangle_r + \|Q_r\|_r^2 = O(|r|\|Q_r\|_0 + \|Q_r\|_0^2).$$

This reference term is kept separately in Corollary 4.1. In particular, an unquantified $o(|r|)$ reference error cannot be absorbed into an $O(|r|^3)$ bound.

C Fixed-Budget Gradient Retuning

This appendix proves Theorem 4.2 and the restriction-cost identity in Proposition 4.1. The whole step box is the comparison family: we identify its derivative polynomials, establish their spectral rank, and rule out better nominal branches elsewhere in the box. The final subsection proves the regular finite-radius rate in Corollary 4.1.

C.1 Filter derivatives

In this subsection, derivatives are evaluated at the nominal schedule $\boldsymbol{\alpha}_0$, whose coordinates are $\alpha_{0,j}$. We write $\mathcal{P}_m$ for real univariate polynomials of degree at most m. For $F_{\boldsymbol{\alpha}}(S) = I - \prod_{k=1}^K (I - \alpha_k S)$, the product rule gives

$$A_j = S_0 \prod_{\ell\neq j}(I - \alpha_{0,\ell}S_0),$$

$$G_K = \sum_{j=1}^K \alpha_{0,j} \left\{\prod_{\ell<j}(I - \alpha_{0,\ell}S_0)\right\} E \left\{\prod_{\ell>j}(I - \alpha_{0,\ell}S_0)\right\}.$$

All factors involving S_0 commute with one another, but generally not with E.

Rank of the adjustment space. For distinct nonzero $\alpha_{0,j}$, the polynomials $x\prod_{\ell\neq j}(1-\alpha_{0,\ell}x)$ form a basis of $x\mathcal{P}_{K-1}$. To prove independence, evaluate a zero linear combination at $x = 1/\alpha_{0,i}$; only its ith coefficient survives. Evaluation at p distinct positive eigenvalues has rank $\min(K, p)$, since the associated Vandermonde matrix has that rank. It follows that

$$T_K = \{S_0 q(S_0) : \deg q \le K-1\}, \qquad \dim T_K = \min(K, p).$$

C.2 Global localization and attainment of the commuting optimum

Let the parameter set be a compact box with every nominal step interior and $0 < \underline{\alpha} < \overline{\alpha} < 1/\lambda_{\max}(S_0)$. For $K \le p$, two degree-at-most-K filters agreeing at the p eigenvalues also agree at zero. Their difference has at least $p+1 > K$ distinct roots and is identically zero. Thus the nominal fiber consists exactly of permutations of the nominal steps. These representations have the same G_K and T_K, and all are regular; the global version of the chart result applies.

Attainment when $K \ge p$. For $K \ge p$, the map from the steps to the p current spectral coefficients has full row rank at the nominal vector. Select p coordinates with invertible Jacobian. The implicit function theorem, with r as an additional variable and the other steps fixed, realizes every commuting filter sufficiently close to F_0. The weighted Gram solve and the moment expansions give a commuting optimum of the form $F_0 + O(r)$. It is therefore in the K-step image for all sufficiently small r. Since every K-step filter commutes with S_r, the two optimal population risks then coincide. The theorem states the resulting coefficient equality.

For completeness, consider any convergent scaled-output sequence $Z_r = (F_r - F_0)/r \to Z$ with $F_r \in \mathcal{F}_r$. Expanding $[F_r, S_r] = 0$ gives $[Z, S_0] + [F_0, E] = 0$. Since Z is symmetric, it follows that $Z \in G_{\mathrm{comm}} + \mathcal{C}_0 = G_K + \mathcal{C}_0$. Conversely, the same full-row-rank chart realizes every smooth commuting path through F_0, and hence every velocity in this affine space. Thus both one-sided attainable velocity sets equal $G_K + \mathcal{C}_0$.

The exact hierarchy at a common reference. For $K < p$, the hierarchy identity gives the exact nonnegative additional term. This compares K update applications after a schedule is chosen; it does not charge the search needed to choose that schedule.

The polynomial identification argument uses values at zero as well as the positive eigenvalues. Every gradient filter vanishes at zero, so the difference of two degree-at-most-K filters has the additional known root zero. This is why p positive spectral nodes identify the polynomial when $K \le p$. The factorization into K linear residual factors then determines the multiset of positive steps, including multiplicities.

The rank result is stated for a fixed nominal filter. For $K < p$, its unavailable adjustment component is the part of $\Pi(D - G_K)$ outside T_K. Increasing the number of available updates can enlarge this space only after one specifies how the same nominal filter is represented. Comparing unrelated early-stopping depths usually changes both the oracle reference and the derivative G_K. No monotonicity statement across such different comparisons follows from rank alone.

C.3 Finite-radius accuracy at distinct steps

We now prove the regular branch of Corollary 4.1. All $O(\cdot)$ constants are uniform on a sufficiently small neighborhood of the fixed instance. They may depend on the minimum nominal step separation, the spectral separation, and the smallest singular value of a regular chart.

For $K \le p$, compactness and nominal polynomial identification reduce every near-optimal schedule to a neighborhood of a permutation of $\boldsymbol{\alpha}_0$. The derivative at each such point is injective. The regular-chart bound therefore gives

$$\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|_2 \le C|r|$$

after a permutation. Write $\delta = \boldsymbol{\alpha} - \boldsymbol{\alpha}_0$, with coordinates $\delta_j = \alpha_j - \alpha_{0,j}$, in this aligned ordering. Taylor expansion with bounded second derivatives yields

$$F_{\boldsymbol{\alpha}}(S_r) = F_0 + rG_K + \sum_j \delta_j A_j + O(r^2 + |r|\|\delta\|_2 + \|\delta\|_2^2).$$

Consequently every relevant scaled output is within $C|r|$ of G_K+T_K. Conversely, let $G_K+\sum_j v_j A_j$ be the weighted projection of D onto this affine space. The schedule $\boldsymbol{\alpha}_0 + rv$ remains in the box and attains this velocity with $O(|r|)$ scaled error. Lemma B.2 gives

$$\inf_{F\in\mathcal{F}_r}\{\mathcal{R}_r(F)-\mathcal{R}_r(F_r^\star)\} = r^2\,\mathrm{dist}_0^2(D, G_K+T_K)+O(|r|^3).$$

Subtracting the exact-reference excess risk proves the result.

For $K \geq p$, choose a full-row-rank spectral coefficient chart as above. The current commuting optimum is $O(|r|)$ from F_0, by the smooth projector basis and the local oracle expansion. The chart attains it exactly for small r. The commuting least-squares comparison has remainder $O(|r|^3)$, so the same bound holds for the step-box infimum. This argument does not require all K step parameters to be individually identifiable.

The moment assumptions needed here are $F_r^\star = F_0 + rD + O(r^2)$ and $\Sigma_r = \Sigma_0 + O(r)$, with bounded constants. A C^2 statistical path supplies the first remainder, and a bounded first derivative supplies the second. A uniformly C^3 path, as used in the numerical constructions, is sufficient but stronger than necessary. The regular-chart constants can deteriorate as nominal steps merge.

D Singular Equal-Step Geometry

This appendix proves Theorem 4.3 for the full independently retunable step box; the steps are not constrained to remain equal after perturbation. An $O(r)$ output displacement need not force an $O(r)$ parameter displacement at the equal schedule. We first isolate that schedule globally, then establish $O(\sqrt{|r|})$ localization, identify both one-sided attainable cones, and transfer them to risk and finite-radius accuracy.

D.1 Equal-step isolation and square-root localization

Fix $K \geq 2$ and a common step a strictly inside a box $[\underline{\alpha}, \overline{\alpha}]^K$ with $0 < \underline{\alpha} < a < \overline{\alpha} < 1/\lambda_{\max}(S_0)$. Assume S_0 has at least two distinct positive eigenvalues; within-eigenvalue multiplicity is allowed. Write $J = I - aS_0$, $A_K = S_0 J^{K-1}$, and $H_K = S_0^2 J^{K-2}$. They are linearly independent: the ratio of their spectral values is $\lambda/(1-a\lambda)$, strictly increasing on the relevant interval.

Global nominal isolation. Suppose $F_{\boldsymbol{\alpha}}(S_0) = F_{a\mathbf{1}}(S_0)$. Choose two distinct eigenvalues $0 < \lambda_1 < \lambda_2$. Set $x_j = \log(1-\alpha_j\lambda_1)$ and $c = \lambda_2/\lambda_1 > 1$. The two equalities of residual products imply

$$\frac{1}{K}\sum_j x_j = \log(1-a\lambda_1), \qquad \frac{1}{K}\sum_j \phi(x_j) = \phi\left(\frac{1}{K}\sum_j x_j\right), \quad \phi(x) = \log(1-c+ce^x).$$

The step box ensures that ϕ is defined throughout the convex hull of the x_j. Moreover $\phi''(x) = ce^x(1-c)/(1-c+ce^x)^2 < 0$. Equality in strict Jensen implies $x_1 = \cdots = x_K$, hence every $\alpha_j = a$. Thus there is a unique nominal parameter vector. Compactness makes every globally near-optimal vector converge to it.

Square-root localization.

Step 2: bound every near-optimal split. Together with nominal isolation above, the following bound proves Lemma 4.2. Put $\delta_j = \alpha_j - a$ and let $e_j(\delta)$ denote the elementary symmetric polynomials. The exact product expansion gives

$$\begin{aligned} F_{a\mathbf{1}+\delta}(S_0) - F_0 &= e_1 A_K - e_2 H_K + O(\|\delta\|_2^3) \\ &= e_1 A_K + \tfrac{1}{2}(\|\delta\|_2^2 - e_1^2) H_K + O(\|\delta\|_2^3). \end{aligned} \tag{D.1}$$

Linear independence gives a constant $c > 0$ with $\|xA_K + yH_K\|_F \geq c(|x|+|y|)$. Apply this with $x = e_1, y = \|\delta\|^2/2$ and move the e_1^2 and cubic terms to the right. For sufficiently small δ, e_1^2 is an arbitrarily small multiple of $|e_1|$ and $\|\delta\|^3$ is an arbitrarily small multiple of $\|\delta\|^2$. Absorbing these terms gives

$$|e_1| + \|\delta\|_2^2 \leq C\|F_{a\mathbf{1}+\delta}(S_0) - F_0\|_F.$$

If $F_{\boldsymbol{\alpha}}(S_r) - F_0 = O(|r|)$, smoothness on the compact step box gives the same bound $O(|r|)$ at S_0. Therefore $e_1 = O(|r|)$ and $\|\delta\| = O(\sqrt{|r|})$. The joint expansion at $(a\mathbf{1}, S_0)$ consequently yields

$$\frac{F_{a\mathbf{1}+\delta}(S_r) - F_0}{r} = G_K + \frac{e_1}{r} A_K + \frac{\|\delta\|^2}{2r} H_K + o(1).$$

The omitted mixed term is $O(|r|\|\delta\|)/r = o(1)$, and the cubic term is $O(|r|^{1/2})$. Every bounded scaled output thus lies asymptotically in $G_K + \operatorname{span}\{A_K\} + s\mathbb{R}_+ H_K$ when $r \to 0$ with sign s.

D.2 Cone attainability and finite-radius accuracy

Step 3: construct every proposed velocity. Conversely, for any $v \in \mathbb{R}$, $w \geq 0$, choose a fixed vector $\zeta \perp \mathbf{1}$ with $\|\zeta\|_2 = 1$ and set

$$\delta(r) = \frac{rv}{K}\mathbf{1} + \sqrt{2|r|w}\,\zeta.$$

The parameters stay in the box for small r. Their scaled outputs tend to $G_K + vA_K + swH_K$. This proves inner attainability, outer completeness, and the exact global one-sided velocity sets in Theorem 4.3.

Step 4: project the oracle drift to obtain cone risk. Let P be the weighted projector onto $\operatorname{span}\{A_K\}$ and $b = (\mathrm{Id} - P)H_K \neq 0$. First minimize over the coefficient of A_K; next minimize $\|(\mathrm{Id} - P)(D - G_K) - swb\|_0^2$ over $w \geq 0$. The reduction relative to the Jacobian-approximation distance is $\{\max(s\langle b, D - G_K\rangle_0, 0)\}^2/\|b\|_0^2$. This proves both one-sided coefficients.

Jacobian failure. If one forces all retuned steps to remain equal, or restricts each step change to $O(r)$, the extra ray disappears. Those are different comparison families from independently retuning the K steps.

Finite-radius bound. The preceding localization is quantitative on bounded-risk sets: $|e_1| + \|\delta\|_2^2 \leq C|r|$. For a bounded scaled output, the joint expansion therefore gives

$$\frac{F_{a\mathbf{1}+\delta}(S_r) - F_0}{r} = G_K + \frac{e_1}{r} A_K + \frac{\|\delta\|_2^2}{2r} H_K + \mathcal{E}_r, \tag{D.2}$$

$$\|\mathcal{E}_r\|_0 \leq C\sqrt{|r|}. \tag{D.3}$$

The discarded e_1^2/r term is $O(|r|)$. The cubic parameter term is $O(|r|^{3/2})$ before division by r. A direct joint Taylor bound on curvature-parameter interactions is also at most $O(|r|^{3/2})$. These estimates are sufficient even without exploiting the symmetry of the schedule polynomial.

Linear independence of A_K and H_K bounds e_1/r and $\|\delta\|_2^2/(2|r|)$ for relevant outputs. Thus the leading part of (D.2) belongs to the correct one-sided cone, at uniformly bounded coordinates. For the opposite approximation, choose one nearest velocity $G_K + vA_K + swH_K$, $w \geq 0$. The explicit path

$$\delta(r) = \frac{rv}{K}\mathbf{1} + \sqrt{2|r|w}\,\zeta, \qquad \mathbf{1}^\top \zeta = 0, \quad \|\zeta\|_2 = 1,$$

has that leading velocity and the same $O(\sqrt{|r|})$ scaled remainder. If the nearest cone point lies on its edge, $w = 0$ is allowed. There is no need for the projection to lie in the relative interior.

Lemma B.2, with $\eta_r = \sqrt{|r|}$, now proves

$$\inf_{F \in \mathcal{F}_r} \{\mathcal{R}_r(F) - \mathcal{R}_r(F_r^\star)\} = r^2 \operatorname{dist}_0^2(D, V_s) + O(|r|^{5/2}).$$

The exact-reference risk has remainder $O(|r|^3)$, which is smaller. Subtraction proves the general singular bound. For a learned reference, append the separate error term from Appendix B.5.

D.3 Why equal two-step schedules can converge faster

For two steps the filter is exactly quadratic in the steps, so the cubic elementary symmetric term vanishes. At S_r, put

$$F_{a,a}(S_r) = I - (I - aS_r)^2, \qquad A_r = S_r(I - aS_r), \qquad H_r = S_r^2.$$

Then the identity

$$F_{a\mathbf{1}+\delta}(S_r) = F_{a,a}(S_r) + e_1 A_r + \tfrac{1}{2}(\|\delta\|_2^2 - e_1^2) H_r$$

holds without a parameter remainder. Since $F_{a,a}(S_r) = F_0 + rG_2 + O(r^2)$, $A_r = A_0 + O(r)$ and $H_r = H_0 + O(r)$, square-root localization gives

$$F_{a\mathbf{1}+\delta}(S_r) = F_0 + rG_2 + e_1 A_0 + \tfrac{1}{2}\|\delta\|_2^2 H_0 + O(r^2).$$

Crucially, the mixed terms multiply $e_1 = O(r)$ or $\|\delta\|^2 = O(|r|)$, not an uncancelled individual step displacement. Both the inner and outer velocity approximations therefore have error $O(|r|)$. Lemma B.2 yields

$$\Delta R_r = r^2 \Gamma_{2,s} + O(|r|^3)$$

for the exact reference on either side.

This improves the general bound in a genuine singular case. The improved rate does not remove the additional cone direction or its risk reversal. The faster risk approximation therefore coexists with singular output geometry. For $K = 3$, the cubic symmetric term can contribute a nonzero order-$|r|^{3/2}$ filter movement; Appendix E.2 gives an explicit example where it affects the minimized risk.

D.4 Two-step transition as the nominal half-gap vanishes

We prove Proposition 4.2 over the full step box, including both task signs and $\rho = \infty$. All statistical quantities in this subsection are evaluated on (r, δ_r); the preserved reference is $\bar{F}_r$, which varies with r.

Exact coefficient coordinates. Order the two retuned steps as $(a + h - v, a + h + v)$ with $v \geq 0$. Set $A_r = S_r(I - aS_r)$ and $H_r = S_r^2$. At $r = 0$, these give $A = A_2$ and $H = H_2$; we use $G = G_2$ at (a, a) as in Proposition 4.2. The exact identity is

$$F_{(a+h-v,a+h+v)}(S_r) - F_{(a-\delta_r,a+\delta_r)}(S_r) = 2hA_r + (v^2 - \delta_r^2 - h^2)H_r.$$

Two distinct positive eigenvalues make A_r, H_r linearly independent with a uniformly nonsingular Gram matrix for small r. Also,

$$F_{(a-\delta_r,a+\delta_r)}(S_r) - \bar{F}_r = rG + O(r^2) + O(|r|\delta_r^2) = rG + o(r).$$

Consequently every filter within $O(|r|)$ of $\bar{F}_r$, wherever its schedule lies in the box, satisfies $h = O(|r|)$ and $v^2 - \delta_r^2 - h^2 = O(|r|)$. This establishes global localization directly from output coefficients.

Outer and inner limits. Write $x_r = 2h/r$ and $q_r = (v^2 - \delta_r^2 - h^2)/r$. Their boundedness and the exact identity give scaled output $G + x_r A + q_r H + o(1)$. Because $v^2 \geq 0$,

$$sq_r \geq -\frac{\delta_r^2}{|r|} - \frac{h^2}{|r|}.$$

Any convergent output sequence therefore has its limit in $V_{s,\rho}$; when $\rho = \infty$, the same expansion still restricts the limit to $G + \operatorname{span}\{A, H\}$. Conversely, for any x, q with $sq \geq -\rho$, take

$$h_r = rx/2, \qquad v_r = \sqrt{\max\{0, \delta_r^2 + rq + h_r^2\}}.$$

For finite ρ, replacing a negative value of $\delta_r^2 + rq + h_r^2$ by zero changes the scaled H-coefficient $(v_r^2 - \delta_r^2 - h_r^2)/r$ by only $o(1)$. This also covers the boundary case $sq = -\rho$. For $\rho = \infty$ and any fixed q, the square-root argument is eventually positive. In both cases $h_r, v_r \to 0$, so these schedules lie inside the fixed box and attain $G + xA + qH$. The two inclusions establish the complete output-set limit for the full step box.

Risk and the explicit penalty. Reusing the nominal steps gives an $O(|r|)$-close candidate, hence $O(r^2)$ oracle excess risk along the triangular path. Positive definiteness localizes every population minimizer to the same output scale. The exact quadratic risk identity then converges uniformly on bounded scaled outputs to $\|Z - D\|_0^2$. The outer and inner limits give $\Delta R_{r,\delta_r}/r^2 \to \operatorname{dist}_0^2(D, V_{s,\rho}) - \|D\|_0^2$. To evaluate this distance, first minimize in the free A direction. The remaining quadratic in

q has unconstrained minimizer $\xi = t/\|b\|_0^2$ and curvature $\|b\|_0^2$. Projection of ξ onto $sq \geq -\rho$ adds $\|b\|_0^2(-s\xi - \rho)_+^2$ to the unrestricted two-direction coefficient $\Gamma_{\mathrm{Jac}} - t^2/\|b\|_0^2$. This proves the stated formula, its equal-step endpoint, and its saturation threshold. For $p > 2$, this two-direction value need not equal unrestricted commuting retuning.

A transition in the existing reversal example. Keep S_0, E, D, a from Appendix E.1, and replace its nominal reference by $\bar{F}_r = F_{(1-\delta_r, 1+\delta_r)}(S_0)$. The independent Gaussian construction in Appendix E remains valid with signal covariance $\bar{F}_r$ and noise covariance $I - \bar{F}_r - r\bar{F}_r^{-1}D$ for small r. For each fixed half-gap its target marginal is unchanged along the task path, and its oracle equals $\bar{F}_r + rD + O(r^2)$ uniformly near zero. Here $\|b\|_0^2 = 1/100$ and $\xi = 5/4$, so

$$\Gamma_{-,\rho} = -\frac{1}{1600} + \frac{1}{100}\left(\frac{5}{4} - \rho\right)_+^2.$$

The coefficient is positive for $0 \leq \rho < 1$, zero at $\rho = 1$, and negative for $\rho > 1$. At $\rho = 0$ this is the equal-step preservation advantage; beyond $\rho = 1$, near-equal retuning is beneficial to leading order. At the zero coefficient, higher-order terms determine the finite-radius decision.

E Exact Reversals and Sharpness

In diagonal examples we identify a diagonal matrix with its diagonal vector; unindexed inner products and norms are Euclidean. This appendix constructs both strict preserve-versus-retune sign reversals with a fixed target distribution before and after perturbation. We first verify that the oracle drifts arise from positive covariance models, then calculate the two coefficients and certify global low-dimensional optima. The final construction attains the general three-step singular remainder rate.

E.1 Statistical realizability and decision reversals

Both examples use diagonal $S_r = S_0 + rE$, so every diagonal matrix is an admissible commuting filter and $G_{\mathrm{comm}} = 0$. Take $W = I$, $u \sim N(0, F_0)$, and independent $\xi_r \sim N(0, I - F_0 - rF_0^{-1}D)$ with $z_r = u + \xi_r$. Here, F_0, D are diagonal and $0 \prec F_0 \prec I$; noise remains positive definite on a neighborhood. The target distribution stays fixed, $\Sigma_r = I - rF_0^{-1}D$, and

$$F_r^\star = F_0(I - rF_0^{-1}D)^{-1} = F_0 + rD + O(r^2).$$

Both reversals admit independent signal-plus-noise realizations with a task-invariant target distribution.

Commuting-benchmark reversal. Take $S_0 = \mathrm{diag}(1/4, 1/2, 3/4)$ and $\alpha_0 = (1, 1/2)$. The nominal filter is

$$f(x) = 3x/2 - x^2/2, \qquad F_0 = \mathrm{diag}(11/32, 5/8, 27/32).$$

Let $v = (3, -3, 1)/10$, $E = \mathrm{diag}(6/25, -3/10, 2/15)$ and $D = \mathrm{diag}(v)/4$. Since $f'(\lambda_i)E_{ii} = v_i$ and v is orthogonal to both (λ_i) and (λ_i^2), $G_K = \mathrm{diag}(v) \perp T_K$ and $\|v\|^2 = 19/100$. Therefore

$$\Gamma_{\mathrm{comm}} = -\|D\|^2 = -19/1600, \qquad \Gamma_K = \|G_K - D\|^2 - \|D\|^2 = 19/200.$$

The additional restriction term is $\|(3/4)G_K\|^2 = 171/1600$. The coefficient gap persists for both signs of r.

Jacobian-approximation reversal. For $S_0 = \mathrm{diag}(1/4, 1/2)$, $K = 2$, and $a = 1$, $F_0 = \mathrm{diag}(7/16, 3/4)$, $A_K = \mathrm{diag}(3/16, 1/4)$ and $H_K = \mathrm{diag}(1/16, 1/4)$. Their projection coefficient is $19/25$, giving $b = H_K - (19/25)A_K = \mathrm{diag}(-2/25, 3/50)$ and $\|b\|^2 = 1/100$. Set $E = \mathrm{diag}(4/75, -3/50)$ and $D = b/4$; then $G_K = -b$. The affine Jacobian coefficient is $3/200$, but the one-sided cone coefficients are

$$\Gamma_{K,+} = \Gamma_{\mathrm{comm}} = -1/1600, \qquad \Gamma_{K,-} = 3/200.$$

For $r > 0$, the schedule $\alpha_{1,2} = 1 - 19r/40 \pm \sqrt{5r/4}$ attains $F_{\boldsymbol{\alpha}}(S_r) = F_0 + rD + O(r^2)$.

Global two-step enumeration. For steps in $[\ell, u]^2$ write $c_1 = \alpha_1 + \alpha_2$, $c_2 = \alpha_1\alpha_2$. The filter at eigenvalue λ is $c_1\lambda - c_2\lambda^2$. The feasible coefficient region is

$$2\ell \le c_1 \le 2u, \qquad \max\{\ell(c_1 - \ell), u(c_1 - u)\} \le c_2 \le c_1^2/4.$$

Its interior corresponds to distinct interior roots. An optimum is either the unconstrained weighted least-squares solution in (c_1, c_2), on the equal-root curve, or on a box edge. On the equal-root curve the objective is quartic in the step; all real roots of its cubic derivative and both endpoints are checked. On each box edge the remaining one-dimensional objective is quadratic. Enumerating these cases yields the global population optimum. The rate experiment uses eight positive radii per control; the population sweep uses seven magnitudes with both signs (Appendix I.1).

The feasible coefficient region can also be derived directly. The roots of $x^2 - c_1x + c_2$ lie in $[\ell, u]$ exactly when the discriminant is nonnegative and the endpoint inequalities hold. The upper boundary $c_2 = c_1^2/4$ corresponds to two equal roots. For $c_1 \le \ell + u$, the active lower boundary is $c_2 = \ell(c_1 - \ell)$; for $c_1 \ge \ell + u$, it is $c_2 = u(c_1 - u)$. These are the two step-box faces after identifying schedule permutations.

For positive weights ω_i and target spectral coefficients t_i, minimize

$$L(c_1, c_2) = \sum_i \omega_i(c_1\lambda_i - c_2\lambda_i^2 - t_i)^2.$$

With at least two distinct positive eigenvalues, its coefficient Gram matrix is positive definite. If the unconstrained minimizer has admissible roots, it is a global lower-bound certificate. Otherwise a minimizer lies on the boundary of the compact feasible coefficient region. On an equal-root boundary, $L(2a, a^2)$ is quartic and its critical points are all real roots of a cubic derivative. On a face with one step fixed, the filter is affine in the remaining step and has one clipped quadratic minimizer. These alternatives exhaust the full step box, including its corners.

E.2 An attained square-root error at three equal steps

This construction distinguishes a worst-case singular approximation rate from the faster two-step special case. Take

$$S_r = S_0 = \mathrm{diag}(1/4, 1/2, 3/4), \quad \boldsymbol{\alpha}_0 = (1, 1, 1), \quad \Theta = [1/4, 5/4]^3, \quad W = \Sigma_r = I.$$

The nominal filter is $F_0 = I - (I - S_0)^3$. Identify diagonal matrices with their vectors. The three polynomial directions are

$$A = (9, 8, 3)/64, \qquad H = (3, 8, 9)/64, \qquad Q = (1, 8, 27)/64.$$

Let

$$R = Q - \mathrm{proj}_{\mathrm{span}(A,H)} Q = (1/17, -3/34, 1/17), \qquad w_0 = 1/5, \qquad D = w_0H + R.$$

Then $\langle R, A\rangle = \langle R, H\rangle = 0$, $\|R\|^2 = \langle R, Q\rangle = 1/68$, and the nearest right-side cone point is w_0H. Therefore

$$\Gamma_{3,+} = \|R\|^2 - \|D\|^2 = -77/51200.$$

A Gaussian response model with a fixed target distribution realizes the exact oracle path. Let $u \sim N(0, 2I)$ and let centered Gaussian noise ξ_r be independent of u. Define

$$z_r = \tfrac{1}{2}(F_0 + rD)^\top u + \xi_r, \qquad \mathrm{Cov}(\xi_r) = I - \tfrac{1}{2}(F_0 + rD)^\top (F_0 + rD).$$

The covariance is positive for sufficiently small r (and on the reported numerical grid). Then $\Sigma_r = I$ and $\mathbb{E}[uz_r^\top] = F_0 + rD$. This is a response-moment construction with a fixed target distribution, rather than the unbiased additive-noise model used by the two reversal witnesses. The curvature is fixed here; oracle movement alone suffices to make the cubic attainable adjustment statistically visible.

Proposition E.1 (Sharpness of the general singular rate)**.** *For this three-step family, with the exact reference and $r \downarrow 0$,*

$$\frac{\Delta R_r}{r^2} = \Gamma_{3,+} - \frac{\sqrt{15}}{3825}\sqrt{r} + o(\sqrt{r}). \tag{E.1}$$

Consequently the general $O(\sqrt{r})$ scaled-error bound cannot be replaced by $o(\sqrt{r})$ uniformly over equal-step instances.

Proof. For $\delta = \boldsymbol{\alpha} - \mathbf{1}$, the filter identity is exact:

$$F_{\mathbf{1}+\delta}(S_0) - F_0 = e_1 A - e_2 H + e_3 Q = e_1 A + \tfrac{1}{2}(\|\delta\|^2 - e_1^2)H + e_3 Q.$$

Near-optimal schedules obey $e_1 = O(r)$ and $\|\delta\|^2 = O(r)$. Write $x = e_1/r$, $w = \|\delta\|^2/(2r)$ and $q = e_3/r^{3/2}$. The scaled filter residual is

$$xA + wH - D + \sqrt{r}\, qQ + O(r)$$

uniformly on the relevant bounded sets.

For a centered three-vector ζ with $\sum_j \zeta_j = 0$ and $\|\zeta\|^2 = 2rw$, the product satisfies

$$|\zeta_1 \zeta_2 \zeta_3| \le \frac{2}{3\sqrt{3}}(rw)^{3/2}.$$

One proof maximizes the product on the compact circle defined by the two constraints. The multiplier equations imply either two coordinates agree or all three agree; the latter is impossible at positive radius. The extremizers are permutations of $(t, t, -2t)$ with $t^2 = rw/3$. Both signs are attained. Centering δ changes its product only by $O(r^2)$ because its mean is $O(r)$. Hence

$$|q| \le \frac{2w^{3/2}}{3\sqrt{3}} + O(\sqrt{r}),$$

with both limiting endpoints attainable.

The leading squared residual is uniquely minimized at $(x, w) = (0, w_0)$. Strict positive definiteness of the Gram matrix of A, H forces every sequence of near-minimizers to converge to these coordinates. At this point the residual is $-R$. The first perturbation of the minimized squared distance is therefore

$$-2\sqrt{r}\,\frac{2w_0^{3/2}}{3\sqrt{3}}\langle R, Q\rangle.$$

For completeness, the lower bound follows by the product inequality, boundedness of the relevant coordinates, and nonnegativity of the excess leading squared residual. For the upper bound choose the product-maximizing centered split at $w = w_0$, $x = 0$. Continuity as $(x, w) \to (0, w_0)$ makes both first-order perturbation bounds agree; the residual $O(r)$ is $o(\sqrt{r})$. The preserved scaled risk is exactly $\|D\|^2$. Finally $4w_0^{3/2}\langle R, Q\rangle/(3\sqrt{3}) = \sqrt{15}/3825$. □

The extremal product has two equal coordinates, which explains why the numerical three-step optimum lies on a repeated-root boundary. This boundary is a property of the current optimum; the nominal parameter remains the triply equal interior schedule. The result does not require estimating a convergence exponent from data.

Global three-step reduction. For three steps, write the filter as

$$c_1 S - c_2 S^2 + c_3 S^3, \qquad (c_1, c_2, c_3) = (e_1(\boldsymbol{\alpha}), e_2(\boldsymbol{\alpha}), e_3(\boldsymbol{\alpha})).$$

With three or more distinct positive eigenvalues, the risk is a strictly convex quadratic in c. Solve its unconstrained weighted least-squares problem. If $x^3 - c_1 x^2 + c_2 x - c_3$ has three real roots in the step box, the resulting schedule attains that unrestricted lower bound.

Otherwise a global minimizer lies on the boundary of the feasible coefficient image. At a schedule with distinct interior steps, the coefficient Jacobian is a nonsingular Vandermonde factor, so its image contains a neighborhood. Such a point cannot be a constrained minimizer unless it is the unrestricted quadratic minimizer. Every remaining candidate either has a step on the box boundary or has a repeated step. On a face with one step fixed at a, the filter is

$$aS + (I - aS)\{(b + c)S - bcS^2\}.$$

Because $I - aS$ is positive in the box, rescaling each spectral target and weight reduces this face exactly to the two-step problem.

For repeated steps (x, x, y), define at each eigenvalue

$$b_i(x) = 2x\lambda_i - x^2\lambda_i^2, \qquad B_i(x) = \lambda_i(1 - x\lambda_i)^2.$$

The filter is $b_i(x) + yB_i(x)$. For weights $\omega_i > 0$, the optimal y at fixed x is the box-clipped ratio $N(x)/Q(x)$, where

$$R_0(x) = \sum_i \omega_i (b_i(x) - t_i)^2, \quad N(x) = \sum_i \omega_i B_i(x)(t_i - b_i(x)), \quad Q(x) = \sum_i \omega_i B_i(x)^2 > 0.$$

On intervals where y is interior, the profiled objective is $R_0 - N^2/Q$. Every interior critical point is a real zero of

$$R_0' Q^2 - 2NN'Q + N^2 Q'. \tag{E.2}$$

When the optimum in y reaches an endpoint, it is already covered by the corresponding step-box face. The endpoints in x are also on those faces. Thus the unrestricted point, the six step faces (identified by permutation), and the real roots of (E.2) exhaust the global comparison.

The implementation isolates all real roots of (E.2) in the box using rational polynomial intervals for the decimal input coefficients. It then evaluates the candidate losses numerically. This certifies the critical-root enumeration for those rounded inputs; reported risks and matrices use floating-point arithmetic. The mathematical reduction establishes globality, while numerical tolerances determine the precision of its computed realization. We classify solutions as unrestricted lower-bound, repeated-boundary, or box-face optima.

E.3 Continuity of strict reversals

Stability of strict reversals. Within either regime, the risk coefficients vary continuously with the defining matrices and schedule parameters as long as the relevant ranks remain unchanged and the risk weights remain positive definite. Strict sign reversals therefore persist under sufficiently small perturbations. In the equal-step regime, these perturbations preserve equality of the nominal steps. For distinct steps, continuity follows from the polynomial derivatives, the divided differences at positive spectral gaps, and the nonsingular weighted Gram matrices defining the projectors. For equal steps, the projector onto A_K, the residual b, and the squared positive-part formulas in (11) are continuous. Positive covariance margins allow the statistical realizations to vary within their legal response-moment families. The neighborhood size depends on the spectral, step, and metric margins.

F Separated Spectral Clusters

This extension replaces scalar commuting coefficients by symmetric blocks within separated spectral groups. We state the block comparison, prove smooth projector transport, and give a repeated-eigenvalue example. The comparison class in this appendix is block symmetric rather than scalar spectral. It allows arbitrary symmetric coefficients within each specified spectral group. This distinction is necessary at internal eigenvalue collisions.

F.1 Separated groups and smooth projector transport

A stable spectral partition can remain identifiable despite multiplicities within a group. Let $P_a(r)$ project onto groups separated by $\Delta_{\mathrm{cl}} > 0$ and set $\mathcal{C}_r^{\mathrm{cl}} = \{\sum_a P_a(r) M_a P_a(r) : M_a = M_a^\top\}$. Group projectors remain smooth under a positive inter-group separation (Kato, 1995; Davis & Kahan, 1970).

Corollary F.1 (Cluster velocities)**.** *Suppose that the spectral groups of S_0 have a positive inter-group gap Δ_{cl} and that F_0 belongs to their symmetric block family; eigenvalue multiplicities within each group are allowed. Write $S_a = P_a(0) S_0 P_a(0)$ restricted to* $\operatorname{range} P_a(0)$*, $E_{ab} = P_a(0) E P_b(0)$, and $E_{\mathrm{cross}} = \sum_{a \neq b} E_{ab}$. Let the skew matrix Ω, with blocks $\Omega_{ab} = P_a(0)\Omega P_b(0)$, have zero diagonal blocks and solve $S_a \Omega_{ab} - \Omega_{ab} S_b = -E_{ab}$ for $a \neq b$. Then $V_s = [\Omega, F_0] + \mathcal{C}_0^{\mathrm{cl}}$. Under the statistical-path and reference-accuracy conditions of Theorem 4.1, the risk coefficient is Γ_{cl} below; the transport bound uses the spectral conditions above:*

$$\Gamma_{\mathrm{cl}} = \operatorname{dist}_0^2(D, [\Omega, F_0] + \mathcal{C}_0^{\mathrm{cl}}) - \|D\|_0^2, \qquad \|[\Omega, F_0]\|_F \leq \frac{2\|F_0\|_{\mathrm{op}} \|E_{\mathrm{cross}}\|_F}{\Delta_{\mathrm{cl}}}.$$

Only cross-group gaps enter the transport bound (Kato, 1995; Bhatia & Rosenthal, 1997). The block family includes matrices that need not commute with curvature within a group. At a simple spectrum

and common nominal filter, $\Gamma_{K,s} \geq \Gamma_{\mathrm{comm}} \geq \Gamma_{\mathrm{cl}}$; at an internal collision, only the block comparison is asserted.

Smooth projector transport. Assume the nominal spectral groups of the symmetric C^2 path S_r have gap $\Delta_{\mathrm{cl}} > 0$. The Riesz projector for each group is a contour integral of the resolvent; differentiation under this finite-dimensional integral shows that $P_a(r)$ is C^2. A simultaneous orthogonal transport exists. For example set $\Omega(r) = \frac{1}{2}\sum_a [P_a'(r), P_a(r)]$ and solve $U_r' = \Omega(r)U_r$, $U_0 = I$. Projector identities give $[\Omega(r), P_a(r)] = P_a'(r)$, so $P_a(r) = U_r P_a(0) U_r^\top$. At zero Ω has zero diagonal blocks. Differentiating the block diagonalization of S_r yields

$$S_a \Omega_{ab} - \Omega_{ab} S_b = -E_{ab}, \qquad a \neq b.$$

In eigenbases of the two blocks this equation divides entries by $\lambda_i - \lambda_j$, whose absolute values exceed Δ_{cl}. It follows that $\|\Omega\|_F \leq \|E_{\mathrm{cross}}\|_F / \Delta_{\mathrm{cl}}$. This is the classical invariant-subspace construction (Kato, 1995; Greenbaum et al., 2020).

F.2 Attainable velocities and the comparison-class boundary

Let $\mathcal{C}_0^{\mathrm{cl}}$ be the full symmetric block-diagonal space and $F_0 \in \mathcal{C}_0^{\mathrm{cl}}$. The entire class is $\mathcal{C}_r^{\mathrm{cl}} = U_r \mathcal{C}_0^{\mathrm{cl}} U_r^\top$. For bounded scaled filters,

$$\frac{U_r A_r U_r^\top - F_0}{r} = [\Omega, F_0] + \frac{A_r - F_0}{r} + o(1).$$

Conjugation gives outer completeness; $A_r = F_0 + rH$ for arbitrary $H \in \mathcal{C}_0^{\mathrm{cl}}$ gives inner attainability. Thus $V = G_{\mathrm{cl}} + \mathcal{C}_0^{\mathrm{cl}}$ with $G_{\mathrm{cl}} = [\Omega, F_0]$. The risk formula follows from Theorem 4.1. The commutator bound gives $\|G_{\mathrm{cl}}\|_F \leq 2\|F_0\|_{\mathrm{op}} \|E_{\mathrm{cross}}\|_F / \Delta_{\mathrm{cl}}$. No internal eigenvalue difference occurs. Since G_{cl} is off-block, its weighted residual vanishes exactly when $G_{\mathrm{cl}} = 0$. For general within-block F_0, this condition is $[\Omega, F_0] = 0$; it need not be equivalent to $[E, F_0] = 0$.

Boundary and repeated-eigenvalue example. The class permits arbitrary symmetric within-cluster matrices, including ones that do not commute with S_r inside a cluster. It is consequently a different, larger comparator. Under a simple spectrum and a common nominal scalar filter, nesting gives $\Gamma_{K,s} \geq \Gamma_{\mathrm{comm}} \geq \Gamma_{\mathrm{cl}}$. At an internal collision the middle coefficient from the simple-spectrum formula is not asserted. For the fixed repeated-eigenvalue check $S_0 = \mathrm{diag}(1, 1, 3)$ and $F_0 = \mathrm{diag}(0.4, 0.4, 0.8)$ with $E_{13} = E_{31} = 1$, one has $\Omega_{13} = 1/2$, $(G_{\mathrm{cl}})_{13} = 0.2$ and $\|G_{\mathrm{cl}}\|_F^2 = 0.08$. The included within-cluster splitting changes no first-order coefficient.

For the repeated-eigenvalue example, partition the first two coordinates as one group and the third as the other. The within-group entries of E may split the repeated eigenvalue at first order. They do not appear in the off-block Sylvester equation. At zero, G_{cl} has entries 0.2 in positions $(1, 3)$ and $(3, 1)$, giving the squared Frobenius norm 0.08. For fixed oracle F_0, the optimal current block approximation has risk gap $0.08r^2 + O(|r|^3)$. This is an example with multiplicity inside a separated group, not an application of a simple-eigenvalue formula with a zero denominator.

Allowing symmetric blocks makes more corrections available than adjusting scalar spectral coefficients. If a group contains several distinct eigenvalues, an arbitrary symmetric block need not commute with the restricted curvature. Consequently the block coefficient is a lower bound for the scalar commuting coefficient when both comparisons share the same nominal filter. At an internal eigenvalue collision the block family remains smooth, while the original commuting family can change dimension discontinuously. The extension obtains stability by making the comparator explicit.

The projector transport uses only cross-group denominators. This controls the matrix derivative independently of eigenvalue differences inside a group. It does not give a uniform bound as the separation between groups tends to zero. The theorem concerns a fixed partition with a positive inter-group gap.

G Finite-Update Derivative and Filter Motion

The main preserve-versus-retune theorem is formulated directly in terms of attainable output families. This appendix connects the filter-level velocity used there to the underlying finite update composition. Finite-pass regularization and learning-rate analyses provide the broader computational setting for this calculation (Rosasco & Villa, 2015; Pillaud-Vivien et al., 2018; Dieuleveut & Bach, 2016; Ge et al., 2019).

Consider a constant-step least-squares update with

$$J_r = I - aS_r, \qquad F_r = I - J_r^K,$$

where $a > 0$ is fixed and $S_r = S_0 + rE + O(r^2)$. At $r = 0$, write $J = I - aS_0$ and define the classical Fréchet derivative of the matrix-power map (Mathias, 1996; Higham & Relton, 2014) by

$$\mathcal{L}_K(B) = \sum_{j=0}^{K-1} J^{K-1-j} B J^j.$$

Because the perturbation of J_r is $B = -aE$, differentiating $F_r = I - J_r^K$ gives the fixed-schedule filter velocity

$$G_K = -\mathcal{L}_K(-aE).$$

This is the constant-step form of the curvature-forced velocity in the main text.

For nonconstant schedules, the same calculation is applied to the product $I - \prod_{k=1}^K (I - \alpha_k S_r)$. The product rule then gives Eq. (5), with the perturbation E inserted once at each update location and the remaining factors kept in their original order. The order matters when E does not commute with S_0.

This calculation connects the filter-level velocity used in the main text to the underlying finite update composition. The preserve-versus-retune risk additionally depends on the oracle drift and on the schedule adjustments attainable under the fixed update budget.

H Estimation of the Commuting-Benchmark Coefficient

This appendix develops perturbation bounds and consistency for the commuting-benchmark coefficient under a fixed positive eigengap. Throughout, $\Gamma = \Gamma_{\mathrm{comm}}$ and G is its curvature transport derivative. Valid simultaneous error bounds yield uncertainty-aware sign certificates; the corresponding numerical results appear in Appendix I.4.

H.1 An aligned commuting-filter estimator

Estimate nearby operators and unrestricted fixed-linear oracles, and form

$$\widehat{E} = \frac{\widehat{S}_{+h} - \widehat{S}_{-h}}{2h}, \qquad \widehat{D} = \frac{\widetilde{F}^{\star}_{+h} - \widetilde{F}^{\star}_{-h}}{2h}.$$

Let $\widehat{\Pi}$ project onto $\widehat{\mathcal{C}}_0 = \{F = F^\top : F\widehat{S}_0 = \widehat{S}_0 F\}$ in the metric $\langle A, B\rangle_{\widehat{0}} = \mathrm{tr}(A^\top \widehat{W} B \widehat{\Sigma}_0)$. Align the nominal estimate by $\widehat{F}_0 = \widehat{\Pi}\widetilde{F}_0$. Writing $\widehat{S}_0 = \sum_i \widehat{\lambda}_i \widehat{P}_i$ and $\widehat{F}_0 = \sum_i \widehat{f}_i \widehat{P}_i$, define

$$\widehat{G} = \sum_{i \neq j} \frac{\widehat{f}_i - \widehat{f}_j}{\widehat{\lambda}_i - \widehat{\lambda}_j} \widehat{P}_i \widehat{E} \widehat{P}_j, \qquad \widehat{g} = (\mathrm{Id} - \widehat{\Pi})\widehat{G},$$

$$\widehat{\Gamma} = \|\widehat{g}\|_{\widehat{0}}^2 - 2\langle \widehat{g}, \widehat{D}\rangle_{\widehat{0}} - \left\|\widehat{\Pi}\widehat{D}\right\|_{\widehat{0}}^2. \tag{H.1}$$

The projection is a weighted Gram solve; it need not equal diagonal extraction from $\widetilde{F}_0$. Estimating the oracle drift requires labelled tasks or another route to identifying $C_r = \mathbb{E}[uz_r^\top]$. Operator refits alone do not identify D or $\Gamma(D)$.

For the perturbation bound, define the nominal eigengap $\Delta = \min_{i \neq j} |\lambda_i - \lambda_j|$. Here, $\otimes$ denotes the Kronecker product. Set $H = \Sigma_0 \otimes W$, $\widehat{H} = \widehat{\Sigma}_0 \otimes \widehat{W}$, $\delta_\Delta = \min\{1, \Delta\}$, and

$$\epsilon_S = \left\|\widehat{S}_0 - S_0\right\|_{\mathrm{op}}, \quad \epsilon_E = \left\|\widehat{E} - E\right\|_F, \quad \epsilon_F = \left\|\widehat{F}_0 - F_0\right\|_F,$$
$$\epsilon_D = \left\|\widehat{D} - D\right\|_F, \qquad \epsilon_H = \left\|\widehat{H} - H\right\|_{\mathrm{op}}.$$

Proposition H.1 (Consistency and margin-separated sign recovery)**.** *Fix the dimension and a simple nominal spectrum with gap $\Delta > 0$. Suppose the population and estimated primitive matrices are locally bounded, $\widehat{S}_0, \widehat{E}$ are symmetric, $\widehat{F}_0 \in \widehat{\mathcal{C}}_0$, and the population and estimated risk metrics are positive definite with eigenvalues uniformly bounded away from zero. There is a constant C, independent of Δ, depending only on dimension and these norm and metric bounds, such that, for $\epsilon_S \leq \Delta/4$,*

$$|\widehat{\Gamma} - \Gamma(D)| \leq C \left\{ \frac{\epsilon_E + \epsilon_F + \epsilon_H}{\delta_\Delta^2} + \frac{\epsilon_S}{\delta_\Delta^3} + \frac{\epsilon_D}{\delta_\Delta} \right\}. \tag{H.2}$$

If $\epsilon_S, \epsilon_E, \epsilon_F, \epsilon_D, \epsilon_H = o_p(1)$, then $\widehat{\Gamma} \to_p \Gamma(D)$. For any fixed margin $|\Gamma(D)| \geq \gamma_0 > 0$, $\Pr\{\mathrm{sign}(\widehat{\Gamma}) = \mathrm{sign}(\Gamma(D))\} \to 1$.

For independent nearby fits with $O_p(n^{-1/2})$ error and paths with bounded third derivatives, central differences give $\epsilon_E, \epsilon_D = O_p(h^2 + n^{-1/2}/h)$. Thus $h \to 0$ and $n^{-1/2}/h \to 0$ suffice at a fixed gap when the nominal estimates and metric are consistent. The result diagnoses the limiting coefficient; its implication for learned-filter risk additionally requires reference error $o_p(|r|)$.

A certificate that includes estimation uncertainty. Let $\widehat{\kappa} = \|\widehat{g}\|_{\widehat{0}}$. On an event where $|\widehat{\kappa} - \kappa| \leq e_\kappa$ and $\|D\|_0 \leq d_U$, set $\kappa_L = (\widehat{\kappa} - e_\kappa)_+$. If

$$\kappa_L > 2d_U, \tag{H.3}$$

then every $\|D'\|_0 \leq d_U$ satisfies $\Gamma(D') \geq \kappa_L(\kappa_L - 2d_U) > 0$. Use an externally justified $d_U = d_{\max}$, or $d_U = \left\|\widehat{D}\right\|_{\widehat{0}} + e_d$ with a valid error bound. If the bounds hold jointly with probability at least $1 - \alpha$, the probability of falsely certifying the actual direction is at most α. An unadjusted plug-in inequality supplies no such finite-sample guarantee. We prove Proposition H.1 with explicit intermediate bounds. The dimension $p \geq 2$ is fixed. For $p = 1$, the commuting class is the entire matrix space, $G = g = 0$, and $\Gamma(D) = -\|D\|_0^2$.

H.2 Aligned estimands and weighted projections

Assume

$$S_0 = S_0^\top = \sum_i \lambda_i P_i, \quad E = E^\top, \quad F_0 = \sum_i f_i P_i, \quad W, \Sigma_0 \succ 0,$$

where $P_i = u_i u_i^\top$ are rank-one spectral projectors with ordered eigenvalues. The oracle drift D is an arbitrary real matrix. The unrestricted fixed-linear oracle is $F_r^\star = C_r \Sigma_r^{-1}$, with $C_r = \mathbb{E}[u z_r^\top]$. Estimating D therefore requires labelled task pairs (u, z_r) or another justified identification of these cross-moments. Unlabelled operator refits alone are insufficient. Estimating an oracle constrained to commute with S_r would estimate a different derivative, so D is not projected into the commuting class.

Let vec stack the columns of a matrix, and write $\mathrm{sym}(A) = (A + A^\top)/2$. Set

$$B = [\mathrm{vec}(P_1), \ldots, \mathrm{vec}(P_p)], \quad H = \Sigma_0 \otimes W, \quad M = B^\top H B.$$

Then $B^\top B = I_p$. The Frobenius projection onto $\mathcal{C}_0$ has matrix $BB^\top$; the weighted projection Π has matrix

$$\mathcal{P} = B M^{-1} B^\top H. \tag{H.4}$$

Hatted quantities use the estimated eigenspaces. For the deterministic argument, suppose

$$\mu I \preceq H \preceq \Lambda I, \quad \mu I \preceq \widehat{H} \preceq \Lambda I, \qquad \max\{\|E\|_F, \left\|\widehat{E}\right\|_F, \|F_0\|_F, \left\|\widehat{F}_0\right\|_F, \|D\|_F, \left\|\widehat{D}\right\|_F\} \leq L \tag{H.5}$$

for $\mu > 0$, $\Lambda \geq \mu$, and finite L. Consequently $\mu I_p \preceq M \preceq \Lambda I_p$ and $\mu I_p \preceq \widehat{M} \preceq \Lambda I_p$ and $\|\mathcal{P}\|_{\mathrm{op}}, \left\|\widehat{\mathcal{P}}\right\|_{\mathrm{op}} \leq q := \Lambda/\mu$. Gram positivity follows from positive-definite risk weights and the linearly independent spectral-projector basis. Estimated weights must preserve this positivity, for example by consistent eigenvalue clipping. The dependence on μ below excludes a metric-singularity limit.

For a raw estimate $\widetilde{F}_0$, define

$$\widehat{F}_0 = \widehat{\Pi}\widetilde{F}_0, \quad \widehat{M}_{ij} = \mathrm{tr}(\widehat{P}_i^\top \widehat{W} \widehat{P}_j \widehat{\Sigma}_0), \quad \widehat{f} = \widehat{M}^{-1}\widehat{v}, \quad \widehat{v}_i = \mathrm{tr}(\widehat{P}_i^\top \widehat{W} \widetilde{F}_0 \widehat{\Sigma}_0).$$

The raw estimate may be nonsymmetric. If $\widetilde{F}_0 = \widehat{C}_0 \widehat{\Sigma}_0^{-1}$, this is the empirical optimal fixed commuting filter under the same moments. Frobenius pinching $\sum_i \widehat{P}_i \,\mathrm{sym}(\widetilde{F}_0)\widehat{P}_i$ also gives an aligned consistent estimate, but generally differs from the empirical risk projection with non-Frobenius weights.

H.3 Perturbation bounds and error envelopes

Eigenvector and subspace perturbation bounds depend on the population eigengap (Davis & Kahan, 1970; Yu et al., 2015). Here, those bounds are combined with a weighted Gram solve for the risk projection.

If $\epsilon_S = \left\|\widehat{S}_0 - S_0\right\|_{\mathrm{op}} \leq \Delta/4$, eigenvalue perturbation gives $|\widehat{\lambda}_i - \lambda_i| \leq \epsilon_S$, and the estimated minimum gap is at least $\Delta - 2\epsilon_S \geq \Delta/2$. Project the residual equation

$$(S_0 - \widehat{\lambda}_i I)\widehat{u}_i = -(\widehat{S}_0 - S_0)\widehat{u}_i$$

onto $\mathrm{span}\{u_j : j \neq i\}$. All eigenvalue magnitudes there are at least $\Delta - \epsilon_S$, hence $\|(I - P_i)\widehat{u}_i\| \leq \epsilon_S/(\Delta - \epsilon_S)$. Since $\left\|\widehat{P}_i - P_i\right\|_F = \sqrt{2}\,\|(I - P_i)\widehat{u}_i\|$,

$$\left\|\widehat{B} - B\right\|_{\mathrm{op}} \leq \left\|\widehat{B} - B\right\|_F \leq t_B := 4\sqrt{p}\,\epsilon_S/\Delta. \tag{H.6}$$

Projectors avoid any eigenvector sign ambiguity.

Writing $\epsilon_H = \left\|\widehat{H} - H\right\|_{\mathrm{op}}$, we have

$$\left\|\widehat{M} - M\right\|_{\mathrm{op}} \leq t_M := \epsilon_H + 2\Lambda t_B, \qquad \left\|\widehat{M}^{-1} - M^{-1}\right\|_{\mathrm{op}} \leq t_M/\mu^2.$$

Telescope the four factors in (H.4) to obtain

$$\left\|\widehat{\mathcal{P}} - \mathcal{P}\right\|_{\mathrm{op}} \leq t_\Pi := 2qt_B + \frac{\Lambda t_M}{\mu^2} + \frac{\epsilon_H}{\mu} \leq C(\epsilon_S/\Delta + \epsilon_H). \tag{H.7}$$

Since $\Pi F_0 = F_0$, the weighted alignment satisfies

$$\epsilon_F = \left\|\widehat{\Pi}\widetilde{F}_0 - F_0\right\|_F \leq q\widetilde{\epsilon}_F + t_\Pi L, \qquad \widetilde{\epsilon}_F = \left\|\widetilde{F}_0 - F_0\right\|_F. \tag{H.8}$$

This includes alignment error from both the estimated spectrum and the estimated metric. For Frobenius pinching, the corresponding bound is $\epsilon_F \leq \widetilde{\epsilon}_F + 2t_B L$. After substitution, (H.2) also holds with the raw nominal error $\widetilde{\epsilon}_F$ in place of ϵ_F, after changing its constant: the extra terms are covered by $\epsilon_S/\delta_\Delta^3$ and $\epsilon_H/\delta_\Delta^2$.

Errors in both weights are included through

$$\epsilon_H \leq \|\Sigma_0\|_{\mathrm{op}} \left\|\widehat{W} - W\right\|_{\mathrm{op}} + \left\|\widehat{W}\right\|_{\mathrm{op}} \left\|\widehat{\Sigma}_0 - \Sigma_0\right\|_{\mathrm{op}}. \tag{H.9}$$

Thus consistent bounded weight estimates imply $\epsilon_H = o_p(1)$. No independence among plug-in estimates is needed for these deterministic bounds or their consistency consequence.

Perturbing the geometric derivative. Let $\mathcal{A}_S(X) = SX - XS$, and denote the Frobenius projection by Π_{F}. The defining equations for G and $\widehat{G}$ are

$$\mathcal{A}_{S_0}(G) = [F_0, E], \quad \Pi_{\mathrm{F}} G = 0, \qquad \mathcal{A}_{\widehat{S}_0}(\widehat{G}) = [\widehat{F}_0, \widehat{E}], \quad \widehat{\Pi}_{\mathrm{F}}\widehat{G} = 0.$$

On the Frobenius-orthogonal complement of the respective commuting space, the commutator maps have inverse norms at most $1/\Delta$ and $2/\Delta$: in an eigenbasis, entry (i,j) is multiplied by $\lambda_i - \lambda_j$. The inequality $\|[A,C]\|_F \le 2\|A\|_{\mathrm{op}}\|C\|_F$ gives

$$\|G\|_F \le B_G := 2L^2/\Delta, \qquad \left\|\widehat{G}\right\|_F \le 2B_G.$$

Subtracting the defining equations yields

$$\mathcal{A}_{\widehat{S}_0}(\widehat{G} - G) = [\widehat{F}_0, \widehat{E}] - [F_0, E] - [\widehat{S}_0 - S_0, G].$$

Its inverse on the estimated off-diagonal space bounds that component by $4L(\epsilon_E+\epsilon_F)/\Delta + 4\epsilon_S B_G/\Delta$. The remaining component is $-\widehat{\Pi}_{\mathrm{F}}G = -(\widehat{\Pi}_{\mathrm{F}} - \Pi_{\mathrm{F}})G$. Since $\left\|\widehat{B}\widehat{B}^\top - BB^\top\right\|_{\mathrm{op}} \le 2t_B$,

$$\left\|\widehat{G} - G\right\|_F \le t_G := \frac{4L(\epsilon_E+\epsilon_F)}{\Delta} + \left(\frac{4\epsilon_S}{\Delta} + 2t_B\right) B_G \le C\{(\epsilon_E+\epsilon_F)/\Delta + \epsilon_S/\Delta^2\}. \quad \text{(H.10)}$$

The alignment condition $\widehat{F}_0 \in \widehat{\mathcal{C}}_0$ is used in the hatted commutator equation and cannot be omitted.

Let $p_D = \Pi D$ and $\widehat{p}_D = \widehat{\Pi}\widehat{D}$. Then

$$\begin{aligned} \|\widehat{g} - g\|_F &\le t_g := (1+q)t_G + t_\Pi B_G, \\ \|\widehat{p}_D - p_D\|_F &\le t_p := q\epsilon_D + t_\Pi L. \end{aligned} \quad \text{(H.11)}$$

Both $\|g\|_F$ and $\|\widehat{g}\|_F$ are bounded by $g_* := 2(1+q)B_G$; both projected-drift norms are bounded by $p_* := qL$.

Explicit error envelopes. For vectorized matrices and $\|H'\|_{\mathrm{op}} \le \Lambda$, $\|H' - H\|_{\mathrm{op}} \le \epsilon_H$, elementary telescoping gives

$$\begin{aligned} |x'^\top H'x' - x^\top Hx| &\le \Lambda(\|x'\| + \|x\|)\,\|x' - x\| + \epsilon_H\,\|x\|^2, \\ |x'^\top H'y' - x^\top Hy| &\le \Lambda\,\|y'\|\,\|x' - x\| + \Lambda\,\|x\|\,\|y' - y\| + \epsilon_H\,\|x\|\,\|y\|. \end{aligned}$$

Applying these bounds to the three terms of Γ proves

$$\begin{aligned} |\widehat{\Gamma} - \Gamma| \le e_\Gamma :=\ & 2\Lambda(g_* + L)t_g + 2\Lambda g_*\epsilon_D + 2\Lambda p_* t_p \\ & + \epsilon_H\{g_*^2 + 2g_*L + p_*^2\}. \end{aligned} \quad \text{(H.12)}$$

The right side is determined by assumed primitive error bounds and norm, metric and gap bounds. With $\delta_\Delta = \min\{1, \Delta\}$,

$$t_\Pi \le C(\epsilon_S/\delta_\Delta + \epsilon_H), \quad t_g \le C\{(\epsilon_E+\epsilon_F+\epsilon_H)/\delta_\Delta + \epsilon_S/\delta_\Delta^2\}, \quad t_p \le C(\epsilon_D + \epsilon_S/\delta_\Delta + \epsilon_H),$$

and $g_* \le C/\delta_\Delta$, $p_* \le C$. Combining these estimates gives (H.2).

For any fixed A, factoring the difference of squared norms and using the metric lower bound gives

$$|\,\|A\|_{\widehat{0}} - \|A\|_0\,| \le \frac{\epsilon_H}{2\sqrt{\mu}}\,\|A\|_F.$$

The triangle inequality therefore yields

$$\begin{aligned} |\widehat{\kappa} - \kappa| &\le e_\kappa := \sqrt{\Lambda}\,t_g + \frac{\epsilon_H g_*}{2\sqrt{\mu}} \le C\{(\epsilon_E+\epsilon_F+\epsilon_H)/\delta_\Delta + \epsilon_S/\delta_\Delta^2\}, \\ |\left\|\widehat{D}\right\|_{\widehat{0}} - \|D\|_0| &\le e_d := \sqrt{\Lambda}\,\epsilon_D + \frac{\epsilon_H L}{2\sqrt{\mu}} \le C(\epsilon_D + \epsilon_H). \end{aligned} \quad \text{(H.13)}$$

For random inputs, these formulas give confidence statements only after valid simultaneous probability bounds for all their inputs (including the eigengap and metric bounds) have been supplied. The deterministic formulas themselves do not produce confidence bounds.

H.4 CONSISTENCY, CENTRAL DIFFERENCES, AND SIGN RECOVERY

For the Gaussian moment estimates used in the diagnostic, sample-covariance concentration supplies the primitive fixed-dimensional rates (Koltchinskii & Lounici, 2017). The bounds below propagate primitive errors through the commuting coefficient.

At a fixed $\Delta > 0$, consistent primitive estimates and positive-definite population weights imply that the bounded event (H.5) and $\epsilon_S \le \Delta/4$ hold with probability tending to one, for fixed constants. Equations (H.8) and (H.2) then prove consistency. If $|\Gamma| \ge \gamma_0 > 0$, a sign error implies $|\widehat{\Gamma} - \Gamma| \ge \gamma_0$, whose probability tends to zero. At $\Gamma = 0$ no sign recovery follows. For a shrinking margin, the entire bound must instead be $o_p(|\Gamma_n|)$, with uniform norm and metric assumptions; a vanishing-gap or growing-dimension extension is not automatic.

Suppose uniformly at $r = \pm h_n$ that

$$\left\|\widehat{S}_r - S_r\right\|_F = O_p(q_n), \qquad \left\|\widetilde{F}_r^\star - F_r^\star\right\|_F = O_p(q_n).$$

If both paths have bounded third derivatives on a common interval, centered Taylor expansions give $\epsilon_E, \epsilon_D = O_p(q_n/h_n + h_n^2)$. The h_n^2 bias uses third-order smoothness. Differentiability at zero suffices for vanishing bias without that rate. Thus $h_n \to 0$ and $q_n/h_n \to 0$ suffice. For $q_n = n^{-1/2}$, balancing at $h_n \asymp n^{-1/6}$ gives $O_p(n^{-1/3})$ derivative error. Undersmoothing at $h_n \asymp n^{-1/4}$ gives $O_p(n^{-1/4})$ error and $O(n^{-1/2})$ bias. These also bound $\widehat{\Gamma}$ at a fixed gap when nominal and metric errors are $O_p(n^{-1/2})$. A fixed nonzero h generally leaves a persistent derivative bias.

Independent nearby fits need not cancel sampling error. Pairing tasks improves the generic q_n/h_n bound only under an additional quantitative cancellation property. For example, an average of independent paired contributions $[V_\ell(h) - V_\ell(-h)]/(2h)$ with uniformly bounded second moment has centered sampling error $O_p(n^{-1/2})$. For nonlinear fitted oracle maps, this property must hold for the fitted map or its justified linearization, including its remainder. Sharing tasks alone does not establish it.

A concrete nominal construction uses moment estimators. If $\widehat{C}_0 - C_0$ and $\widehat{\Sigma}_0 - \Sigma_0$ are $O_p(n^{-1/2})$ and $\lambda_{\min}(\Sigma_0) > 0$, the inverse identity gives $\widetilde{F}_0 = \widehat{C}_0\widehat{\Sigma}_0^{-1} = F_0 + O_p(n^{-1/2})$ in fixed dimension. Its weighted alignment obeys (H.8). Analogous conditions uniform at $\pm h_n$ can supply q_n. These rates require suitable moments of observed products; the geometry theorem alone does not imply them.

The scope of an uncertainty-aware certificate. Suppose an event $\mathcal{E}$, with probability at least $1 - \alpha$, provides $|\widehat{\kappa} - \kappa| \le e_\kappa$ and $\|D\|_0 \le d_U$. On $\mathcal{E}$, $\kappa \ge \kappa_L$. The sharp drift-ball formula gives, for every $\|D'\|_0 \le d_U$,

$$\Gamma(D') \ge (\kappa - d_U)_+^2 - d_U^2 \ge (\kappa_L - d_U)_+^2 - d_U^2.$$

If $\kappa_L > 2d_U$, this is $\kappa_L(\kappa_L - 2d_U) > 0$. A false certificate for the actual direction is therefore contained in $\mathcal{E}^c$ and has probability at most α. The probability bound controls false certification unconditionally. It requires the primitive error bounds to hold jointly. Separate error bounds may be combined by a union bound with total failure probability at most α.

A valid e_Γ also gives a direction-aware rule: $\widehat{\Gamma} - e_\Gamma > 0$ certifies a positive coefficient, and $\widehat{\Gamma} + e_\Gamma < 0$ certifies a negative one. Otherwise the interval does not determine the sign. For approximate bootstrap intervals, the validity of this rule depends on joint coverage of the underlying error bounds. For inference on nonsmooth functionals, see Dümbgen (1993); Fang & Santos (2019).

An upper bound for the actual drift certifies that direction and all directions in the ball of radius d_U. Coverage of an externally specified scientific family additionally requires that all its directions lie in that ball. One estimated drift norm does not establish this family bound, and unlabelled refits do not supply it.

From coefficient sign to learned-filter risk sign. Let $\widehat{F}_{\mathrm{ref}}$ be fixed conditionally on training information independent of the current evaluation response, and write $\epsilon = \left\|\widehat{F}_{\mathrm{ref}} - F_0\right\|_0$. Under the

smooth expansions of the local risk theorem,

$$\inf_{F\in\mathcal{C}_r}\mathcal{R}_r(F)-\mathcal{R}_r(\widehat{F}_{\mathrm{ref}})=r^2\Gamma(D)+O(|r|^3+|r|\epsilon+\epsilon^2). \tag{H.14}$$

The result at F_0 has remainder $O(|r|^3)$. Changing F_0 by a matrix of norm ϵ changes quadratic excess risk by at most $C(|r|\epsilon+\epsilon^2)$, since $F_r^\star - F_0 = O(r)$ and the nearby metrics are equivalent. For $r_n \to 0$ and $\epsilon_n = o_p(|r_n|)$, the scaled learned-reference risk gap converges to $\Gamma(D)$. This additional condition is separate from estimating the coefficient. For a specified nonzero r, a finite-sample risk certificate also needs valid numerical bounds on the remainder and reference error in (H.14).

What transfers to the algorithm-specific comparison. For a family contained in the commuting class and sharing the nominal filter, the population hierarchy gives $\Gamma_s \geq \Gamma_{\mathrm{comm}}$. A valid positive lower confidence bound for the latter therefore also gives a positive lower bound for the algorithm coefficient. This implication uses containment; it does not require estimating the restriction term. A negative commuting coefficient supplies no analogous algorithm-specific decision, as the regular witness shows.

The reference correction in Proposition B.2 applies to the target risk comparison, whether or not the coefficient is estimated. At a perturbation scale comparable to reference error, the corresponding squared distance to $D - H_r$ must be included. Consistency of an estimator for $\Gamma_{\mathrm{comm}}(D;0)$ does not remove this reference contribution. An uncertainty statement for the algorithm decision would need to account for both.

The numerical study in Appendix I.4 uses approximate bootstrap intervals. At small eigengaps, empirical coverage falls below the nominal level. We report these failures together with the complete replication results.

I Complete Numerical Results and Reproducibility

This appendix gathers the prespecified population protocol, every radius-level summary, all finite-radius failures, the fixed rate controls, and reproduction instructions after the theorem proofs and extensions. We evaluate the population quantities in the main theorems using exact response moments and the nominal oracle reference. These comparisons require neither sampled training sets nor estimated algorithm-specific coefficients. We study reference learning and diagnostic uncertainty in the separate commuting-benchmark estimation experiment.

I.1 Prespecified population protocol

Before generating instances, we fix the seed, matrix distribution, schedule box, radii, comparisons and failure policy. The seed is 2026091707. There are 18 instances in each of six distinct-step cells $(p,K)\in\{3,4,5\}\times\{2,3\}$ and each of three equal-step cells $(p,K)\in\{2,3,4\}\times\{2\}$. Thus the prespecified population sweep contains 162 instances. No instance or perturbation radius is discarded after observing its risk.

For each instance, draw a Gaussian $p\times p$ matrix and factor it into an orthogonal matrix Q and an upper-triangular matrix R (QR factorization), taking $U=Q$. Write $\mathrm{linspace}(a,b,p)$ for the p equally spaced points from a to b, including both endpoints. Set

$$\lambda_i=\mathrm{linspace}(0.18,0.72,p)_i+\varepsilon_i,\qquad \varepsilon_i\sim\mathrm{Unif}(-0.012,0.012),\qquad S_0=U\,\mathrm{diag}(\lambda_i)U^\top.$$

The perturbations ε_i are independent and identically distributed. For distinct schedules, the nominal steps are $(0.35,0.95)$ for $K=2$ and $(0.25,0.65,1.05)$ for $K=3$, with independent uniform jitter in $[-0.015,0.015]$. Equal two-step schedules use $(0.7,0.7)$. The common retuning box is $[0.1,1.2]^K$. The nominal spectra and jitter preserve distinctness and positive residuals.

Draw independent symmetric Gaussian directions and normalize their Frobenius norms to $\|E\|_F=0.04$ and $\|N'\|_F=0.03$. With $F_0=F_{\boldsymbol{\alpha}_0}(S_0)$, use

$$u\sim N(0,F_0),\qquad z_r=u+\xi_r,\qquad \xi_r\sim N(0,I-F_0+rN'),\qquad S_r=S_0+rE. \tag{I.1}$$

Signal and noise are independent, and the target distribution is fixed along the task path. Then

$$W=\Sigma_0=I,\quad \Sigma_r=I+rN',\quad F_r^\star=F_0(I+rN')^{-1},\quad D=-F_0N'.$$

The oracle drift need not be symmetric. All sampled matrices and nominal schedules are included in the source data. Positivity of the noise, curvature and residual factors is checked at every evaluated radius.

The radius magnitudes are 0.1×2^{-j}, $j = 0, \dots, 6$, with both signs. This gives $162 \times 7 \times 2 = 2268$ comparisons. The numerator is computed from oracle-centered quadratic excess risks, so the constant irreducible risk cancels before numerical evaluation. The reported quantity is

$$Y_r = \frac{\inf_{\boldsymbol{\alpha} \in [0.1,1.2]^K} \mathcal{R}_r(F_{\boldsymbol{\alpha}}(S_r)) - \mathcal{R}_r(F_0)}{r^2}.$$

The theoretical coefficient is calculated from the affine-space or cone projection at the nominal instance. No regression significance, fitted convergence exponent or data-dependent stopping rule is used.

Exact reduction from matrix risk to spectral fitting. The response model permits noncommuting curvature and noise perturbations. The actual step filter is diagonal in the current eigenbasis, but the reference and oracle generally are not. Let $S_r = U_r \operatorname{diag}(\lambda_i(r)) U_r^\top$, $\widetilde{\Sigma} = U_r^\top \Sigma_r U_r$ and $\widetilde{C} = U_r^\top F_0 U_r$. Since $W = I$, the part of risk depending on a diagonal current filter is

$$\sum_i \widetilde{\Sigma}_{ii} f_i^2 - 2 \sum_i \widetilde{C}_{ii} f_i.$$

It is a weighted fit to $t_i = \widetilde{C}_{ii} / \widetilde{\Sigma}_{ii}$ with weights $\omega_i = \widetilde{\Sigma}_{ii} > 0$. This reduction is exact for the generated model. Terms independent of the current diagonal coefficients are common to all retuned candidates.

The global algorithms are proved in Appendices E.1 and E.2. Two-step candidates comprise the unrestricted coefficient optimum, equal-root boundary and step-box faces. For three steps, a feasible unrestricted quadratic optimum is already a global lower bound. If it is infeasible, repeated-root critical points and step-box faces exhaust the boundary. Rational polynomial root isolation is applied to the decimal inputs of the repeated-root profile. This isolates roots of the polynomial built from rounded decimal inputs, not of an exact real-input model. The numerical matrices, roots and evaluated risks are reported at floating-point precision; candidate completeness and arithmetic accuracy are separate checks.

I.2 Complete budget-matched sweep

Table 1: All nine cells at the smallest radius. Each cell has 18 instances and 36 signed comparisons.

Schedule	p	K	Median error	Max. error	Signs
Distinct	3	2	7.44×10^{-8}	1.83×10^{-7}	36/36
Distinct	3	3	8.50×10^{-8}	4.80×10^{-7}	36/36
Distinct	4	2	7.78×10^{-8}	2.78×10^{-7}	36/36
Distinct	4	3	8.61×10^{-8}	4.02×10^{-7}	36/36
Distinct	5	2	7.18×10^{-8}	2.93×10^{-7}	36/36
Distinct	5	3	9.46×10^{-8}	3.37×10^{-7}	36/36
Equal	2	2	8.82×10^{-8}	2.89×10^{-7}	36/36
Equal	3	2	7.25×10^{-8}	2.12×10^{-7}	36/36
Equal	4	2	4.42×10^{-8}	2.38×10^{-7}	36/36

At the smallest radius, 8 of the 324 predicted coefficients are negative and 316 are positive. The random generator is not a balanced sign-classification design. Its purpose is to compare numerical coefficients with their formulas across admissible matrices and schedules. The exact signed witnesses supply controlled reversals in both directions.

Across the full grid, 3 of the 2268 finite-radius risk signs differ from the limiting coefficient sign. This is allowed at nonasymptotic radii; every such comparison is reported below. The smallest-radius maximum absolute error is 4.80×10^{-7}.

Among the global optima, 1134 lie in the coefficient interior and 57 on a repeated-step boundary found by root isolation. Another 681 attain the unrestricted lower bound, 18 lie on a box face, and

378 on the equal-step boundary. Numerical commuting-minus-step inequalities hold to a minimum residual of -2.26×10^{-15}, consistent with floating-point rounding at equality.

Radius-level summaries for every cell. The following table pools the two signs only within a specified cell and radius. It reports both median and maximum coefficient error. All 2,268 individual rows, including step coordinates and solver status, are supplied as machine-readable source data.

Table 2: Complete radius-level summaries. Each row contains 36 signed comparisons.

Schedule	(p, K)	$\lvert r\rvert$	Median error	Max. error	Signs
Distinct	(3,2)	0.1	4.76×10^{-6}	1.21×10^{-5}	36/36
Distinct	(3,2)	0.05	2.38×10^{-6}	5.94×10^{-6}	36/36
Distinct	(3,2)	0.025	1.19×10^{-6}	2.95×10^{-6}	36/36
Distinct	(3,2)	0.0125	5.95×10^{-7}	1.47×10^{-6}	36/36
Distinct	(3,2)	0.00625	2.98×10^{-7}	7.34×10^{-7}	36/36
Distinct	(3,2)	0.003125	1.49×10^{-7}	3.67×10^{-7}	36/36
Distinct	(3,2)	0.0015625	7.44×10^{-8}	1.83×10^{-7}	36/36
Distinct	(3,3)	0.1	5.71×10^{-5}	9.45×10^{-4}	34/36
Distinct	(3,3)	0.05	7.38×10^{-6}	4.60×10^{-4}	35/36
Distinct	(3,3)	0.025	1.36×10^{-6}	7.72×10^{-6}	36/36
Distinct	(3,3)	0.0125	6.80×10^{-7}	3.85×10^{-6}	36/36
Distinct	(3,3)	0.00625	3.40×10^{-7}	1.92×10^{-6}	36/36
Distinct	(3,3)	0.003125	1.70×10^{-7}	9.61×10^{-7}	36/36
Distinct	(3,3)	0.0015625	8.50×10^{-8}	4.80×10^{-7}	36/36
Distinct	(4,2)	0.1	4.99×10^{-6}	1.79×10^{-5}	36/36
Distinct	(4,2)	0.05	2.49×10^{-6}	8.92×10^{-6}	36/36
Distinct	(4,2)	0.025	1.25×10^{-6}	4.45×10^{-6}	36/36
Distinct	(4,2)	0.0125	6.23×10^{-7}	2.22×10^{-6}	36/36
Distinct	(4,2)	0.00625	3.11×10^{-7}	1.11×10^{-6}	36/36
Distinct	(4,2)	0.003125	1.56×10^{-7}	5.56×10^{-7}	36/36
Distinct	(4,2)	0.0015625	7.78×10^{-8}	2.78×10^{-7}	36/36
Distinct	(4,3)	0.1	1.45×10^{-5}	9.66×10^{-4}	36/36
Distinct	(4,3)	0.05	4.29×10^{-6}	4.59×10^{-4}	36/36
Distinct	(4,3)	0.025	1.44×10^{-6}	6.43×10^{-6}	36/36
Distinct	(4,3)	0.0125	6.89×10^{-7}	3.21×10^{-6}	36/36
Distinct	(4,3)	0.00625	3.44×10^{-7}	1.61×10^{-6}	36/36
Distinct	(4,3)	0.003125	1.72×10^{-7}	8.03×10^{-7}	36/36
Distinct	(4,3)	0.0015625	8.61×10^{-8}	4.02×10^{-7}	36/36
Distinct	(5,2)	0.1	4.62×10^{-6}	1.88×10^{-5}	36/36
Distinct	(5,2)	0.05	2.31×10^{-6}	9.38×10^{-6}	36/36
Distinct	(5,2)	0.025	1.15×10^{-6}	4.69×10^{-6}	36/36
Distinct	(5,2)	0.0125	5.75×10^{-7}	2.35×10^{-6}	36/36
Distinct	(5,2)	0.00625	2.87×10^{-7}	1.17×10^{-6}	36/36
Distinct	(5,2)	0.003125	1.44×10^{-7}	5.86×10^{-7}	36/36
Distinct	(5,2)	0.0015625	7.18×10^{-8}	2.93×10^{-7}	36/36
Distinct	(5,3)	0.1	6.10×10^{-6}	1.96×10^{-4}	36/36
Distinct	(5,3)	0.05	2.72×10^{-6}	1.08×10^{-5}	36/36
Distinct	(5,3)	0.025	1.52×10^{-6}	5.39×10^{-6}	36/36
Distinct	(5,3)	0.0125	7.58×10^{-7}	2.69×10^{-6}	36/36
Distinct	(5,3)	0.00625	3.79×10^{-7}	1.35×10^{-6}	36/36
Distinct	(5,3)	0.003125	1.89×10^{-7}	6.73×10^{-7}	36/36
Distinct	(5,3)	0.0015625	9.46×10^{-8}	3.37×10^{-7}	36/36
Equal	(2,2)	0.1	5.60×10^{-6}	1.86×10^{-5}	36/36
Equal	(2,2)	0.05	2.81×10^{-6}	9.26×10^{-6}	36/36
Equal	(2,2)	0.025	1.41×10^{-6}	4.63×10^{-6}	36/36
Equal	(2,2)	0.0125	7.05×10^{-7}	2.31×10^{-6}	36/36
Equal	(2,2)	0.00625	3.53×10^{-7}	1.16×10^{-6}	36/36
Equal	(2,2)	0.003125	1.76×10^{-7}	5.78×10^{-7}	36/36
Equal	(2,2)	0.0015625	8.82×10^{-8}	2.89×10^{-7}	36/36

Schedule	(p, K)	$\lvert r\rvert$	Median error	Max. error	Signs
Equal	(3,2)	0.1	4.60×10^{-6}	1.37×10^{-5}	36/36
Equal	(3,2)	0.05	2.31×10^{-6}	6.81×10^{-6}	36/36
Equal	(3,2)	0.025	1.16×10^{-6}	3.39×10^{-6}	36/36
Equal	(3,2)	0.0125	5.80×10^{-7}	1.70×10^{-6}	36/36
Equal	(3,2)	0.00625	2.90×10^{-7}	8.47×10^{-7}	36/36
Equal	(3,2)	0.003125	1.45×10^{-7}	4.23×10^{-7}	36/36
Equal	(3,2)	0.0015625	7.25×10^{-8}	2.12×10^{-7}	36/36
Equal	(4,2)	0.1	2.73×10^{-6}	1.52×10^{-5}	36/36
Equal	(4,2)	0.05	1.39×10^{-6}	7.61×10^{-6}	36/36
Equal	(4,2)	0.025	7.02×10^{-7}	3.80×10^{-6}	36/36
Equal	(4,2)	0.0125	3.53×10^{-7}	1.90×10^{-6}	36/36
Equal	(4,2)	0.00625	1.77×10^{-7}	9.51×10^{-7}	36/36
Equal	(4,2)	0.003125	8.84×10^{-8}	4.76×10^{-7}	36/36
Equal	(4,2)	0.0015625	4.42×10^{-8}	2.38×10^{-7}	36/36

All individual limiting-scale comparisons. Each row is one seeded instance, indexed by its instance identifier (ID). The one-sided coefficients are Γ_+ and Γ_-, and the observed scaled risks at $r = \pm.0015625$ are Y_+ and Y_-. Regular instances have identical limiting coefficients on both sides but need not have identical finite-radius risks. Values in the four coefficient columns are multiplied by 10^3. Full-precision values and optimizer coordinates are included in the supplementary data.

Table 3: All 162 instances at the smallest perturbation magnitude. Coefficients and risks are in units of 10^{-3}.

ID	Schedule	(p, K)	Γ_+	Y_+	Γ_-	Y_-
1	Distinct	(3,2)	0.275497	0.275442	0.275497	0.275553
2	Distinct	(3,2)	1.376194	1.376012	1.376194	1.376375
3	Distinct	(3,2)	1.518613	1.518734	1.518613	1.518493
4	Distinct	(3,2)	1.137540	1.137598	1.137540	1.137482
5	Distinct	(3,2)	1.938700	1.938883	1.938700	1.938517
6	Distinct	(3,2)	1.008126	1.008042	1.008126	1.008210
7	Distinct	(3,2)	0.634151	0.634129	0.634151	0.634172
8	Distinct	(3,2)	1.158909	1.158755	1.158909	1.159062
9	Distinct	(3,2)	1.133292	1.133213	1.133292	1.133372
10	Distinct	(3,2)	0.773693	0.773618	0.773693	0.773768
11	Distinct	(3,2)	0.989658	0.989505	0.989658	0.989811
12	Distinct	(3,2)	0.732310	0.732353	0.732310	0.732266
13	Distinct	(3,2)	0.755466	0.755540	0.755466	0.755392
14	Distinct	(3,2)	0.790368	0.790307	0.790368	0.790429
15	Distinct	(3,2)	1.015059	1.015127	1.015059	1.014991
16	Distinct	(3,2)	0.474781	0.474771	0.474781	0.474791
17	Distinct	(3,2)	1.715776	1.715706	1.715776	1.715846
18	Distinct	(3,2)	0.932982	0.932822	0.932982	0.933142
19	Distinct	(3,3)	0.553858	0.553767	0.553858	0.553950
20	Distinct	(3,3)	0.041561	0.041473	0.041561	0.041649
21	Distinct	(3,3)	0.593230	0.593197	0.593230	0.593263
22	Distinct	(3,3)	0.381816	0.382084	0.381816	0.381548
23	Distinct	(3,3)	0.459946	0.459991	0.459946	0.459900
24	Distinct	(3,3)	0.184704	0.184693	0.184704	0.184715
25	Distinct	(3,3)	0.250053	0.249971	0.250053	0.250134
26	Distinct	(3,3)	0.601634	0.601927	0.601634	0.601340
27	Distinct	(3,3)	0.929349	0.929740	0.929349	0.928959
28	Distinct	(3,3)	1.393190	1.393534	1.393190	1.392847
29	Distinct	(3,3)	-0.089188	-0.089146	-0.089188	-0.089231
30	Distinct	(3,3)	0.195266	0.195332	0.195266	0.195201
31	Distinct	(3,3)	1.229043	1.229016	1.229043	1.229070
32	Distinct	(3,3)	0.064789	0.064870	0.064789	0.064709
33	Distinct	(3,3)	1.175731	1.175962	1.175731	1.175501
34	Distinct	(3,3)	0.770800	0.771068	0.770800	0.770533

ID	Schedule	(p, K)	Γ_+	Y_+	Γ_-	Y_-
35	Distinct	(3,3)	0.847051	0.846571	0.847051	0.847531
36	Distinct	(3,3)	1.714326	1.714336	1.714326	1.714316
37	Distinct	(4,2)	1.408938	1.409004	1.408938	1.408873
38	Distinct	(4,2)	0.785237	0.785156	0.785237	0.785317
39	Distinct	(4,2)	1.981459	1.981372	1.981459	1.981546
40	Distinct	(4,2)	1.873876	1.873904	1.873876	1.873848
41	Distinct	(4,2)	1.025968	1.026011	1.025968	1.025925
42	Distinct	(4,2)	1.230807	1.230790	1.230807	1.230825
43	Distinct	(4,2)	1.145297	1.145523	1.145297	1.145071
44	Distinct	(4,2)	0.594017	0.593994	0.594017	0.594040
45	Distinct	(4,2)	1.433670	1.433792	1.433670	1.433547
46	Distinct	(4,2)	0.830072	0.830046	0.830072	0.830098
47	Distinct	(4,2)	1.844458	1.844590	1.844458	1.844326
48	Distinct	(4,2)	1.251827	1.251770	1.251827	1.251884
49	Distinct	(4,2)	1.361637	1.361499	1.361637	1.361775
50	Distinct	(4,2)	1.323225	1.323335	1.323225	1.323116
51	Distinct	(4,2)	1.093777	1.093701	1.093777	1.093852
52	Distinct	(4,2)	1.604574	1.604296	1.604574	1.604852
53	Distinct	(4,2)	2.315328	2.315269	2.315328	2.315388
54	Distinct	(4,2)	0.626112	0.626207	0.626112	0.626016
55	Distinct	(4,3)	0.719979	0.720062	0.719979	0.719897
56	Distinct	(4,3)	0.843695	0.843637	0.843695	0.843752
57	Distinct	(4,3)	2.213728	2.213441	2.213728	2.214015
58	Distinct	(4,3)	1.040802	1.040533	1.040802	1.041071
59	Distinct	(4,3)	0.744721	0.744955	0.744721	0.744487
60	Distinct	(4,3)	1.412199	1.412004	1.412199	1.412393
61	Distinct	(4,3)	1.369659	1.369694	1.369659	1.369624
62	Distinct	(4,3)	2.052635	2.052233	2.052635	2.053036
63	Distinct	(4,3)	0.244086	0.244015	0.244086	0.244157
64	Distinct	(4,3)	1.582930	1.582957	1.582930	1.582903
65	Distinct	(4,3)	0.909932	0.909842	0.909932	0.910021
66	Distinct	(4,3)	1.663675	1.663813	1.663675	1.663536
67	Distinct	(4,3)	1.066727	1.066742	1.066727	1.066712
68	Distinct	(4,3)	0.523267	0.523203	0.523267	0.523330
69	Distinct	(4,3)	0.590302	0.590403	0.590302	0.590200
70	Distinct	(4,3)	0.318383	0.318423	0.318383	0.318343
71	Distinct	(4,3)	0.971933	0.971711	0.971933	0.972155
72	Distinct	(4,3)	0.710161	0.710191	0.710161	0.710131
73	Distinct	(5,2)	1.855358	1.855435	1.855358	1.855280
74	Distinct	(5,2)	1.230294	1.230405	1.230294	1.230183
75	Distinct	(5,2)	1.585203	1.585276	1.585203	1.585131
76	Distinct	(5,2)	1.691096	1.691071	1.691096	1.691120
77	Distinct	(5,2)	0.482922	0.482853	0.482922	0.482990
78	Distinct	(5,2)	1.092400	1.092412	1.092400	1.092387
79	Distinct	(5,2)	1.621140	1.621283	1.621140	1.620998
80	Distinct	(5,2)	1.034459	1.034166	1.034459	1.034752
81	Distinct	(5,2)	1.276601	1.276584	1.276601	1.276618
82	Distinct	(5,2)	1.478860	1.478981	1.478860	1.478739
83	Distinct	(5,2)	1.573655	1.573544	1.573655	1.573766
84	Distinct	(5,2)	1.777239	1.777322	1.777239	1.777155
85	Distinct	(5,2)	1.202036	1.202020	1.202036	1.202052
86	Distinct	(5,2)	1.291008	1.290943	1.291008	1.291074
87	Distinct	(5,2)	1.392821	1.392912	1.392821	1.392730
88	Distinct	(5,2)	1.118948	1.119002	1.118948	1.118894
89	Distinct	(5,2)	1.407511	1.407582	1.407511	1.407440
90	Distinct	(5,2)	1.355125	1.355129	1.355125	1.355121
91	Distinct	(5,3)	1.690882	1.690806	1.690882	1.690959
92	Distinct	(5,3)	1.528857	1.529046	1.528857	1.528667
93	Distinct	(5,3)	1.614809	1.614603	1.614809	1.615015
94	Distinct	(5,3)	1.372007	1.371744	1.372007	1.372269
95	Distinct	(5,3)	2.227153	2.227294	2.227153	2.227011
96	Distinct	(5,3)	2.053739	2.053403	2.053739	2.054076
97	Distinct	(5,3)	1.604513	1.604436	1.604513	1.604591

ID	Schedule	(p, K)	Γ_+	Y_+	Γ_-	Y_-
98	Distinct	(5,3)	1.751556	1.751415	1.751556	1.751697
99	Distinct	(5,3)	1.816189	1.816270	1.816189	1.816109
100	Distinct	(5,3)	1.585296	1.585446	1.585296	1.585145
101	Distinct	(5,3)	1.696131	1.696061	1.696131	1.696202
102	Distinct	(5,3)	2.127685	2.127567	2.127685	2.127802
103	Distinct	(5,3)	1.587305	1.587390	1.587305	1.587220
104	Distinct	(5,3)	1.017205	1.017101	1.017205	1.017309
105	Distinct	(5,3)	1.616262	1.616281	1.616262	1.616243
106	Distinct	(5,3)	2.018579	2.018534	2.018579	2.018624
107	Distinct	(5,3)	2.313962	2.313949	2.313962	2.313976
108	Distinct	(5,3)	1.032848	1.032787	1.032848	1.032908
109	Equal	(2,2)	1.693778	1.693824	0.835050	0.835246
110	Equal	(2,2)	1.110873	1.110716	1.085230	1.085437
111	Equal	(2,2)	2.478549	2.478634	0.869028	0.869203
112	Equal	(2,2)	-0.142220	-0.142220	0.770602	0.770733
113	Equal	(2,2)	0.698152	0.698240	1.110585	1.110717
114	Equal	(2,2)	1.981089	1.981096	1.959139	1.959195
115	Equal	(2,2)	1.050715	1.050734	1.524353	1.524525
116	Equal	(2,2)	-0.238821	-0.238815	1.502415	1.502503
117	Equal	(2,2)	0.736445	0.736388	0.683099	0.683181
118	Equal	(2,2)	1.144497	1.144386	1.144318	1.144431
119	Equal	(2,2)	1.653181	1.653341	2.252303	2.252435
120	Equal	(2,2)	1.104766	1.104703	0.236645	0.236693
121	Equal	(2,2)	2.120808	2.121097	0.281393	0.281441
122	Equal	(2,2)	1.909757	1.909907	-0.354334	-0.354326
123	Equal	(2,2)	0.164332	0.164330	0.358050	0.358189
124	Equal	(2,2)	1.586235	1.586456	-0.071848	-0.071839
125	Equal	(2,2)	1.144946	1.145221	-0.258880	-0.258876
126	Equal	(2,2)	-0.047675	-0.047608	1.322986	1.323124
127	Equal	(3,2)	0.460076	0.459979	0.996386	0.996527
128	Equal	(3,2)	0.528901	0.528996	0.485213	0.485143
129	Equal	(3,2)	0.570892	0.570944	0.367446	0.367480
130	Equal	(3,2)	1.122197	1.122306	1.084172	1.084095
131	Equal	(3,2)	1.406089	1.406054	0.636512	0.636724
132	Equal	(3,2)	0.325123	0.325209	0.963145	0.963198
133	Equal	(3,2)	1.611652	1.611636	0.853454	0.853631
134	Equal	(3,2)	1.469138	1.469266	0.915910	0.915825
135	Equal	(3,2)	2.061456	2.061460	0.267410	0.267355
136	Equal	(3,2)	1.088326	1.088188	1.115851	1.115996
137	Equal	(3,2)	1.081390	1.081371	1.045902	1.045976
138	Equal	(3,2)	1.897501	1.897689	2.160015	2.159990
139	Equal	(3,2)	1.676201	1.676201	1.569332	1.569444
140	Equal	(3,2)	1.514680	1.514744	1.440436	1.440464
141	Equal	(3,2)	1.657420	1.657414	1.123622	1.123786
142	Equal	(3,2)	0.605603	0.605580	1.897301	1.897467
143	Equal	(3,2)	0.421583	0.421573	0.670370	0.670471
144	Equal	(3,2)	0.976929	0.976900	0.956699	0.956747
145	Equal	(4,2)	1.902325	1.902302	1.899181	1.899228
146	Equal	(4,2)	1.115016	1.115117	1.272855	1.272839
147	Equal	(4,2)	1.348863	1.348875	1.348374	1.348370
148	Equal	(4,2)	1.516152	1.516067	1.517539	1.517629
149	Equal	(4,2)	1.024915	1.025000	1.035425	1.035375
150	Equal	(4,2)	0.644997	0.644963	0.674224	0.674270
151	Equal	(4,2)	1.704206	1.704189	1.720834	1.720898
152	Equal	(4,2)	1.499112	1.499154	1.531663	1.531660
153	Equal	(4,2)	0.990857	0.990968	0.943616	0.943554
154	Equal	(4,2)	0.549956	0.549962	1.039387	1.039456
155	Equal	(4,2)	1.706191	1.706180	1.682855	1.682901
156	Equal	(4,2)	1.187342	1.187318	1.178297	1.178357
157	Equal	(4,2)	2.126538	2.126569	2.098041	2.098050
158	Equal	(4,2)	0.366336	0.366329	0.865915	0.865983
159	Equal	(4,2)	1.231773	1.231784	0.773264	0.773502
160	Equal	(4,2)	1.164004	1.163954	1.148031	1.148139

ID	Schedule	(p, K)	Γ_+	Y_+	Γ_-	Y_-
161	Equal	(4,2)	0.423518	0.423586	0.194589	0.194551
162	Equal	(4,2)	1.388577	1.388556	1.387238	1.387270

All finite-radius sign disagreements.

ID	Schedule	r	Γ	Y_r	Error
29	regular	-0.1	-8.92×10^{-5}	3.22×10^{-4}	4.11×10^{-4}
29	regular	0.1	-8.92×10^{-5}	1.75×10^{-5}	1.07×10^{-4}
29	regular	-0.05	-8.92×10^{-5}	4.74×10^{-5}	1.37×10^{-4}

Table 4: Every sign disagreement on the complete population grid.

I.3 Finite-perturbation rate controls and complete values

The rate experiment uses three fixed controls rather than additional random instances. The regular and equal two-step controls are the exact reversal witnesses of Appendix E. Their moments are evaluated without replacing the nonlinear oracle by its first-order approximation. The equal three-step control uses the exact linear oracle path and fixed curvature from Proposition E.1. All use the exact reference F_0.

Table 5: Configurations of the three fixed controls. Every control uses the exact reference F_0 and the positive grid $r = 0.1 \times 4^{-j}$, $j = 0, \dots, 7$. These boxes differ from the population-sweep box $[0.1, 1.2]^K$.

Control	Nominal schedule	Step box	Construction	Global solver
Distinct $K = 2$	$(1, 1/2)$	$[1/4, 5/4]^2$	App. E.1	Two-step enumeration
Equal $K = 2$	$(1, 1)$	$[1/2, 3/2]^2$	App. E.1	Two-step enumeration
Equal $K = 3$	$(1, 1, 1)$	$[1/4, 5/4]^3$	App. E.2	Faces and repeated roots

At $r = 0.1$, the equal two-step control attains steps $(0.5871845366, 1.3157290231)$ inside its original box $[0.5, 1.5]^2$, giving $\Delta R_r / r^2 = -0.000623630651519$. For comparison, imposing the different box $[0.1, 1.2]^2$ gives $(1.2, 0.7626860553)$ and $+0.005733124860479$. The latter is a different feasible family; it neither replaces the original protocol nor contradicts its local asymptotic result. All 24 reported control optima satisfy their own step bounds and have positive curvature, noise covariance and residual factors on the evaluated grid.

Numerical precision. Risk differences are evaluated as oracle-centered excess quadratics, with residuals divided by r before squaring. The rate calculations in Figure 2 use 80- and 120-digit arithmetic; the two precisions agree in scaled risk to better than 10^{-60}. For both two-step controls, the unconstrained coefficient optimum has feasible roots at all eight radii and certifies the global value. For three steps, high precision refines the repeated-root candidate selected by global boundary enumeration; the root-isolation implementation operates on decimal-rounded inputs.

The grid is $r = 0.1 \times 4^{-j}$, $j = 0, \dots, 7$. The measured error is $e(r) = |Y_r - \Gamma|$. Reference slopes are 1 for the first two controls and $1/2$ for the third. No exponent is fitted. For the three-step control, $e(r)/\sqrt{r}$ is compared with the analytic limit $\sqrt{15}/3825 \approx 0.001012545$.

Table 6: Complete rate experiment. The last column divides by the stated theoretical rate.

Control	r	Y_r	$e(r)$	$e(r)/r^\nu$
Distinct $K = 2, \nu = 1$	0.1	0.094134433	8.66×10^{-4}	0.0086556694
Distinct $K = 2, \nu = 1$	0.025	0.094881315	1.19×10^{-4}	0.0047473847
Distinct $K = 2, \nu = 1$	0.00625	0.094977608	2.24×10^{-5}	0.0035827771
Distinct $K = 2, \nu = 1$	0.0015625	0.094994877	5.12×10^{-6}	0.0032785835

Control	r	Y_r	$e(r)$	$e(r)/r^\nu$
Distinct $K = 2, \nu = 1$	0.000390625	0.094998749	1.25×10^{-6}	0.0032016971
Distinct $K = 2, \nu = 1$	9.76563×10^{-5}	0.094999689	3.11×10^{-7}	0.0031824227
Distinct $K = 2, \nu = 1$	2.44141×10^{-5}	0.094999922	7.76×10^{-8}	0.0031776009
Distinct $K = 2, \nu = 1$	6.10352×10^{-6}	0.094999981	1.94×10^{-8}	0.0031763952
Equal $K = 2, \nu = 1$	0.1	-0.000623631	1.37×10^{-6}	1.3693485×10^{-5}
Equal $K = 2, \nu = 1$	0.025	-0.000624656	3.44×10^{-7}	1.3762589×10^{-5}
Equal $K = 2, \nu = 1$	0.00625	-0.000624914	8.61×10^{-8}	1.3779929×10^{-5}
Equal $K = 2, \nu = 1$	0.0015625	-0.000624978	2.15×10^{-8}	1.3784268×10^{-5}
Equal $K = 2, \nu = 1$	0.000390625	-0.000624995	5.38×10^{-9}	1.3785353×10^{-5}
Equal $K = 2, \nu = 1$	9.76563×10^{-5}	-0.000624999	1.35×10^{-9}	1.3785624×10^{-5}
Equal $K = 2, \nu = 1$	2.44141×10^{-5}	-0.000625000	3.37×10^{-10}	1.3785692×10^{-5}
Equal $K = 2, \nu = 1$	6.10352×10^{-6}	-0.000625000	8.41×10^{-11}	1.3785709×10^{-5}
Equal $K = 3, \nu = 1/2$	0.1	-0.001825553	3.22×10^{-4}	0.0010171348
Equal $K = 3, \nu = 1/2$	0.025	-0.001665574	1.62×10^{-4}	0.0010224778
Equal $K = 3, \nu = 1/2$	0.00625	-0.001584524	8.06×10^{-5}	0.0010197479
Equal $K = 3, \nu = 1/2$	0.0015625	-0.001544097	4.02×10^{-5}	0.0010167432
Equal $K = 3, \nu = 1/2$	0.000390625	-0.001523963	2.01×10^{-5}	0.0010147975
Equal $K = 3, \nu = 1/2$	9.76563×10^{-5}	-0.001513924	1.00×10^{-5}	0.0010137099
Equal $K = 3, \nu = 1/2$	2.44141×10^{-5}	-0.001508912	5.01×10^{-6}	0.0010131371
Equal $K = 3, \nu = 1/2$	6.10352×10^{-6}	-0.001506409	2.50×10^{-6}	0.0010128433

I.4 COMMUTING-BENCHMARK ESTIMATION STUDY

The following study estimates Γ_{comm} from independent nearby refits and labelled tasks. Its prespecified settings, replication counts and adverse cells are reported in full. We report the root mean squared error (RMSE) of the estimated coefficient across replications. It does not estimate the finite-budget restriction term or singular cone coefficient.

The two-dimensional controls use $u \sim N(0, T)$ and independent additive noise with covariance N_r. For anisotropic controls, $T = \text{diag}(0.8, 0.2)$, $N_0 = \text{diag}(0.2, 0.8)$ and $W = I$. The nominal filter is $F_0 = I - (I - S_0)^4$ with $(S_0)_{ii} = 1 - (1 - f_i)^{1/4}$, $(f_1, f_2) = (0.8, 0.2)$. For a planar rotation $R(r)$, use

$$S_r = R(r) S_0 R(r)^\top, \qquad N_r = N_0 + r \begin{pmatrix} 0 & -a \\ -a & 0 \end{pmatrix}. \tag{I.2}$$

The choices $a = 0.2, 1$ give $\Gamma_{\text{comm}} = 0.72 - 1.2a = \pm .48$. The isotropic control has $T = N_0 = 0.5I$, no noise drift and $F_0 = 0.5I$. At $r = 0.0015625$, the exact commuting-minus-preserved risk differences divided by r^2 are 0.479999, 0 and -0.480001. Only the baseline curvature in the later gap sweep retains the same four-step representation of F_0.

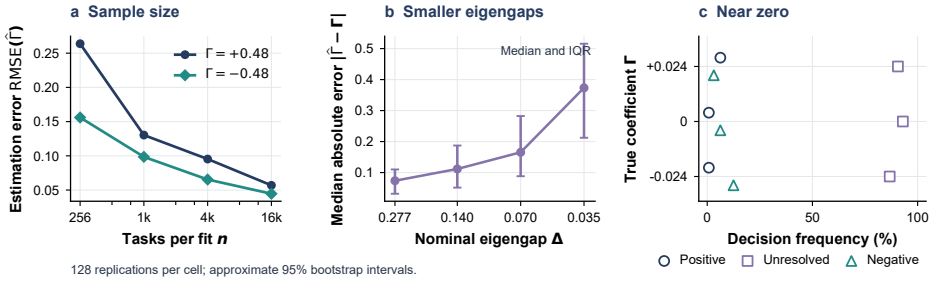


Figure 4: Commuting-benchmark coefficient experiment, with 128 independent replications per cell. (a) Estimation RMSE versus sample size. (b) Median absolute error and interquartile range (IQR) versus eigengap. (c) Approximate bootstrap abstention near zero; decision frequencies use 128 replications as the denominator. These 199-draw intervals fail at the adverse small-gap setting; the complete results follow.

Sampling and separation of fitting from evaluation. The Gaussian family is (I.2), with $W = I$ known and the target distribution fixed. At each of $-h, 0, +h$, one independent sample of n labelled pairs (u, z_r) estimates

$$\widehat{C}_r = n^{-1} \sum_{\ell} u_\ell z_{r,\ell}^\top, \qquad \widehat{\Sigma}_r = n^{-1} \sum_{\ell} z_{r,\ell} z_{r,\ell}^\top, \qquad \widetilde{F}_r^\star = \widehat{C}_r \widehat{\Sigma}_r^{-1}.$$

A separate independent sample $x_\ell \sim N(0, S_r)$ gives $\widehat{S}_r = n^{-1} \sum_\ell x_\ell x_\ell^\top$. The three labelled samples and three design samples are mutually independent. These labelled synthetic tasks identify the cross-moment; unlabelled operator estimates alone would not do so.

Nominal alignment uses $\widehat{F}_0 = \widehat{\Pi} \widetilde{F}_0^\star$. For $W = I$ and $\widehat{S}_0 = V \operatorname{diag}(\widehat{\lambda}_i) V^\top$, the weighted projection has coefficients

$$(\widehat{\Pi} A) = V \operatorname{diag}\left(\frac{(V^\top A \widehat{\Sigma}_0 V)_{ii}}{(V^\top \widehat{\Sigma}_0 V)_{ii}} \right) V^\top.$$

The derivative estimates use independent central differences with $h = n^{-1/4}$, and no coefficient clipping. The target is the derivative at zero; finite-bandwidth bias is present.

Experimental settings and interval construction. All 15 cells were fixed before the full sample was generated. Cells S1–S8 use $a = 0.2, 1.0$, each at $n = 256, 1024, 4096, 16384$, and baseline gap $\Delta_0 = 0.2770013040$. Cells G1–G4 use $n = 4096$, $a = 0.2$ and gaps $\Delta_0, 0.14, 0.07, 0.035$. The eigenvalue center, statistical moments and rotation speed stay fixed. The nominal filter remains $\operatorname{diag}(0.8, 0.2)$; only the baseline gap retains the same four-step representation $I - (I - S_0)^4$. Cells Z1–Z3 use $n = 16384$, baseline gap and $a = 0.58, 0.60, 0.62$, yielding $\Gamma = 0.024, 0, -0.024$.

Each cell has 128 independent replications. Within a replication, 199 parametric Gaussian bootstrap draws resample the six uncentered second-moment estimates through their Wishart distributions with n degrees of freedom, divided by n. For each labelled sample at $r \in \{-h, 0, h\}$, we resample the joint second-moment matrix of (u, z_r). We extract C_r^* and Σ_r^* from the same draw, preserving their joint dependence. The other three Wishart draws resample the design second moments for S_r. Each draw recomputes the complete aligned diagnostic. If q_ℓ is its empirical ℓ-quantile, the basic interval for Γ is $[2\widehat{\Gamma} - q_{0.975},\ 2\widehat{\Gamma} - q_{0.025}]$. A sign is declared only when this interval excludes zero. The direction-free screen uses

$$\kappa_L = (2\widehat{\kappa} - q^\kappa_{0.975})_+, \qquad d_U = (2\widehat{d} - q^d_{0.025})_+, \qquad \kappa_L > 2 d_U.$$

We assess the joint coverage of these approximate one-sided bounds empirically. The worst coefficient-interval coverage is 83/128 at G4. Its eight negative declarations are all wrong although 114/128 raw signs are positive. At G3 the two resolved declarations include one correct and one wrong sign. All failures remain in the tables and replication file.

Fresh evaluation and reference estimation. Every replication receives 32,768 independent new pairs (u, z_h). The population-optimal filter in $\mathcal{C}_h$ is evaluated against known F_0 and fitted $\widehat{F}_0$. It is computed from population moments solely as the comparison benchmark, independently of the diagnostic fit. We retain each paired empirical mean risk gap, its standard error, and its exact population counterpart. Positive gaps favor preservation. Direct sample losses and quadratic empirical-moment formulas agree within 10^{-12} in all 3,840 comparisons.

A fresh sign agreement compares $\operatorname{sign}(\widehat{\Gamma})$ with the empirical finite-h gap sign; it is not automatically a resolved risk decision. A separate column records how often the paired gap exceeds 1.96 estimated standard errors in magnitude. The near-zero cases seldom resolve even with the large fresh sample. At $\Gamma = 0$, binary coefficient sign accuracy is undefined; interval declarations and coverage remain meaningful. A nonzero finite-h sign can reflect higher-order terms.

The independent reference error $\epsilon = \|\widehat{F}_0 - F_0\|_0$ is recorded relative to h. Its median decreases with n in the separated baseline cases, while shrinking the gap at fixed n worsens it. This exposes the distinction between consistency of $\widehat{\Gamma}$ and the additional $\epsilon = o_p(h)$ condition for the learned-filter risk comparison.

Complete numerical summaries. Tables 7–8 include every prespecified cell. Percentages use 128 replications as denominator unless otherwise stated. For the eight S cells, raw sign accuracy is

128/128; the marginal descriptive Wilson interval is $[97.1\%, 100\%]$ per cell, without a simultaneous-coverage claim. All replication-level uncertainty and component estimates are supplied in the source CSV.

Table 7: Coefficient estimates and approximate interval decisions. Cov.: empirical coverage of the nominal 95% coefficient interval. Unres.: interval contains zero. Accuracy is undefined when the true coefficient is zero.

Cell	n	Δ	Γ	RMSE	Sign (%)	Cov. (%)	Unres. (%)
S1	256	0.277	+0.480	0.264	100.0	82.0	82.0
S2	1024	0.277	+0.480	0.130	100.0	93.0	3.9
S3	4096	0.277	+0.480	0.095	100.0	91.4	0.0
S4	16384	0.277	+0.480	0.057	100.0	95.3	0.0
S5	256	0.277	-0.480	0.156	100.0	92.2	3.1
S6	1024	0.277	-0.480	0.098	100.0	89.8	0.0
S7	4096	0.277	-0.480	0.065	100.0	91.4	0.0
S8	16384	0.277	-0.480	0.045	100.0	94.5	0.0
G1	4096	0.277	+0.480	0.099	100.0	90.6	0.0
G2	4096	0.140	+0.480	0.157	100.0	86.7	14.8
G3	4096	0.070	+0.480	0.295	97.7	81.2	98.4
G4	4096	0.035	+0.480	0.813	89.1	64.8	93.8
Z1	16384	0.277	+0.024	0.052	63.3	93.0	90.6
Z2	16384	0.277	+0.000	0.049	–	93.0	93.0
Z3	16384	0.277	-0.024	0.048	74.2	95.3	86.7

Table 8: Screen frequency, its empirical simultaneous bound coverage, and fresh-task risk agreement. Known/learned columns compare the raw diagnostic sign with empirical gaps for the respective references. Resolved is the fraction of learned-reference fresh gaps exceeding 1.96 standard errors. The screen is approximate.

Cell	Screen (%)	Joint cov. (%)	Known (%)	Learned (%)	Resolved (%)	Median ϵ/h
S1	39.1	92.2	100.0	100.0	100.0	0.161
S2	88.3	93.0	100.0	100.0	100.0	0.115
S3	100.0	93.8	100.0	100.0	100.0	0.082
S4	100.0	95.3	100.0	100.0	100.0	0.051
S5	0.0	99.2	100.0	100.0	100.0	0.150
S6	0.0	99.2	100.0	100.0	100.0	0.114
S7	0.0	96.9	100.0	100.0	100.0	0.079
S8	0.0	96.9	100.0	100.0	100.0	0.053
G1	100.0	93.0	100.0	100.0	100.0	0.066
G2	90.6	93.0	100.0	100.0	100.0	0.118
G3	35.9	94.5	97.7	94.5	98.4	0.235
G4	4.7	96.1	89.1	70.3	97.7	0.368
Z1	0.0	96.9	56.2	54.7	14.8	0.053
Z2	0.0	96.9	53.9	51.6	9.4	0.052
Z3	0.0	95.3	54.7	55.5	16.4	0.055

Decision directions and finite-bandwidth risks. Table 9 separates positive and negative declarations and records the finite-h population risk gaps. Accuracy conditional on acting uses the number of resolved decisions as denominator, and can be unstable when few decisions are issued. At zero, this accuracy is undefined. The learned-reference gap is averaged over the independently trained references.

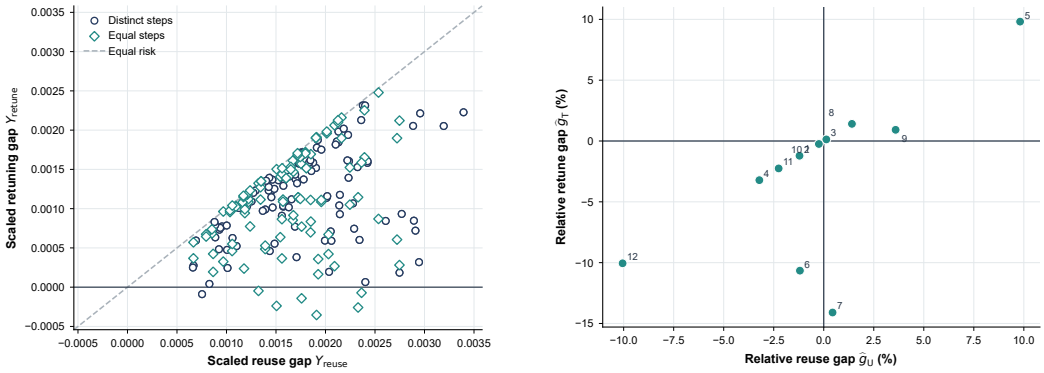
I.5 Three adaptation choices

For every prespecified population comparison, we rerun the nominal schedule on the perturbed operator. Writing $Y_{\mathrm{reuse}} = \{R_r(F_{\boldsymbol{\alpha}_0}(S_r)) - R_r(F_0)\}/r^2$ and $Y_{\mathrm{retune}} = \{\inf_{\boldsymbol{\alpha}} R_r(F_{\boldsymbol{\alpha}}(S_r)) - R_r(F_0)\}/r^2$, all 2,268 records satisfy $Y_{\mathrm{retune}} \le Y_{\mathrm{reuse}} + 10^{-10}$. Among the 324 signed comparisons from 162 instances at $|r| = 0.0015625$, the reuse gap is positive in 324/324 and the retuning gap in 316/324 (97.53%), and reuse is positive while retuning is negative in 8/324 (2.47%). These are

Table 9: All declared directions and population risk gaps. The last two columns divide the commuting-minus-reference risk by h^2; they are finite-bandwidth quantities, not exactly Γ. Declared accuracy is conditional on a nonzero interval decision.

Cell	Positive (%)	Negative (%)	Unres. (%)	Declared acc. (%)	Known/h^2	Learned/h^2
S1	18.0	0.0	82.0	100.0	+0.4540	+0.4173
S2	96.1	0.0	3.9	100.0	+0.4668	+0.4527
S3	100.0	0.0	0.0	100.0	+0.4734	+0.4634
S4	100.0	0.0	0.0	100.0	+0.4767	+0.4720
S5	0.0	96.9	3.1	100.0	-0.5114	-0.5257
S6	0.0	100.0	0.0	100.0	-0.4953	-0.5201
S7	0.0	100.0	0.0	100.0	-0.4876	-0.5013
S8	0.0	100.0	0.0	100.0	-0.4838	-0.4797
G1	100.0	0.0	0.0	100.0	+0.4734	+0.4677
G2	85.2	0.0	14.8	100.0	+0.4734	+0.4525
G3	0.8	0.8	98.4	50.0	+0.4734	+0.3685
G4	0.0	6.2	93.8	0.0	+0.4734	+0.1318
Z1	6.2	3.1	90.6	66.7	+0.0239	+0.0268
Z2	0.8	6.2	93.0	–	-0.0001	-0.0015
Z3	0.8	12.5	86.7	94.1	-0.0242	-0.0269

empirical fractions, not population probabilities or 324 independently generated instances. The median $Y_{\text{reuse}} - Y_{\text{retune}}$ is 2.90×10^{-4}.



(a) Prespecified population sweep

(b) Bike Sharing month shifts

Figure 5: Preservation, nominal-schedule reuse and fixed-budget retuning are distinct adaptation choices. Panel (a) uses the population-optimal retuning family on the prespecified synthetic sweep; its diagonal is the feasibility bound. Panel (b) uses the centered Bike Sharing sensitivity with a common target intercept and reuse included among target candidates.

For each calendar month, the source domain is the corresponding month in 2011 and the target domain is the same month in 2012. The response is $\log(1 + \text{count})$. The covariates cover calendar, weather, temperature, humidity, wind, and periodic hour, weekday, and month terms; the component counts `casual` and `registered` are excluded to prevent target leakage. Within each year-month domain, the first 60%, next 20% and final 20% in time are used for training, validation and testing.

There are 13 columns: holiday, working-day and weather-category codes; temperature, apparent temperature, humidity and wind; and sine/cosine pairs for hour, weekday and month. There is no one-hot expansion. The feature transform is fitted on source-training covariates and applied throughout.

Strictly constant source-training columns are centered at their observed constant value and assigned unit scale; other columns retain their source mean and population standard deviation (denominator n_s). No near-zero threshold is introduced: every nonconstant source column has scale well above machine precision. Constant columns therefore remain exactly zero after centering, without floating-point amplification.

We compare three experimental protocols: the original protocol, the same protocol with the nominal schedule added to the target candidates, and training-set centering with this augmented candidate family. Source validation selects an equal two-step nominal schedule; target validation selects from a 101×101 grid augmented by that complete nominal schedule. The original protocol uses no intercept; the centered sensitivity adds a shared target intercept outside the regularized coordinates.

For the third version, each domain uses only its own training means $\bar{x}_d$ and $\bar{y}_d$ of the transformed features and response. Define $X_{d,c} = X_d - \mathbf{1}\bar{x}_d^\top$ and $y_{d,c} = y_d - \bar{y}_d\mathbf{1}$; in the following formulas X_d, y_d denote these centered quantities. All methods predict

$$\widehat{y}_M(x) = \bar{y}_t + (x - \bar{x}_t)^\top \widehat{\beta}_M, \qquad M \in \{\text{preserve}, \text{reuse}, \text{retune}\}.$$

The intercept is shared and is not a schedule-regularized coordinate; validation and test observations never determine centering parameters. For $d \in \{s, t\}$, the training quantities are

$$H_d = X_d^\top X_d/n_d, \quad b_d = X_d^\top y_d/n_d, \quad S_d = H_d + \rho_d I, \quad z_d = S_d^{-1} b_d.$$

Here, $\rho_d = 10^{-6}$ if $\mathrm{cond}(H_d) > 10^8$ or a direct solve fails, and zero otherwise; the code uses a linear solve, not a pseudoinverse. The same S_d is used in the filters, corresponding to gradient descent on $\ell_d(\beta) = \|X_d\beta - y_d\|_2^2/(2n_d) + \rho_d\|\beta\|_2^2/2$, whose update is $\beta_{k+1} = \beta_k + \alpha_{k+1}(b_d - S_d\beta_k)$.

In the centered sensitivity, monthly source and target designs have numerical ranks 10, 11 in 13 dimensions; every monthly fit uses $\rho_d = 10^{-6}$. We use $\mathrm{cond}(S) = \|S\|_{\mathrm{op}}\|S^{-1}\|_{\mathrm{op}}$ for the two-norm condition number (infinite for a singular matrix), and double-precision machine epsilon $\epsilon_{\mathrm{mach}} = 2^{-52}$. The rank is computed from the standardized design X_d, after any domain centering and before forming H_d or adding ridge: singular values exceeding $\epsilon_{\mathrm{mach}} \max(n_d, 13)\sigma_{\max}(X_d)$ are counted. Both month columns are exactly zero after transformation; other constant or dependent columns can reduce the rank further. The all-year comparison has rank 13 in both domains and uses no ridge. For each domain, we report the minimum eigenvalue, condition number, and rank before and after stabilization. Source validation selects the equal schedule $\boldsymbol{\alpha}_0 = (a_0, a_0)$ and $F_0 = F_{\boldsymbol{\alpha}_0}(S_s)$. The three target predictors are

$$\widehat{\beta}_{\mathrm{preserve}} = F_0 z_t, \qquad \widehat{\beta}_{\mathrm{reuse}} = F_{\boldsymbol{\alpha}_0}(S_t) z_t, \qquad \widehat{\beta}_{\mathrm{retune}} = F_{\widehat{\boldsymbol{\alpha}}_t}(S_t) z_t,$$

where $\widehat{\boldsymbol{\alpha}}_t$ is selected from the target grid and the added nominal schedule. This diagnostic compares filters applied to the same target-domain input z_t. The update budget counts the K gradient updates defining the current schedule family; construction of z_t and validation search are excluded, so it is not an end-to-end runtime benchmark. Preservation keeps the nominal filter, not the source-domain fitted coefficient vector. Indeed, if $F_\alpha(S) = Sq_\alpha(S)$, then $F_\alpha(S_t)z_t = q_\alpha(S_t)b_t$, whereas $F_0 S_t^{-1} b_t$ generally has no such cancellation.

Validation grids and boundary selections. For each pair, set $a_{\max} = 0.95/\max\{\lambda_{\max}(S_s), \lambda_{\max}(S_t)\}$ using the training covariates from both domains. This defines a paired-domain diagnostic rather than a source-only schedule-transfer experiment.

The source grid contains 200 equally spaced absolute learning rates on $[0.02a_{\max}, a_{\max}]$ for an equal-step schedule; the target grid contains 101 equally spaced values on the same interval for each coordinate.

Let $\mathcal{G}_{101} = \{0.02a_{\max} + j(0.98a_{\max})/100 : j = 0, \ldots, 100\}$ be the one-coordinate target grid; $\mathcal{G}_{101}^2$ denotes its Cartesian square. The original implementation retains the first grid minimum, while the candidate-only and centered versions use $\mathcal{G}_{101}^2 \cup \{\boldsymbol{\alpha}_0\}$. If reuse loss is at most the grid minimum plus $10^{-12} + 10^{-12}\max\{1, |L_{\mathrm{val}}(\boldsymbol{\alpha}_0)|, |L_{\mathrm{val}}^{\mathrm{grid}}|\}$, reuse is selected; otherwise the grid minimum is selected. Thus selected validation loss never exceeds reuse loss beyond this fixed tolerance. This nesting does not imply a test-risk ordering.

In the centered sensitivity, both source and target select the common upper endpoint in months 1, 2, 3, 4, 5, 8, 10, 11, 12; the remaining 3 months use different schedules. Target selections lie on a box boundary in 11 months. Supplementary data give the matrices, grid endpoints, schedules, and validation and test losses for all three protocols.

The original protocol is reproduced to a maximum stored-summary discrepancy of 3.6×10^{-15}; adding the reuse candidate alone leaves its selected predictors unchanged. It favors retuning over

preservation in 9 months and preservation in 3, with median absolute relative gap 17.0%. After domain centering and a shared intercept, these counts become 8 and 4, and the median absolute relative gap is 1.83%. Relative to reuse, centered retuning has lower test loss in 3 months, higher loss in 0, and equal loss in 9. This sensitivity identifies the contribution of mean handling; we report results for all three protocols.

For each month, let ℓ_i^{P} and ℓ_i^{T} denote the held-out squared losses of preservation and retuning. Write $\widehat{R}_{\mathrm{P}} = n^{-1}\sum_i \ell_i^{\mathrm{P}}$ and $\widehat{\Delta} = n^{-1}\sum_i(\ell_i^{\mathrm{T}} - \ell_i^{\mathrm{P}})$, where n is the number of held-out hours. Paired circular block resampling uses the same hourly indices for both losses, producing mean differences $\widehat{\Delta}^{*(b)}$ over 2,000 draws with 24-hour blocks. If q_a^* denotes their empirical a-quantile, the displayed point and interval are

$$\widehat{g}_{\mathrm{T}} = 100\frac{\widehat{\Delta}}{\widehat{R}_{\mathrm{P}}}, \qquad I_{\mathrm{display}} = \frac{100}{\widehat{R}_{\mathrm{P}}}[q^*_{0.025}, q^*_{0.975}].$$

Quantiles use linear interpolation (`numpy.quantile`). Each version uses its own original-sample preservation MSE as a denominator, fixed across bootstrap draws. The interval therefore measures uncertainty in the absolute risk gap on a fixed normalized scale, rather than a percentile interval for a resampled ratio; it may extend below -100%.

For reuse, $\widehat{g}_{\mathrm{U}}$ replaces the retuning losses, and figures use $\widehat{g}_{\mathrm{M}}$ for $\mathrm{M} \in \{\mathrm{U}, \mathrm{T}\}$. The 48-hour sensitivity uses the same construction without refitting or retuning. These intervals are conditional on the fitted schedules and are not simultaneous across months.

For the centered sensitivity, both the 24-hour and 48-hour intervals exclude zero in months 4, 7, and 12. All versions use the same saved resampling starts and their own fitted losses and fixed preservation denominator. Supplementary data also include the full-year comparison and all selected schedules.

Table 10: Bike Sharing month-shift test risks for the centered sensitivity. The final column displays a nominal 95% paired circular block-bootstrap confidence interval (CI) for the absolute retune-minus-preserve MSE gap, divided by the centered version's original-sample preservation MSE (in percent), held fixed across bootstrap draws; block length is 24 hours with 2,000 resamples.

Month	Preserve	Reuse	Retune	T–P (%)	95% CI (%)
1	1.014758	1.002630	1.002630	-1.2	[-3.1, +0.8]
2	1.155841	1.141833	1.141833	-1.2	[-3.4, +1.1]
3	1.235272	1.236873	1.236873	+0.1	[-3.0, +2.7]
4	0.990111	0.958251	0.958251	-3.2	[-5.6, -0.8]
5	1.056261	1.159917	1.159917	+9.8	[-0.4, +21.4]
6	1.303713	1.288199	1.164771	-10.7	[-26.8, +2.9]
7	1.132348	1.137336	0.972705	-14.1	[-22.0, -6.4]
8	0.981333	0.995161	0.995161	+1.4	[-4.0, +7.1]
9	0.874631	0.906046	0.882672	+0.9	[-2.8, +4.2]
10	0.786590	0.784674	0.784674	-0.2	[-3.4, +3.2]
11	1.159580	1.133441	1.133441	-2.3	[-7.4, +2.4]
12	1.757292	1.580628	1.580628	-10.1	[-21.0, -1.4]

Table 11: Monthly design-matrix properties for the centered sensitivity. Ranks refer to standardized, domain-centered training designs, before ridge; all have 13 columns and use $\rho_d = 10^{-6}$. The last column is $\max\{\mathrm{cond}(S_s), \mathrm{cond}(S_t)\}$, and the grid lower endpoint is $0.02a_{\max}$. Supplementary data give the full matrices and selected schedules.

Month	Source rank	Target rank	$a_{\max}$	Max. stabilized condition
1	11	11	0.221335	4.29×10^6
2	10	10	0.301393	3.15×10^6
3	10	10	0.280254	3.39×10^6
4	11	11	0.291705	3.26×10^6
5	10	10	0.316352	3×10^6
6	10	10	0.317554	2.99×10^6
7	11	11	0.228912	4.15×10^6
8	10	10	0.199989	4.75×10^6
9	11	11	0.308535	3.08×10^6
10	11	11	0.282293	3.37×10^6
11	11	11	0.319207	2.98×10^6
12	10	10	0.295400	3.22×10^6

I.6 Near-equal finite-radius diagnostic

We examine finite-radius behavior as the nominal step half-gap and task radius vary. For this diagnostic, we use instance 127 from the population sweep, with $a = 0.7$, retuning box $\mathcal{A} = [0.1, 1.2]^2$, and fixed matrices

$$S_0 = \begin{bmatrix} 0.57738994 & -0.11251870 & 0.12411033 \\ -0.11251870 & 0.22463203 & -0.06934887 \\ 0.12411033 & -0.06934887 & 0.55978492 \end{bmatrix},$$

$$E = \begin{bmatrix} 0.00430056 & 0.02290907 & -0.00479506 \\ 0.02290907 & 0.01715514 & 0.00660038 \\ -0.00479506 & 0.00660038 & 0.01021960 \end{bmatrix},$$

$$N' = \begin{bmatrix} -0.00710477 & -0.01323135 & 0.00026810 \\ -0.01323135 & -0.01738365 & -0.00598961 \\ 0.00026810 & -0.00598961 & -0.01119371 \end{bmatrix}.$$

The displayed matrices are rounded; the supplementary data give full-precision S_0, E, N'. For each half-gap define

$$\boldsymbol{\alpha}_{0,\delta} = (a - \delta, a + \delta), \quad F_{0,\delta} = F_{\boldsymbol{\alpha}_{0,\delta}}(S_0) = F_{(a,a)}(S_0) + \delta^2 S_0^2.$$

The code keeps N' fixed across half-gaps and uses the model

$$u_\delta \sim N(0, F_{0,\delta}), \quad z_{r,\delta} = u_\delta + \xi_{r,\delta}, \quad \xi_{r,\delta} \sim N(0, I - F_{0,\delta} + rN'),$$

with independent signal and noise and $S_r = S_0 + rE$. For each fixed δ, the target distribution is invariant along r; it differs between half-gap curves. Consequently

$$\Sigma_{r,\delta} = I + rN', \quad F^\star_{r,\delta} = F_{0,\delta}(I + rN')^{-1}, \quad D_\delta = -F_{0,\delta}N'.$$

Thus D_δ generally varies with δ and can be nonsymmetric. The exact preserved reference and risk comparison are

$$\Delta R_{r,\delta} = \inf_{\boldsymbol{\alpha} \in \mathcal{A}} \mathcal{R}_{r,\delta}(F_{\boldsymbol{\alpha}}(S_r)) - \mathcal{R}_{r,\delta}(F_{0,\delta}).$$

Each curve holds δ fixed and varies r on one side. The original diagnostic used $r > 0$; this extension adds the same radius magnitudes with $r < 0$. The ten half-gaps are $\{0, 10^{-5}, 3 \times 10^{-5}, 10^{-4}, 3 \times 10^{-4}, 10^{-3}, 3 \times 10^{-3}, 10^{-2}, 3 \times 10^{-2}, 10^{-1}\}$; the nine radii are $\{10^{-5}, 3 \times 10^{-5}, 10^{-4}, 3 \times 10^{-4}, 10^{-3}, 3 \times 10^{-3}, 10^{-2}, 3 \times 10^{-2}, 10^{-1}\}$. All 180 signed cells use the complete two-step solver; the scaled risks in the 90 positive cells agree with the positive-side calculation to within 5.4×10^{-13}. Step-box feasibility, positive curvature, positive noise covariance and positive residual factors are checked at each cell.

Two distinct approximation references. Let $G_{2,\delta}$ and $A_{1,\delta}, A_{2,\delta}$ be the operator and step derivatives at $\boldsymbol{\alpha}_{0,\delta}$, as in (5) and (6). For $\delta > 0$, the distinct-step reference is

$$\Gamma_{\mathrm{reg}}(\delta) = \mathrm{dist}_0^2(D_\delta, G_{2,\delta} + \mathrm{span}\{A_{1,\delta}, A_{2,\delta}\}) - \|D_\delta\|_0^2.$$

For numerical stability at $\delta > 0$, its correction space is evaluated in the basis $\{S_0, S_0^2\}$. At $\delta = 0$, the separate benchmark is $\Gamma_{\mathrm{Jac}}(0)$ from Theorem 4.3. The equal-step reference is the auxiliary benchmark

$$\Gamma_{\mathrm{eq},s}(\delta) = \mathrm{dist}_0^2\big(D_\delta, G_{2,0} + \mathrm{span}\{A_{2,0}^{\mathrm{eq}}\} + s\mathbb{R}_+ H_{2,0}\big) - \|D_\delta\|_0^2,$$

where $s = \pm 1$, $A_{2,0}^{\mathrm{eq}} = S_0(I - aS_0)$ and $H_{2,0} = S_0^2$. Here, the code holds the equal-step geometry at (a, a) but substitutes the changed drift D_δ. For $\delta > 0$ this remains an auxiliary equal-geometry reference, not the near-equal coefficient in Proposition 4.2, which also depends on $\delta_r^2/|r|$. At $\delta = 0$ it equals the established equal-step coefficient on side s. Both distances use $W = \Sigma_{0,\delta} = I$.

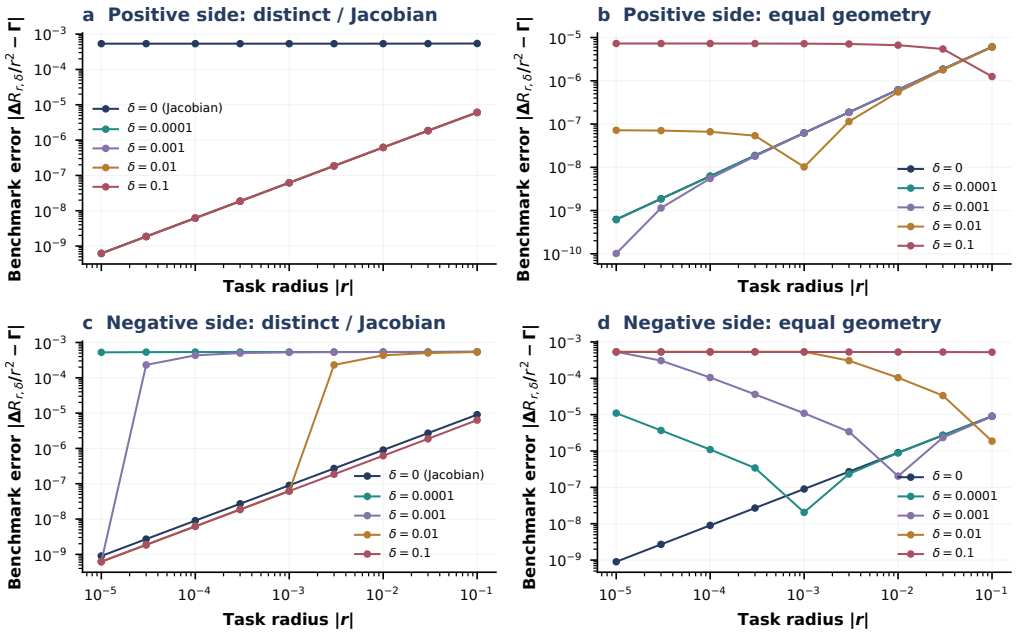


Figure 6: Near-equal sensitivity on both sides of the task path: positive side (top) and negative side (bottom). Left panels use $\Gamma_{\mathrm{reg}}(\delta)$ only for $\delta > 0$, and $\Gamma_{\mathrm{Jac}}(0)$ for equal steps. Right panels use $\Gamma_{\mathrm{eq},s}(\delta)$ with fixed equal geometry and drift D_δ. Five of ten half-gaps are shown; the numerical results cover all 180 cells.

Direction and constraint checks. Full-precision instance 127 gives $t_0 = \langle(\mathrm{Id} - P)H_{2,0}, D_0 - G_{2,0}\rangle_0 = 0.00553032466632$ and

$$\Gamma_{\mathrm{Jac}}(0) = 0.000996386484164,$$
$$\Gamma_{\mathrm{eq},+}(0) = 0.000460076232535, \qquad \Gamma_{\mathrm{eq},-}(0) = 0.000996386484164.$$

Thus the positive ray contains the two-direction affine optimum; the negative ray does not. All 90 positive-side optima are interior in coefficient space; the negative side has 67 equal-step boundary optima and 23 interior optima. The latter supplies direct evidence of changing constraint activity across the finite-radius grid, whereas the positive-side errors alone do not establish that mechanism. For optimized steps $\alpha_1 \le \alpha_2$, define the common displacement $h = (\alpha_1 + \alpha_2)/2 - a$ and the nonnegative half-split $v = (\alpha_2 - \alpha_1)/2$. Thus $(\alpha_1, \alpha_2) = (a + h - v, a + h + v)$, and the implementation checks

$$F_{(a+h-v,a+h+v)}(S_0) - F_{(a-\delta,a+\delta)}(S_0) = 2hA_{2,0}^{\mathrm{eq}} + (v^2 - \delta^2 - h^2)H_{2,0}.$$

The maximum Frobenius residual is 2.8×10^{-16} over all 180 cells. Figure 7 reports constraint activity, the normalized optimal split, and the scaled risk gap against $\delta^2/|r|$.

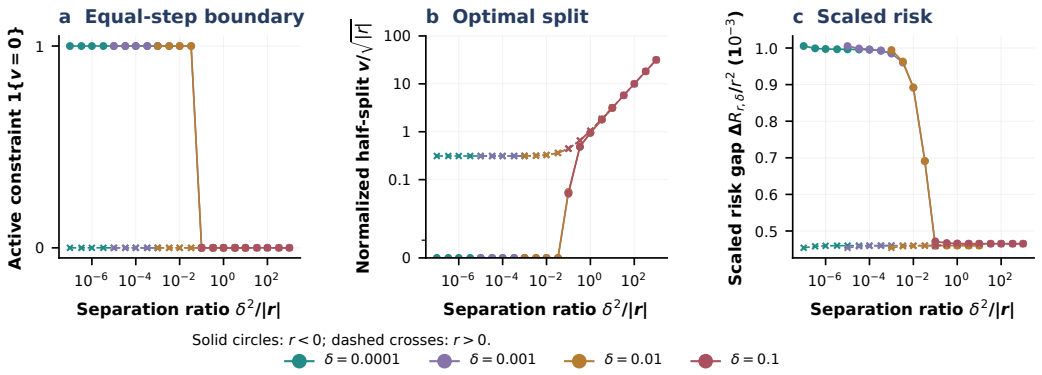


Figure 7: Signed sensitivity of the fixed near-equal instance. Four positive half-gaps are shown. (a) Indicator of the equal-step boundary. (b) Optimal half-split normalized by $\sqrt{|r|}$. (c) Scaled risk difference. Solid circles use $r < 0$; dashed crosses use $r > 0$. Colors distinguish half-gaps; $\delta = 0$ is excluded from the logarithmic ratio axis.

I.7 Noncommuting weighted-metric controls

These two additional deterministic controls were fixed before evaluation, without screening coefficient signs. Both have $p = 3$, $K = 2$, $S_0 = \mathrm{diag}(0.2, 0.45, 0.7)$ and step box $[0.1, 1.2]^2$. The distinct schedule is $(0.4, 0.95)$ and the equal schedule is $(0.7, 0.7)$. We use

$$W = \begin{bmatrix} 1.3 & 0.25 & -0.1 \\ 0.25 & 0.9 & 0.15 \\ -0.1 & 0.15 & 1.1 \end{bmatrix}, \quad \Sigma_0 = \begin{bmatrix} 1 & 0.2 & 0.1 \\ 0.2 & 1.4 & -0.15 \\ 0.1 & -0.15 & 0.8 \end{bmatrix},$$

$$E = \begin{bmatrix} 0.012 & 0.016 & -0.008 \\ 0.016 & -0.010 & 0.007 \\ -0.008 & 0.007 & 0.015 \end{bmatrix}, \quad D = \begin{bmatrix} 0.006 & -0.008 & 0.003 \\ 0.004 & -0.005 & 0.007 \\ -0.002 & 0.006 & 0.009 \end{bmatrix}.$$

Both weights are positive definite and neither commutes with S_0. For each nominal filter F_0, take $z_0 \sim N(0, \Sigma_0)$, $u = F_0 z_0 + \varepsilon$ with independent $\varepsilon \sim N(0, I)$, and $z_r = (F_0 + rD)^{-1} F_0 z_0$ as in Appendix B.4. This keeps the target distribution fixed and gives $F_r^\star = F_0 + rD$ exactly. The grid is $r = \pm 0.1 \times 2^{-j}$, $j = 0, \ldots, 7$, totaling 32 cells. We minimize the full matrix quadratic $\|c_1 S_r - c_2 S_r^2 - F_r^\star\|_r^2$ using its 2×2 coefficient Gram matrix and exhaustive two-step boundary enumeration. Scalar spectral weights are not substituted for this quadratic. The distinct coefficient is 0.001170413074, with commuting coefficient 0.001064689773 and restriction cost 0.000105723301. The equal coefficients are 0.001113196308 on the positive side and 0.001213195209 on the negative side; direct weighted ray projection agrees with the closed formula. The hierarchy identity residual is below 7×10^{-20}, and the maximum coefficient error at the smallest radius is 1.90×10^{-8}. All curvature and input second-moment matrices remain positive definite. These controls verify the weighted implementation.

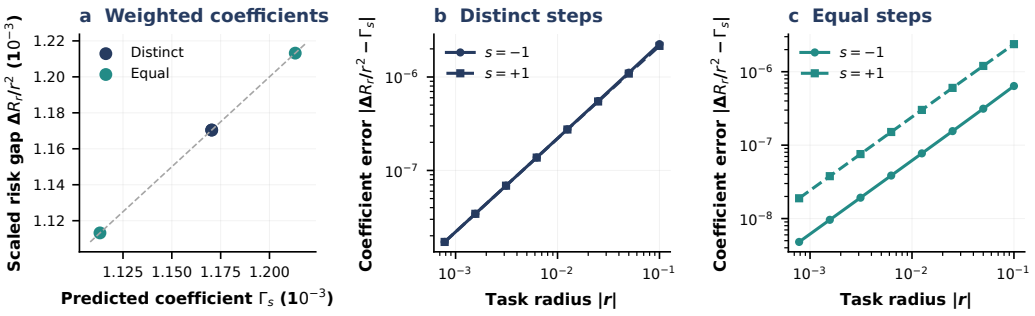


Figure 8: Fixed controls for noncommuting statistical weights. (a) Predictions versus scaled risks at $|r| = 0.00078125$. (b,c) Finite-radius coefficient errors for distinct and equal steps, respectively, on both sides. All 32 cells use the full matrix Gram objective.

I.8 Bike Sharing sensitivity, reproducibility, and scope

Figure 9 compares all three protocols. Main Figure 3(c), Figure 5(b), and Tables 10–11 consistently use the centered sensitivity. Supplementary data give the point estimates, intervals, and selected schedules for the original and candidate-only protocols. Here, P, U and T denote preservation, reuse and validation-selected retuning. Validation nesting is verified within the fixed tolerance in the candidate-only and centered protocols; no test-risk ordering is imposed.

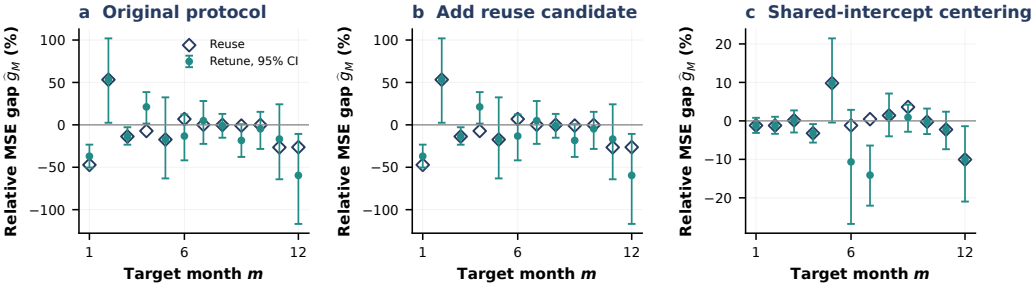


Figure 9: Implementation sensitivities, with the original protocol reproduced in (a). (b) Adding the complete reuse schedule to target candidates leaves the selected predictors unchanged. (c) Domain training-set centering with a shared target intercept substantially reduces relative gaps; its vertical scale differs. Points and 24-hour block-bootstrap intervals come from the same version. Each version uses its own original-sample preservation MSE as a denominator, fixed across bootstrap draws; resampling starts are shared across versions. MSE denotes mean squared error and CI denotes confidence interval.

Table 12: All three Bike Sharing versions. Counts compare held-out MSE across twelve fixed months. The final column is the median absolute retune-minus-preserve gap divided by the corresponding version's original-sample preservation MSE.

Version	T < P	T > P	T = U	Median gap
Original	9	3	3	17.05%
Add reuse	9	3	3	17.05%
Centered + reuse	8	4	9	1.83%

Reproducibility and scope. For the separate commuting estimation study only, seed 20260916 indexes independent cell/replication/role streams; it is distinct from population-sweep seed 2026091707 and Bike Sharing bootstrap seed 20260919. The rate controls are deterministic, and the near-equal diagnostic reuses instance 127 without a new random seed. The estimation study covers all 15 cells and seven radii, with 1,920 replications in total. It uses 382,080 bootstrap draws and 62,914,560 fresh task pairs. The finite-perturbation controls use a separate grid of eight positive radii. We chose the root mean squared error (RMSE) display in Figure 4(a) after observing 128/128 sign agreement in

the separated cases; the statistical settings were prespecified. The commuting-benchmark estimation study uses labelled two-dimensional Gaussian tasks with independent fits.

Reproducing the population comparisons. To verify numerical reproducibility, we repeat the 162-instance sweep with the same seed and recompute the nominal coefficients and current global optima. The three rate controls are deterministic, and all figures use the resulting numerical values.

The 324 smallest-radius values are shown directly rather than summarized by a fitted correlation. The full-radius table, solver status and individual-instance table distinguish the asymptotic comparison from behavior at a larger perturbation.

The population-validation calculations use low-dimensional fixed-budget models. The global solver exploits the polynomial structure of two and three steps. These experiments validate Γ_s on the specified instances.