MAINTAINING THE RIGHT ABSTRACTION: A FINITE-SAMPLE THEORY OF ADAPTIVE LOW-RANK REPRESENTATIONS UNDER DRIFT

Anonymous authors
Paper under double-blind review

ABSTRACT

When the dimension and orientation of a representation are learned simultaneously under distribution drift, their errors form a feedback loop. Off-equilibrium subspace error perturbs the statistic used to select rank, while each rank change moves the learner to a new Grassmann manifold and creates new orientation error. We develop a finite-sample theory that closes this loop. For a linear Gaussian channel, a single spectral benefit function determines the globally optimal rank and orientation, and its Riemannian Hessian determines the local recovery rate. Our main theorem for blockwise empirical updates gives an explicit invariant tracking radius, margins that prevent erroneous rank changes, re-entry guarantees after every accepted birth or death, and a finite recovery time while orientation remains out of equilibrium. At spectral crossings, we bound detection delay. Near a selection boundary, a two-point lower bound identifies an unavoidable $n^{-1/2}$ ambiguity in dimension. Experiments follow the structure of the theory by testing the spectral selection boundary, curvature-controlled recovery, hybrid reorganization, and low-rank scaling; scientific snapshots and a gated MNIST relaxation place the same mechanisms in broader representation settings. The resulting picture is a temporarily stable projection between excessive complexity and an evolving, noisy world.

1 INTRODUCTION

We ask when a data-driven low-rank representation can maintain its dimension and orientation under finite-sample distribution drift. Orientation error changes the rank statistic, while a rank change places the learner on a different Grassmann manifold and introduces new orientation error. The individual operations are understood: fixed-rank trackers follow orientation under prescribed capacity (Oja, 1982; Balzano et al., 2010; Chi et al., 2013); static rank selection chooses capacity from a batch (Minka, 2000; Gavish & Donoho, 2014; Perry & Wolfe, 2010); and rank-adaptive matrix optimization changes rank during optimization (Rabideau, 1996; Zhou et al., 2016). Their composition requires an invariant that survives the feedback: tracking must remain accurate enough for rank decisions to be reliable, and every accepted rank change must enter a basin from which tracking can recover.

A representation is an intervention on complexity. It replaces an unmanageably rich observation by a projection that is small enough to operate on and rich enough to remain useful. Under drift, the relevant object is a projection whose relation to the data can be *maintained*. Maintenance couples the choice of dimension, the orientation of the corresponding subspace, and the evidence required for reorganization.

We develop *adaptive low-rank representation dynamics* (ALRD). In a general representation family, its coupled state consists of dimension r, orientation $[V] \in \mathrm{Gr}(r, p)$, and remaining within-rank features $\vartheta \in \Theta_r$. These states form the rank-stratified space

$$\mathcal{M} = \bigsqcup_{r=0}^{p} \big(\mathrm{Gr}(r, p) \times \Theta_r\big).$$

Within a component, a Riemannian step rotates the representation. Between components, birth and death maps add or remove one direction. Hysteresis separates statistical fluctuation from structural change.

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.

To determine when this feedback can remain stable, we study a linear Gaussian channel. Here Θ_r is absent and the state reduces to $\bigsqcup_r \mathrm{Gr}(r,p)$. A scalar spectral benefit selects the static optimum. The Riemannian Hessian determines how quickly its orientation can be restored. The main theorem then shows that empirical block updates maintain a moving subspace inside an explicit tube; inside that tube, rank scores remain correct; after a rank jump, spectral perturbation puts the new state inside the next tube. Thus the output of each part of the argument supplies the hypothesis needed by the next.

Our contributions are:

- We prove a finite-sample maintenance theorem across rank strata for the implemented blockwise polar update. It includes covariance drift, finite samples, nonzero orientation error at rank checks, hysteretic birth/death decisions, adjacent-stratum re-entry, finite recovery time, endpoint ranks, and finite eigensolver accuracy.
- We derive an exact variational selection law for both rank and orientation. Every covariance mode has a signed benefit $\Delta_i = b(\lambda_i) - \gamma$, and the complete Grassmann Hessian gives the restoring curvature of every active–inactive rotation.
- We characterize both kinds of failure. A transversal spectral crossing produces a controlled detection delay, while a two-point lower bound identifies an $n^{-1/2}$ neighborhood in which no estimator can reliably maintain the correct dimension.

The experiments follow the same sequence: Gaussian streams expose the spectral and geometric mechanisms, scientific snapshots test the adaptive spectral rule, and gated autoencoders trace reconstruction across latent dimensions.

Related work. Prescribed-rank representation learning includes PCA and probabilistic PCA (Eckart & Young, 1936; Tipping & Bishop, 1999), while the information bottleneck and its variational forms trade retained information against compression in a fixed latent architecture (Tishby et al., 1999; Alemi et al., 2017). Riemannian optimization supplies the geometry for fixed-rank matrix problems (Edelman et al., 1998; Absil et al., 2008). Online PCA and subspace tracking adapt orientation, and classical rank-adaptive trackers or Riemannian methods also alter order. Prior guarantees do not close the feedback whereby off-equilibrium subspace error perturbs the rank statistic, while the resulting rank change alters the manifold on which subsequent tracking must re-enter. ALRD controls this feedback through a joint invariant that couples empirical orientation tracking, spectral rank decisions, and re-entry across adjacent rank strata. The flow and jump construction is a hybrid system in the standard sense (Goebel et al., 2012).

2 A RANK-PRICED REPRESENTATION CHANNEL

Let $X_k \sim \mathcal{N}(0, \Sigma_k)$ in $\mathbb{R}^p$, and write the eigenvalues of $\Sigma_k \succeq 0$ as $\lambda_1(k) \geq \cdots \geq \lambda_p(k)$. At rank r, an orthonormal matrix $V \in \mathrm{St}(p,r)$ encodes

$$Z = V^\top X + \xi, \qquad \xi \sim \mathcal{N}(0, sI_r), \quad s > 0,$$

and the decoder is the population MMSE estimator. We price reconstruction distortion, representation information, and integer capacity by

$$\mathbb{E}\,\|X - \mathbb{E}[X \mid Z]\|^2 + \gamma r - \beta I(X;Z), \qquad \beta, \gamma \geq 0.$$

Eliminating the decoder gives the Grassmann-invariant energy

$$\begin{aligned}\mathcal{E}_{r,\Sigma}(V) = {} & \operatorname{tr}\Sigma - \operatorname{tr}\big[(V^\top\Sigma^2 V)(V^\top\Sigma V + sI)^{-1}\big] \\ & - \frac{\beta}{2}\log\det(I + s^{-1}V^\top\Sigma V) + \gamma r. \end{aligned} \tag{1}$$

Nondegenerate encoder noise makes the information finite and keeps all resolvents uniformly regular. The integer price γr prevents the trivial full-dimensional solution.

The noise and the hybrid state resolve two features of the objective. For a continuously distributed X, a nonconstant deterministic encoder typically gives infinite $I(X;Z)$. Nondegenerate noise yields a finite information term. Rank is integer valued, so an expression such as $\dot{r} = -\partial_r \mathcal{E}$ has no intrinsic meaning. Birth and death maps place dimension changes between adjacent strata.

Hybrid state. For fixed r, the state is the equivalence class represented by $P = VV^\top$. A birth appends a unit direction $u \in \mathrm{range}(P)^\perp$; a death removes the weakest active Ritz line. These are incidence maps $\mathrm{Gr}(r,p) \leftrightarrow \mathrm{Gr}(r \pm 1, p)$ between manifolds. ALRD alternates these discrete maps with tangent orientation steps on the current component of $\mathcal{M}$.

3 What must be maintained

Define the benefit of retaining a covariance mode:

$$b(\lambda) = \frac{\lambda^2}{\lambda + s} + \frac{\beta}{2}\log(1 + \lambda/s), \qquad \Delta_i = b(\lambda_i) - \gamma. \tag{2}$$

The first term is the reduction in MMSE distortion; the second is the information reward. Since b is strictly increasing on $(0,\infty)$, the signs of the ordered Δ_i's change at most once.

Theorem 3.1 (Spectral variational law)**.** *For fixed r, every global minimizer of equation 1 spans a leading r-dimensional invariant subspace of Σ, with the usual freedom inside repeated eigenspaces, and*

$$\mathcal{E}_r^* = \mathrm{tr}\,\Sigma - \sum_{i=1}^{r} b(\lambda_i) + \gamma r.$$

Across all ranks, the smallest globally optimal rank is

$$r^* = \#\{i : b(\lambda_i) > \gamma\}.$$

Modes satisfying $b(\lambda_i) = \gamma$ are indifferent.

The theorem turns a model-selection problem into a local law: a dimension exists precisely when its benefit exceeds its price. In the limit $\beta = 0$, $s \downarrow 0$, this reduces to penalized Eckart–Young approximation. Here it also determines where a representation must reorganize: the rank boundary for mode i is $b(\lambda_i) = \gamma$.

Selection raises the next question: whether the projection returns to the selected eigenspace after perturbation. Let P_r^* denote the leading rank-r projector. A tangent direction at P_r^* rotates an active eigenvector $i \le r$ toward an inactive eigenvector $j > r$. The complete Riemannian Hessian is diagonal in these elementary rotations.

Theorem 3.2 (Restoring curvature)**.** *If $\lambda_r > \lambda_{r+1}$, the Hessian eigenvalue associated with the (i,j) active–inactive rotation is*

$$k_{ij} = \frac{2(\lambda_i - \lambda_j)[\lambda_i\lambda_j + s(\lambda_i + \lambda_j)]}{(\lambda_i + s)^2} + \frac{\beta(\lambda_i - \lambda_j)}{\lambda_i + s} > 0. \tag{3}$$

Hence P_r^* is an isolated local minimum and Grassmann gradient flow converges locally at a rate controlled by $\min_{i \le r < j} k_{ij}$.

The boundary eigengap determines identifiability. The slowest restoring mode may involve another active and inactive pair, so maintenance depends on the full spectrum together with the information weight and channel noise.

4 How the projection is maintained

At block k, let $\widehat{\Sigma}_k$ be the covariance of n fresh observations. With $P = VV^\top$, $Q = I - P$, define

$$\widehat{\Delta}_r^+ = b\Big(\lambda_{\max}(Q\widehat{\Sigma}_k Q)\Big) - \gamma, \qquad r < p, \tag{4}$$

$$\widehat{\Delta}_r^- = b\Big(\lambda_{\min}(V^\top\widehat{\Sigma}_k V)\Big) - \gamma, \qquad r > 0. \tag{5}$$

A birth is allowed when $\widehat{\Delta}_r^+ > h$; a death is allowed when $\widehat{\Delta}_r^- < -h$. If neither event occurs, rank is held fixed. After a jump, rank checks are paused for a guard interval while orientation updates continue. The empirical orientation map is

$$\widehat{\mathcal{T}}_{r,k}(V) = \mathrm{polar}\Big(V - \eta\ \mathrm{grad}\,\mathcal{E}_{r,\widehat{\Sigma}_k}(V)\Big). \tag{6}$$

Algorithm 1 Blockwise adaptive low-rank representation dynamics

Require: blocks X_k, (r, V), step η, threshold h, check period K, guards $\{G_q\}$
$g \leftarrow 0$
for $k = 0, 1, \ldots$ **do**
  form $\widehat{\Sigma}_k$; set $P = VV^\top, Q = I - P$
  if $g = 0$ and $k \equiv 0 \pmod K$ **then**
    compute available benefits $s^+ = \widehat{\Delta}_r^+$, $s^- = -\widehat{\Delta}_r^-$
    if $\max(s^+, s^-) > h$ **then**
      birth along the leading residual Ritz line if $s^+ \geq s^-$; otherwise remove the weakest active Ritz line
      update r; $g \leftarrow G_r$
    end if
  end if
  $V \leftarrow \widehat{\mathcal{T}}_{r,k}(V)$; $g \leftarrow \max(g - 1, 0)$
end for

QR orthogonalization gives the same subspace.

The hysteresis threshold also supplies a pathwise switching certificate. In an autonomous population problem, if every accepted jump decreases the true energy by at least $h_{\mathrm{dec}} > 0$ and $\mathcal{E} \geq \mathcal{E}_{\min}$, then

$$N_{\mathrm{jump}} \leq \left\lfloor \frac{\mathcal{E}_0 - \mathcal{E}_{\min}}{h_{\mathrm{dec}}} \right\rfloor. \tag{7}$$

Thus rank adaptation cannot chatter indefinitely. If an estimated jump increment has error at most δ_J, accepting only increments below $-h$ gives $h_{\mathrm{dec}} = h - \delta_J > 0$. Under drift, the guard and score margins below replace this global descent argument with a local invariant.

Proposition 4.1 (Low-rank cost). *For a block $X \in \mathbb{R}^{n \times p}$, current rank $r \ll p$, and q residual power iterations, one full-data orientation sweep costs $O(npr + pr^2)$, a birth check costs $O(npq)$, and the maintained representation state uses $O(pr + r^2)$ memory. No $p \times p$ covariance or full SVD is required.*

4.1 The two errors that must balance

Use chordal distance

$$d_r(P, P_r^*) = 2^{-1/2} \left\| P - P_r^* \right\|_F.$$

There are two disturbances. Sampling perturbs the gradient and the rank scores; drift moves the target itself. Both admit explicit constants. Put $L_b = 1 + \beta/(2s)$. If $\left\| \widehat{\Sigma} - \Sigma \right\|_{\mathrm{op}} \leq \epsilon$ and both covariance norms are at most M, then, uniformly over $V \in \mathrm{St}(p, r)$,

$$\left\| \operatorname{grad} \mathcal{E}_{r,\widehat{\Sigma}}(V) - \operatorname{grad} \mathcal{E}_{r,\Sigma}(V) \right\|_F \leq C_E(r, M, s, \beta)\epsilon, \tag{8}$$

$$C_E(r, M, s, \beta) = \sqrt{r}\left[\frac{4M + \beta}{s} + \frac{8M^2 + \beta M}{s^2} + \frac{4M^3}{s^3} \right]. \tag{9}$$

Moreover, if $d_r(P, P_r^*) \leq e$, each available empirical score differs from its population value by at most

$$\zeta_r(e) = L_b(\epsilon + 2Me). \tag{10}$$

The first bound controls orientation; the second says when that orientation is accurate enough to trust a structural decision.

Let $g_r = \inf_k(\lambda_r(k) - \lambda_{r+1}(k)) > 0$, let ρ_{d} bound $\|\Sigma_{k+1} - \Sigma_k\|_{\mathrm{op}}$, and let $\underline{k}_r$ be a uniform lower bound on equation 3. For a sufficiently small local radius R_r^{d} and $\eta \leq 1/\overline{k}_r$, the population polar map contracts by

$$a_r = 1 - \frac{\eta \underline{k}_r}{2} < 1.$$

Combining this contraction with equation 8 and the sin–theta perturbation of the moving target gives the scalar recursion

$$e_{r,k+1} \le a_r e_{r,k} + 2\eta C_E \epsilon + \frac{2\sqrt{r}\,\rho_{\mathrm{d}}}{g_r}. \tag{11}$$

Its fixed point is the *maintenance radius*

$$R_r^{\mathrm{maint}} = \frac{4C_E\epsilon}{\underline{k}_r} + \frac{4\sqrt{r}\,\rho_{\mathrm{d}}}{\eta \underline{k}_r g_r}. \tag{12}$$

Equation equation 12 is the central balance. Sampling error and drift enlarge the maintained projection; curvature and eigengap stabilize it; a faster orientation step suppresses lag from drift but cannot remove the statistical floor.

It also exposes a feasibility boundary before any rank decision is made. Let m_r be the smallest population benefit margin relevant at rank r. Before imposing post-jump re-entry, the admissible interval for an inner decision radius is nonempty precisely when

$$R_r^{\mathrm{maint}} < \min\left\{ R_r^{\mathrm{d}}, \frac{m_r - h - L_b\epsilon}{2ML_b} \right\}. \tag{13}$$

More samples enlarge the window by reducing ϵ, while faster drift closes it through R_r^{maint}. Weak curvature or a small eigengap has the same effect even when the static optimum remains unique. A population optimum can therefore be identifiable and still lie beyond the dynamical resolution of the learner.

4.2 From local control to rank recovery

Choose an inner decision radius

$$R_r^{\mathrm{maint}} < \bar{R}_r < R_r^{\mathrm{d}}, \qquad \bar{a}_r = \epsilon + 2M\bar{R}_r, \qquad \zeta_r = L_b \bar{a}_r.$$

A rank decision made in this inner tube has score error at most ζ_r. It remains to ensure that a decision does not eject the state from the next orientation basin. Let ω_r^+ be the eigengap isolating the leading inactive residual line used for birth, and let ω_r^- isolate the weakest active line used for death. A Davis–Kahan bound gives the post-jump error

$$E_r^{\pm} = \bar{R}_r + \frac{2\bar{a}_r}{\omega_r^{\pm}}. \tag{14}$$

Thus $E_r^{\pm} < R_{r\pm1}^{\mathrm{d}}$ closes the geometric part of the jump. The next guard length is the number of contractions needed to return from $E_r^{\pm}$ to the next inner tube:

$$G_\ell(E) = \left\lceil \frac{\log\big((E - R_\ell^{\mathrm{maint}})/(\bar{R}_\ell - R_\ell^{\mathrm{maint}})\big)}{\log(1/a_\ell)} \right\rceil_+, \quad \ell = r \pm 1, \tag{15}$$

where the value is zero when $E \le \bar{R}_\ell$.

Theorem 4.2 (Representation maintenance across rank strata)**.** *Consider Algorithm 1 for N Gaussian blocks, with blockwise-constant optimal rank r^*. Let $\mathcal{Q}$ contain the ranks between the initial r_0 and r^*. For every nontrivial $q \in \mathcal{Q}$, assume:*

1. *the target eigengap is at least $g_q > 0$, the restoring curvature is at least $\underline{k}_q > 0$, the largest local curvature is at most $\overline{k}_q$, $\eta \le 1/\overline{k}_q$, and $\rho_{\mathrm{d}} < g_q/2$;*
2. *$R_q^{\mathrm{maint}} < \bar{R}_q < R_q^{\mathrm{d}}$, and $d_{r_0}(P_0, P_{r_0}^*(0)) \le \bar{R}_{r_0}$;*
3. *for every required move, $\bar{a}_q < \omega_q^{\pm}/2$ and $E_q^{\pm} < R_{q\pm1}^{\mathrm{d}}$;*
4. *the population score margin satisfies*

$$m := \min_{\substack{0 \le k < N \\ 1 \le i \le p}} |b(\lambda_i(k)) - \gamma| > h + \max_{q \in \mathcal{Q}} \zeta_q.$$

Use guards at least as long as equation 15, and check rank at least once every K updates outside a guard. Assume N is at least

$$K|r_0 - r^*| + \sum_{\text{required } q \to \ell} G_\ell(E_q^\pm).$$

On the event $\max_{k<N} \left\| \widehat{\Sigma}_k - \Sigma_k \right\|_{\text{op}} \le \epsilon$*:*

1. *every accepted jump moves exactly one rank toward r^*; no spurious or opposite jump is accepted;*
2. *the algorithm reaches r^* after exactly $|r_0 - r^*|$ jumps and within*

$$K|r_0 - r^*| + \sum_{\text{required } q \to \ell} G_\ell(E_q^\pm)$$

updates;

3. *after recovery, rank remains r^* and*

$$e_{r^*,k} \le a_{r^*}^{k-k_0} e_{r^*,k_0} + (1 - a_{r^*}^{k-k_0}) R_{r^*}^{\text{maint}}.$$

The endpoint strata $r = 0, p$ have zero orientation error and require no orientation guard.

For an ambient-dimension bound, set

$$u_{n,\delta} = \sqrt{p/n} + \sqrt{2\log(2N/\delta)/n}, \qquad \epsilon_n^{\text{amb}} = \Lambda(2u_{n,\delta} + u_{n,\delta}^2).$$

This bound holds for all n, p. A sharper high-dimensional form uses

$$r_{\text{eff}} = \max_{k<N} \frac{\operatorname{tr} \Sigma_k}{\|\Sigma_k\|_{\text{op}}}, \quad t_\delta = \log(eN/\delta), \quad \epsilon_n^{\text{eff}} = C_0 \Lambda \left[\sqrt{\frac{r_{\text{eff}} + t_\delta}{n}} + \frac{r_{\text{eff}} + t_\delta}{n} \right],$$

where C_0 is a universal numerical constant from Gaussian covariance concentration (Koltchinskii & Lounici, 2017). If $\|\Sigma_k\|_{\text{op}} \le \Lambda$, all conclusions hold with probability at least $1 - \delta$ using either valid radius $\epsilon \in \{\epsilon_n^{\text{amb}}, \epsilon_n^{\text{eff}}\}$ and $M = \Lambda + \epsilon$.

Propagation across rank strata. The proof is an induction over rank strata. First, the Hessian and polar retraction give population contraction. Gradient perturbation and target drift then give equation 11, whose invariant solution is equation 12. Inside the inner tube, equation 10 and the margin select only the move toward r^*. Equation equation 14 puts the jumped state inside the neighboring outer tube, and equation 15 returns it to the next decision tube. The induction therefore cannot lose the hypothesis it needs. Finally, Gaussian singular-value concentration and a union bound establish the covariance event. Complete proofs are in Appendix A.

Finite eigensolvers. Suppose an approximate Ritz value has error at most τ_λ and its rank-one projector is within τ_u of the exact empirical Ritz line. Theorem 4.2 remains valid after replacing ζ_q by $L_b(\bar{a}_q + \tau_\lambda)$ and $E_q^\pm$ by

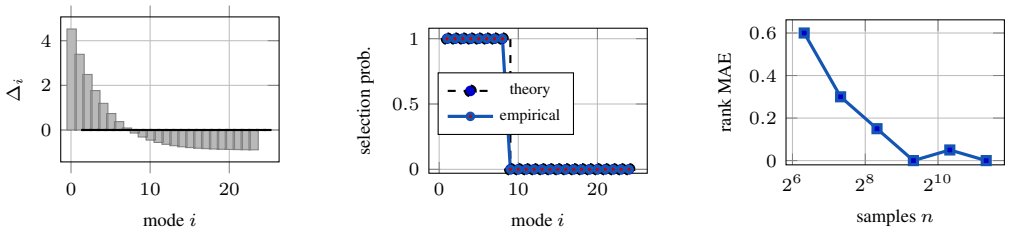
$$\bar{R}_q + 2\bar{a}_q/\omega_q^\pm + \tau_u.$$

For m power iterations with leading residual eigenvalues $\theta_1 > \theta_2$ and initial angle α_0, one may take $\tau_u \le (\theta_2/\theta_1)^m \tan\alpha_0$.

5 Reorganization and breakdown

When a population mode crosses its price, the stable projection changes. Hysteresis holds the current stratum until the evidence dominates estimation error.

Corollary 5.1 (Controlled reorganization)**.** *Suppose a single signed move benefit $\Psi_j(k)$ crosses zero at block τ, is negative immediately before the crossing, and satisfies $\Psi_j(k+1) - \Psi_j(k) \ge \nu_{\text{d}} > 0$ afterward. Let $\zeta = \max_q \zeta_q$. If the previous guard has expired, $h > \zeta$, all other modes retain margin*



(a) Population benefit. (b) Selected modes. (c) Sampling recovery.

Figure 1: The spectral law predicts both the active modes and the finite-sample transition to exact rank recovery.

greater than $h + \zeta$, and the re-entry condition in Theorem 4.2 holds, then no jump occurs before the population crossing. The unique required jump is accepted within

$$K + \left\lceil \frac{h + \zeta}{\nu_{\mathrm{d}}} \right\rceil$$

updates after τ, and its state enters the neighboring invariant tube.

Here $h + \zeta$ is the evidence required for a reliable structural change, and ν_{d} is the rate at which the environment supplies it.

There is also a boundary at which no maintenance rule can succeed.

Proposition 5.2 (Unavoidable sampling boundary)**.** *Let $\lambda_c > 0$ satisfy $b(\lambda_c) = \gamma$. For every n, there exist two centered Gaussian distributions whose optimal ranks differ by one and whose spectral benefit margins are $O(n^{-1/2})$, such that every rank estimator $\widetilde{r}$ based on n samples satisfies*

$$\max_{\theta \in \{-,+\}} \Pr_{\theta}(\widetilde{r} \neq r^*_{\theta}) \geq \frac{3}{8}.$$

Away from the boundary, finite-sample perturbation fits inside the score margin and the projection is maintainable. Within an $n^{-1/2}$ neighborhood, the underlying rank states are statistically indistinguishable. The apparent *raw noise* outside a stable representation appears when the data no longer contain enough evidence to support a unique dimension. For fixed p and bounded spectra, the upper radius and this local ambiguity both have $n^{-1/2}$ order. When p grows, the upper radius is governed by r_{eff}; it retains the same order when $r_{\mathrm{eff}} = O(1)$, and records the additional difficulty when the effective rank also grows.

6 Theory Diagnostics

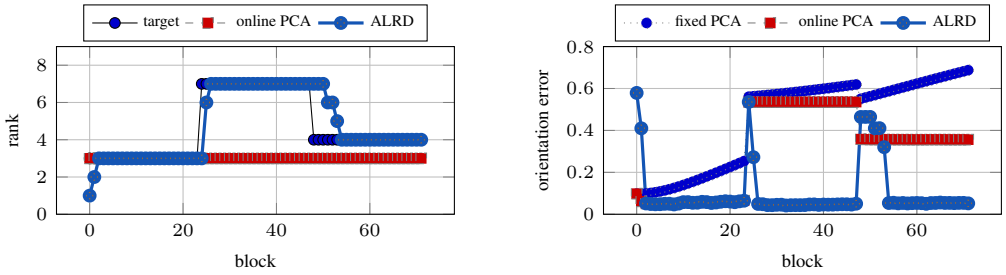
Centered Gaussian streams test the spectral law, joint maintenance, curvature, and hysteresis. Scientific snapshots examine the adaptive spectral rule, and a gated autoencoder supplies a nonlinear relaxation. Stochastic synthetic results use 20 fixed seeds and MNIST uses ten paired runs. Appendix B gives the complete protocols.

Spectral selection and sampling. For $p = 24$, the population law equation 2 selects rank 8. Exact-rank recovery probabilities for $n = 80, 160, 320, 640, 1280, 2560$ are respectively $0.40, 0.70, 0.85, 1.00, 0.95, 1.00$, and rank MAE falls from 0.60 to zero. Across seeds, spectral rank MAE is 0.050 ± 0.224. On a (s, γ) transition grid, empirical and population phases agree on $95.1\% \pm 1.1\%$ of points; discrepancies concentrate near $b(\lambda) = \gamma$, exactly where Proposition 5.2 predicts instability. Increasing encoder noise produces the predicted monotone rank decrease, with $93.8\% \pm 3.3\%$ rank agreement.

Joint maintenance. We separate three streams: A changes rank $3 \to 7 \to 4$ at fixed orientation; B holds rank 5 while rotating the target by 0.85 radians; C combines both. Fixed PCA changes neither coordinate, online PCA changes orientation only, and ALRD changes both. Table 1 reports

Table 1: Synthetic maintenance diagnostics. Errors are time averages; lower is better. Each cell is rank error / normalized chordal orientation error.

Method	A: rank drift	B: rotation	C: both
Fixed PCA	1.667/0.340	0/0.414	1.667/0.456
Online PCA	1.667/0.310	**0/0.053**	1.667/0.316
ALRD	**0.292/0.081**	0.056/0.063	**0.306/0.104**



(a) Dimension coordinate.

(b) Grassmann coordinate.

Figure 2: Case C changes rank and orientation simultaneously. Fixed and fixed-rank trackers each miss one coordinate; ALRD maintains both.

Table 2: Scientific snapshot ensembles. Each cell is spectral-target rank error / relative reconstruction error / normalized subspace error.

Data	Fixed POD	Adaptive POD	ALRD
Diffusion	7.750/.244/.516	1.143/.368/.264	.**107**/.318/.**146**
Allen–Cahn	3.379/.592/.533	13.828/.**344**/.581	**1.759**/.510/.**464**
Navier–Stokes	.**920**/.209/.**373**	5.600/.322/.422	2.400/.**187**/.392

a representative run. Across all 20 runs and cases, ALRD obtains rank MAE 0.2120 ± 0.0090 and orientation error 0.0829 ± 0.0017. Rank adaptation accounts for the gain in A, Grassmann motion accounts for the gain in B, and their combination accounts for the gain in C.

Curvature and hysteresis. For a spectrum with predicted least-curvature rate 0.7825, fitted convergence rates from initial angles $5°, 15°, 30°, 60°$ are $0.8081, 0.8081, 0.8080, 0.8077$, all within 3.3% of Theorem 3.2. In a rank-changing stream, full ALRD makes seven switches with rank/orientation errors $0.178/0.080$. Removing rank dynamics raises them to $1.333/0.277$; removing Grassmann motion raises orientation error to 0.422; removing hysteresis increases switching from 7 to 17. In a separate $3 \to 8 \to 5$ hybrid trajectory, eight accepted jumps have minimum energy decrement $0.180 > h = 0.055$, while the finite-switching certificate equation 7 permits at most 44. Together, the four ablations separate rank motion, orientation motion, and hysteresis.

Low-rank scaling and scientific snapshots. At retained rank 12, increasing p from 500 to 5000 changes one ALRD sweep from 0.52 to 2.84 ms and its $p \times r$ state from 0.046 to 0.458 MiB. A full thin SVD takes 41.80 and 57.93 ms at the same endpoints. We also apply the spectral rule to evolving snapshot ensembles from diffusion, Allen–Cahn, and Navier–Stokes solvers, where POD is the conventional reduced representation (Holmes et al., 2012). Table 2 reports rank, reconstruction, and subspace errors separately. ALRD most closely follows the spectral target for diffusion and Allen–Cahn and has the smallest Navier–Stokes reconstruction residual.

Gated nonlinear relaxation. A width-32 autoencoder replaces discrete birth and death maps with gates $a_j = \mathrm{sigmoid}(g_j/\tau)$, soft dimension $r_{\mathrm{soft}} = \sum_j a_j$, a price γr_{soft}, and a polarization penalty; stable gates are thresholded and fine-tuned. On ten MNIST subsets (LeCun et al., 1998), it selects 16.50 ± 1.27 dimensions and stabilizes at epoch 15.90 ± 2.77. Table 3 reports the original comparisons (Baldi & Hornik, 1989; Kingma & Welling, 2014; Han et al., 2015). Gated ALRD

Table 3: MNIST over ten runs: retained dimension, test reconstruction MSE, and frozen-code linear-probe accuracy.

Method	Dimension	MSE	Accuracy (%)
Fixed AE $r = 16$	16	$.03208 \pm .00047$	$\mathbf{83.36} \pm 1.05$
VAE $r = 16$	16	$.03411 \pm .00045$	82.96 ± 1.12
Post-train pruning	16.50 ± 1.27	$.03389 \pm .00090$	82.08 ± 1.14
Gated ALRD	16.50 ± 1.27	$\mathbf{.02966} \pm .00134$	81.74 ± 1.37

improves reconstruction at approximately 16 dimensions. The fixed autoencoder has higher linear-probe accuracy. Reconstruction and label separation thus vary differently along the dimension path.

Varying γ over $\{1, 2, 3, 4, 6, 8, 12\} \times 10^{-4}$ decreases the selected dimension from 23.9 to 5.2. Between 4 and 6×10^{-4}, dimension falls from 16.2 to 12.9 while MSE changes only from 0.02913 to 0.02926; accuracy decreases along the path. Table 8 reports the full regularization path.

7 Discussion

The spectral theorem says which abstraction minimizes the population energy. The maintenance theorem says whether a finite-sample learner can keep that abstraction operational while the world moves. Their combination reveals three regimes. When curvature, eigengap, and score margin dominate drift and sampling, the projection is maintained inside an invariant tube. At an isolated crossing, that tube reorganizes after a controlled delay. At the sampling boundary, no method can identify a unique dimension reliably.

The rank price defines excessive complexity. Grassmann flow repairs orientation at fixed capacity. Birth and death repair capacity across adjacent strata. Hysteresis converts uncertain evidence into delayed, reliable reorganization. Stability emerges from the repeated restoration of the relation between a projection and its environment.

For the linear Gaussian channel, isotropic encoder noise and a rank price make the structural benefit, restoring curvature, and re-entry error explicit. Other representation families lead to the same three questions: which coordinates merit existence, how strongly the current representation returns after perturbation, and whether a structural change enters a stable basin.

8 Conclusion

The theory identifies a feasibility boundary for adaptive representation learning. A benefit margin makes dimension statistically resolvable; curvature and eigengap restore orientation; sampling error and drift consume the available margin. Their balance determines whether a projection remains inside an invariant tube, whether a rank decision can be trusted, and whether the state after that decision enters the next recovery basin. This boundary also describes how representations reorganize. At an isolated spectral crossing, evidence accumulates until it exceeds the statistical and hysteretic margins, producing a controlled delay. Near the sampling boundary, the data do not support a unique dimension. Representation dimension is thus a learned state with its own resolution and recovery time, alongside model parameters and features. The same problem appears in streaming encoders, continual learning systems, scientific reduced-order models, and resource-constrained edge learners. Each must determine which distinctions the current evidence can support and how long they can be preserved as conditions change. Understanding, in this sense, is the maintained agreement between an evolving world and a projection simple enough to use.

AI use statement

Generative AI tools were used to assist with language editing, LaTeX organization, and algebraic cross-checks. All derivations, claims, references, and experimental outputs were reviewed and verified by the authors, who take responsibility for the complete content of the submission.

Reproducibility statement

All assumptions and constants used in the main claims are stated in Sections 2–5; complete proofs appear in Appendix A. Appendices B and C specify the synthetic, scientific, and MNIST data-generating processes, metrics, hyperparameters, seeds, and stopping rules. The supplementary material contains the scripts and raw CSV outputs for every reported experiment.

References

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. In *International Conference on Learning Representations*, 2017. URL https://arxiv.org/abs/1612.00410.

Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, 2(1):53–58, 1989. doi: 10.1016/0893-6080(89)90014-2.

Laura Balzano, Robert Nowak, and Benjamin Recht. Online identification and tracking of subspaces from highly incomplete information. In *Allerton Conference on Communication, Control, and Computing*, pp. 704–711, 2010. doi: 10.1109/ALLERTON.2010.5706976.

Yuejie Chi, Yonina C. Eldar, and Robert Calderbank. PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations. *IEEE Transactions on Signal Processing*, 61(23):5947–5959, 2013. doi: 10.1109/TSP.2013.2282910.

Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. doi: 10.1007/BF02288367.

Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998. doi: 10.1137/S0895479895290954.

Matan Gavish and David L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014. doi: 10.1109/TIT.2014.2323359.

Rafal Goebel, Ricardo G. Sanfelice, and Andrew R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, 2012.

Song Han, Jeff Pool, John Tran, and William J. Dally. Learning both Weights and Connections for Efficient Neural Networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015. URL https://arxiv.org/abs/1506.02626.

Philip Holmes, John L. Lumley, Gal Berkooz, and Clarence W. Rowley. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, 2 edition, 2012. doi: 10.1017/CBO9780511919701.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. URL https://arxiv.org/abs/1312.6114.

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.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Thomas P. Minka. Automatic choice of dimensionality for PCA. In *Advances in Neural Information Processing Systems*, volume 13, pp. 598–604, 2000. URL https://proceedings.neurips.cc/paper/1853-automatic-choice-of-dimensionality-for-pca.pdf.

Erkki Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273, 1982. doi: 10.1007/BF00275687.

Patrick O. Perry and Patrick J. Wolfe. Minimax rank estimation for subspace tracking. *IEEE Journal of Selected Topics in Signal Processing*, 4(3):504–513, 2010. doi: 10.1109/JSTSP.2010.2048070.

Daniel J. Rabideau. Fast, rank adaptive subspace tracking and applications. *IEEE Transactions on Signal Processing*, 44(9):2229–2244, 1996. doi: 10.1109/78.536680.

Michael E. Tipping and Christopher M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B*, 61(3):611–622, 1999. doi: 10.1111/1467-9868.00196.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. In *Allerton Conference on Communication, Control, and Computing*, pp. 368–377, 1999. URL `https://arxiv.org/abs/physics/0004057`.

Guifang Zhou, Wen Huang, Kyle A. Gallivan, Paul Van Dooren, and P.-A. Absil. A Riemannian rank-adaptive method for low-rank optimization. *Neurocomputing*, 192:72–80, 2016. doi: 10.1016/j.neucom.2016.02.030.

A Proofs

A.1 Proof of the spectral variational law

Let $A = \Sigma^2$, $B = \Sigma + sI$, and $V^\top V = I_r$. The matrices A, B, Σ are simultaneously diagonalizable. The generalized eigenvalues of (A, B) are

$$f(\lambda_i) = \frac{\lambda_i^2}{\lambda_i + s},$$

and f is increasing on $[0, \infty)$. The generalized Ky Fan principle therefore gives

$$\operatorname{tr}\big[(V^\top A V)(V^\top B V)^{-1}\big] \le \sum_{i=1}^{r} \frac{\lambda_i^2}{\lambda_i + s}.$$

If μ_i are the ordered Ritz values of $V^\top \Sigma V$, Poincaré separation gives $\mu_i \le \lambda_i$. Since $x \mapsto \log(1 + x/s)$ is increasing,

$$\log\det(I + s^{-1} V^\top \Sigma V) \le \sum_{i=1}^{r} \log(1 + \lambda_i/s).$$

The leading eigenspace attains both bounds simultaneously. Substitution into equation 1 gives $\mathcal{E}_r^* = \operatorname{tr}\Sigma - \sum_{i \le r} b(\lambda_i) + \gamma r$. Consequently $\mathcal{E}_r^* - \mathcal{E}_{r-1}^* = \gamma - b(\lambda_r)$, proving Theorem 3.1. Equality cases give uniqueness when $\lambda_r > \lambda_{r+1}$.

A.2 Hessian audit

Write $V_* = U_1$ for the leading eigenbasis and U_2 for its orthogonal complement. Every horizontal tangent vector is $\xi = U_2 K$, with $\|\xi\|_F = \|K\|_F$. The Grassmann geodesic with this initial velocity has expansion

$$V(t) = U_1 + t U_2 K - \tfrac{1}{2} t^2 U_1 K^\top K + O(t^3).$$

Because $\Sigma U_a = U_a \Lambda_a$, the first variations of $V^\top \Sigma V$ and $V^\top \Sigma^2 V$ vanish at V_*. The coefficient can be checked on one elementary rotation. Replace the active eigenvector u_i by $v_i(u) = u_i \cos u + u_j \sin u$, where $i \le r < j$, and put

$$q = \sin^2 u, \qquad \mu(q) = \lambda_i + (\lambda_j - \lambda_i) q, qquad \nu(q) = \lambda_i^2 + (\lambda_j^2 - \lambda_i^2) q.$$

All varying terms of the energy are

$$\phi_{ij}(q) = -\frac{\nu(q)}{\mu(q) + s} - \frac{\beta}{2} \log\!\left(1 + \frac{\mu(q)}{s}\right).$$

Direct differentiation at $q = 0$ gives

$$\frac{d}{dq}\left[-\frac{\nu(q)}{\mu(q)+s}\right]\Big|_{q=0} = \frac{(\lambda_i-\lambda_j)[\lambda_i\lambda_j + s(\lambda_i+\lambda_j)]}{(\lambda_i+s)^2},$$
$$\frac{d}{dq}\left[-\frac{\beta}{2}\log\left(1+\frac{\mu(q)}{s}\right)\right]\Big|_{q=0} = \frac{\beta(\lambda_i-\lambda_j)}{2(\lambda_i+s)}.$$

Since $q = u^2 + O(u^4)$, twice their sum is the second derivative of $u \mapsto \phi_{ij}(\sin^2 u)$ at zero, namely k_{ij}. Expanding the inverse and log determinant in equation 1 yields

$$\mathcal{E}_r(V(t)) = \mathcal{E}_r(V_*) + \frac{t^2}{2}\sum_{i\le r<j} k_{ij}|K_{j-r,i}|^2 + O(t^3),$$

where k_{ij} is exactly equation 3. There are no mixed quadratic terms: the elementary active–inactive rotations diagonalize the full Riemannian Hessian. Each coefficient is positive when $\lambda_i > \lambda_j$, since $s > 0$ and $\beta \ge 0$. Hessian continuity then supplies a normal neighborhood in which the objective is strongly geodesically convex with any constant below $\min_{i\le r<j} k_{ij}$. Linearizing gradient flow, or its polar retraction, proves the claimed local convergence.

A.3 Explicit covariance perturbation

For fixed V, set

$$D_\Sigma = (V^\top \Sigma V + sI)^{-1}, \qquad C_\Sigma = V^\top \Sigma^2 V.$$

Direct differentiation of equation 1 gives

$$\nabla_V \mathcal{E}_{r,\Sigma} = -2\Sigma^2 V D_\Sigma + 2\Sigma V D_\Sigma C_\Sigma D_\Sigma - \beta \Sigma V D_\Sigma. \tag{16}$$

Suppose $\left\|\widehat{\Sigma} - \Sigma\right\|_{\mathrm{op}} \le \epsilon$ and both covariance norms are bounded by M. The resolvent identity gives

$$\left\|D_{\widehat{\Sigma}} - D_\Sigma\right\|_{\mathrm{op}} \le \epsilon/s^2, \qquad \left\|C_{\widehat{\Sigma}} - C_\Sigma\right\|_{\mathrm{op}} \le 2M\epsilon.$$

Using $\|V\|_F = \sqrt{r}$, the three differences in equation 16, before their leading factors $2, 2, \beta$, are bounded respectively by

$$\sqrt{r}\,\epsilon(2M/s + M^2/s^2),$$
$$\sqrt{r}\,\epsilon(3M^2/s^2 + 2M^3/s^3),$$
$$\sqrt{r}\,\epsilon(1/s + M/s^2).$$

Combining them produces exactly C_E in equation 9. The Grassmann projection $Z \mapsto (I - VV^\top)Z$ is nonexpansive, proving equation 8.

For the scores, $|b'(x)| \le L_b = 1 + \beta/(2s)$. If $P = VV^\top$, $P^* = P_r^*$, and $Q = I - P$, then

$$\begin{aligned}\left\|Q\widehat{\Sigma}Q - (I-P^*)\Sigma(I-P^*)\right\|_{\mathrm{op}} &\le \epsilon + 2M\left\|P - P^*\right\|_{\mathrm{op}} \\ &\le \epsilon + 2M d_r(P, P^*).\end{aligned}$$

The same inequality holds for the compressed active operators. Weyl's inequality followed by the Lipschitz bound for b proves equation 10.

A.4 Discrete contraction and the invariant tube

At $P_r^*(k)$, the derivative of the population map

$$\mathcal{T}_{r,k}(V) = \mathrm{polar}(V - \eta\,\mathrm{grad}\,\mathcal{E}_{r,\Sigma_k}(V))$$

is diagonal in the Hessian eigenbasis, with eigenvalues $1 - \eta k_{ij}(k)$. Under $0 < \eta \le 1/\overline{k}_r$, its norm is at most $1 - \eta\underline{k}_r$. Continuity of this derivative, uniformly over the finite horizon, gives a radius $R_r^{\mathrm{d}} > 0$ on which the contraction factor is at most $a_r = 1 - \eta\underline{k}_r/2$. More explicitly, in horizontal normal coordinates,

$$D\mathcal{T}_{r,k}(V_*)[U_2K] = U_2\widetilde{K}, \qquad \widetilde{K}_{j-r,i} = (1 - \eta k_{ij}(k))K_{j-r,i}.$$

The differential of chordal distance agrees with the canonical Grassmann metric at the origin. Shrinking the common normal neighborhood transfers the same bound, with the stated factor a_r, to chordal distance.

The population and empirical polar arguments have smallest singular value at least one: the Riemannian gradient is horizontal, so

$$(V - \eta G)^\top (V - \eta G) = I + \eta^2 G^\top G.$$

For full-column-rank matrices A, B, the polar perturbation inequality reads

$$\|\mathrm{polar}(A) - \mathrm{polar}(B)\|_F \le \frac{2\,\|A - B\|_F}{\sigma_{\min}(A) + \sigma_{\min}(B)}.$$

Applying it to the two update arguments, using equation 8, and then passing from aligned frames to their projectors gives

$$d_r(\widehat{\mathcal{T}}_{r,k}(P), \mathcal{T}_{r,k}(P)) \le 2\eta C_E \epsilon.$$

The sin–theta theorem and $\|\Sigma_{k+1} - \Sigma_k\|_{\mathrm{op}} \le \rho_{\mathrm{d}} < g_r/2$ give

$$d_r(P_r^*(k+1), P_r^*(k)) \le \frac{2\sqrt{r}\,\rho_{\mathrm{d}}}{g_r}.$$

Indeed, $d_r(P, P') = \|\sin\Theta(P, P')\|_F$, and Davis–Kahan gives $\|\sin\Theta\|_F \le \sqrt{r}\,\rho_{\mathrm{d}}/(g_r - \rho_{\mathrm{d}})$; the assumption $\rho_{\mathrm{d}} < g_r/2$ yields the displayed bound. The triangle inequality proves equation 11. Iterating it gives

$$e_{r,k} \le a_r^k e_{r,0} + (1 - a_r^k) R_r^{\mathrm{maint}},$$

because $1 - a_r = \eta \underline{k}_r/2$. For completeness, suppose the first exit occurs at update $m+1$. The recursion is valid through update m, and hence

$$e_{r,m+1} \le a_r^{m+1} e_{r,0} + (1 - a_r^{m+1}) R_r^{\mathrm{maint}} \le \max\{e_{r,0}, R_r^{\mathrm{maint}}\} < R_r^{\mathrm{d}},$$

which contradicts the definition of the first exit. This proves invariance.

A.5 Jump re-entry and proof of Theorem 4.2

Suppose a rank check is made with $d_r(P, P_r^*) \le \bar{R}_r$. The residual and compressed empirical covariance operators are within $\bar{a}_r = \epsilon + 2M\bar{R}_r$ of their population counterparts. If the relevant Ritz line is isolated by $\omega_r^\pm$ and $\bar{a}_r < \omega_r^\pm/2$, the one-dimensional sin–theta bound gives

$$\sin\angle(\widehat{u}, u^*) \le 2\bar{a}_r/\omega_r^\pm.$$

Let u_i be population covariance eigenvectors. For birth and death, respectively, the neighboring population targets satisfy

$$P_{r+1}^* = P_r^* + u_{r+1}u_{r+1}^\top, \qquad P_{r-1}^* = P_r^* - u_r u_r^\top.$$

The empirical jump uses the corresponding residual or compressed Ritz line $\widehat{u}$, which is orthogonal to the retained projector in the relevant space. Since the chordal distance between rank-one projectors equals the sine of their principal angle, the triangle inequality gives, for either sign,

$$d_{r\pm1}(P^\pm, P_{r\pm1}^*) \le d_r(P, P_r^*) + d_1(\widehat{u}\widehat{u}^\top, u^* u^{*\top}) \le \bar{R}_r + \frac{2\bar{a}_r}{\omega_r^\pm}.$$

This is the post-jump bound $d_{r\pm1}(P^\pm, P_{r\pm1}^*) \le E_r^\pm$ in equation 14. The assumed strict inequality $E_r^\pm < R_{r\pm1}^{\mathrm{d}}$ places the new state inside the neighboring contraction tube. Iterating the scalar recursion from $E_r^\pm$ proves the guard formula equation 15.

It remains to verify the direction of every decision. Write the birth and death desirabilities as $s_q^+ = \widehat{\Delta}_q^+$ and $s_q^- = -\widehat{\Delta}_q^-$. Inside the decision tube each differs from its population value by at most ζ_q. Therefore

$$\begin{aligned}
q < r^*: &\quad s_q^+ \ge m - \zeta_q > h, && s_q^- \le -m + \zeta_q < -h,\\
q > r^*: &\quad s_q^- \ge m - \zeta_q > h, && s_q^+ \le -m + \zeta_q < -h,\\
q = r^*: &\quad s_q^+ \le -m + \zeta_q < -h, && s_q^- \le -m + \zeta_q < -h,
\end{aligned}$$

whenever the indicated score exists. Thus exactly the move toward r^* is accepted away from the target rank, and neither move is accepted at r^*. At $q = 0$ only the birth score exists; at $q = p$ only the death score exists.

Induct on $|q - r^*|$. Tube invariance brings the state to a decision; the score argument decreases $|q - r^*|$ by one; re-entry and the guard restore the next decision-tube hypothesis. This proves the exact jump count and the recovery-time sum. After the last jump, the scalar recursion gives the stated tracking bound. For $q = 0, p$, the Grassmann component is a singleton, so its orientation error and guard are zero.

For the explicit ambient bound, write each Gaussian block as

$$\widehat{\Sigma}_k = \Sigma_k^{1/2}(Z_k^\top Z_k/n)\Sigma_k^{1/2}.$$

The rectangular Gaussian singular-value inequalities imply, for all n, p,

$$\left\|Z_k^\top Z_k/n - I\right\|_{\mathrm{op}} \le 2u_{n,\delta} + u_{n,\delta}^2$$

simultaneously over N blocks with probability at least $1 - \delta$. Multiplication by $\Sigma_k^{1/2}$ gives $\left\|\widehat{\Sigma}_k - \Sigma_k\right\|_{\mathrm{op}} \le \epsilon_n^{\mathrm{amb}}$. For the effective-rank form, Gaussian covariance concentration gives

$$\left\|\widehat{\Sigma}_k - \Sigma_k\right\|_{\mathrm{op}} \le C_0 \left\|\Sigma_k\right\|_{\mathrm{op}} \left[\sqrt{\frac{r_{\mathrm{eff}} + t_\delta}{n}} + \frac{r_{\mathrm{eff}} + t_\delta}{n}\right]$$

with failure probability at most e^{-t_δ} per block (Koltchinskii & Lounici, 2017). A union bound gives $\epsilon_n^{\mathrm{eff}}$ simultaneously over the stream. Either event also gives $\left\|\widehat{\Sigma}_k\right\|_{\mathrm{op}} \le \Lambda + \epsilon = M$.

A.6 Crossing delay and lower bound

For Corollary 5.1, the empirical signed score differs from $\Psi_j(k)$ by at most ζ. Since $h > \zeta$, a pre-crossing score cannot pass the acceptance threshold. Once $\Psi_j(k) \ge h + \zeta$, acceptance is certain. The growth condition bounds the hitting time by $\lceil (h + \zeta)/\nu_{\mathrm{d}} \rceil$, and the next check adds at most K updates. Other-mode margins make the move unique, and the re-entry result places it in the adjacent tube.

For Proposition 5.2, keep all covariance eigenvalues except one fixed and separated from the threshold. Set

$$\lambda_\pm = \lambda_c \pm d, \qquad d = \frac{\lambda_c}{8\sqrt{n}}.$$

The two optimal ranks differ by one. Their product Gaussian distributions differ in only one scalar variance. With $x = 2d/(\lambda_c - d)$ and $x - \log(1 + x) \le x^2/2$,

$$D_{\mathrm{KL}}(P_+^n \| P_-^n) = \frac{n}{2}\{x - \log(1 + x)\} \le \frac{4nd^2}{\lambda_c^2} = \frac{1}{16}.$$

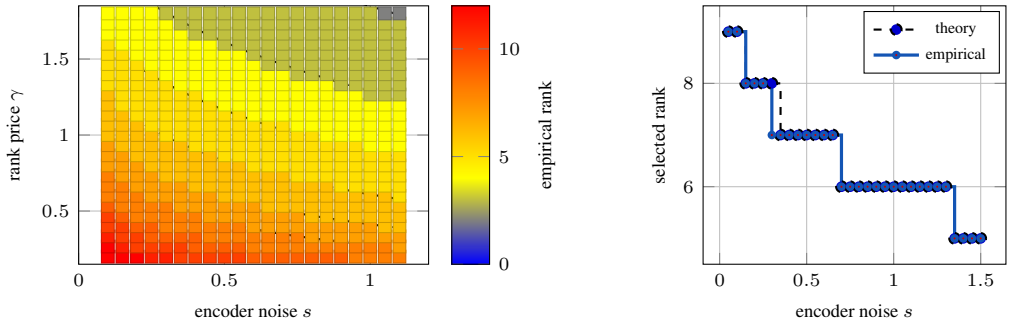
Pinsker's inequality bounds total variation by $1/\sqrt{32} < 1/4$. Le Cam's two-point inequality then gives minimax error greater than $3/8$. The mean-value theorem and boundedness of b' near λ_c give the $O(n^{-1/2})$ benefit margin.

B Synthetic protocol

Common implementation. Exact-rank ALRD uses polar-equivalent QR-retracted Riemannian updates with learning rate 0.05, covariance blocks of size 150, hysteresis $h = 0.05$, four power iterations, and a rank check every five updates. Runs use seeds $100, \ldots, 119$. Orientation error is normalized chordal distance

$$\frac{\left\|\widehat{P}_k - P_k^*\right\|_F}{\sqrt{2\max(\widehat{r}_k, r_k^*)}},$$

and rank error is $|\widehat{r}_k - r_k^*|$.



(a) Dimension-transition diagram.

(b) Noise sensitivity.

Figure 3: Further checks of the spectral law. Empirical phase boundaries follow $b(\lambda_i) = \gamma$; increasing encoder noise lowers selected rank.

Table 4: Finite-sample spectral selection over 20 seeds.

n	Mean selected rank	Exact recovery	Rank MAE
80	7.50	.40	.60
160	7.70	.70	.30
320	7.85	.85	.15
640	8.00	1.00	.00
1280	7.95	.95	.05
2560	8.00	1.00	.00

Table 5: Component ablation on the coupled stream.

Variant	Rank error	Orientation error	Switches
Full ALRD	.178	.080	7
No rank dynamics	1.333	.277	0
No Grassmann adaptation	.178	.422	7
No information term	.289	.109	9
No hysteresis	.222	.095	17

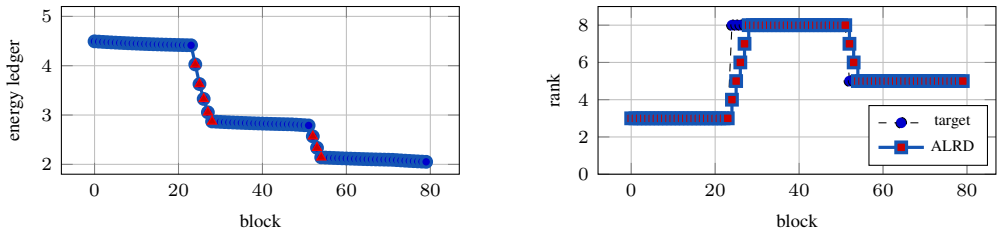
Spectral experiment. We set $p = 24$, $\lambda_i = 5.4\exp[-0.23(i-1)] + 0.04$, $s = 0.42$, $\beta = 0.30$, and $\gamma = 0.92$, for which Theorem 3.1 gives $r^* = 8$. For every sample size, each seed draws an independent Gaussian sample and applies the empirical spectral rule. The transition diagnostic evaluates 525 (s, γ) points. The noise diagnostic uses an 18-mode spectrum, $\beta = 0.26$, $\gamma = 0.88$, and 30 values of s between 0.05 and 1.50.

Maintenance experiment. All three streams are in $\mathbb{R}^{24}$. Case A changes target rank $3 \to 7 \to 4$ with fixed eigenvectors. Case B holds rank 5 and rotates the target through 0.85 radians. Case C applies both schedules. Fixed PCA freezes both initial rank and orientation. Online PCA performs the same Grassmann updates as ALRD but holds the initial rank fixed. The ablations remove rank transitions, orientation updates, information reward, or hysteresis one at a time.

Curvature experiment. The flow is initialized in the elementary rotation attaining the minimum in equation 3. We fit the slope of log orientation error on the common post-transient interval for initial angles $5°, 15°, 30°, 60°$. The predicted exponent is computed directly from the audited Hessian coefficient, without fitting.

Table 6: Computational scaling at retained rank 12 and block size 192.

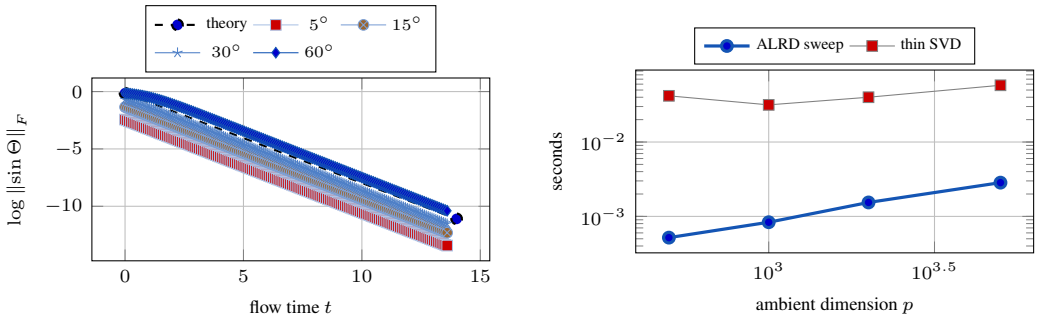
p	ALRD (ms)	Thin SVD (ms)	State (MiB)	Block (MiB)
500	.52	41.80	.046	.732
1000	.83	31.59	.092	1.465
2000	1.54	40.01	.183	2.930
5000	2.84	57.93	.458	7.324



(a) Jump-energy certificate.

(b) Finite birth/death sequence.

Figure 4: Hybrid switching: accepted jumps reduce the energy ledger and track the changing target rank.



(a) Curvature-predicted decay.

(b) Measured scaling.

Figure 5: Within-stratum convergence and low-rank computation.

C Additional Experimental Results

C.1 Spectral Recovery, Ablation, and Scaling

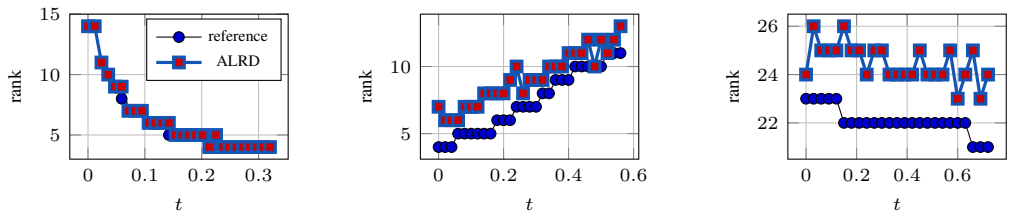
The cost in Proposition 4.1 follows by forming XV and $X^\top(XV)$ in $O(npr)$, projecting and retracting in $O(pr^2)$, and using q residual matrix–vector products for a birth proposal in $O(npq)$. Storing V and its small Gram-space quantities costs $O(pr + r^2)$.

C.2 Scientific Snapshot Protocol

The diffusion experiment uses 72 interior grid points, 14 sine modes, 80 snapshots per ensemble, diffusivity 0.025, 28 output times on $[0, 0.32]$, and additive noise variance 0.03. Allen–Cahn uses 96 periodic grid points, 80 snapshots, 29 output times, interface parameter 0.035, and noise variance 0.02. The dealiased vorticity solver uses a 12×12 grid, viscosity 0.015, 25 output times, 60 snapshots, and noise variance 0.006. We compare initial fixed POD, adaptive POD retaining 90% noise-corrected energy, and the ALRD spectral rule. The reference rank is

$$r_{\mathrm{ref}}(t) = \#\{i : b(\lambda_i(t)) > \gamma\}.$$

Rank error is measured against this reference; reconstruction and projector errors are reported separately in Table 2. The observations are time-indexed snapshot ensembles, and all three metrics concern their reduced representations.



(a) Diffusion. (b) Allen–Cahn. (c) Navier–Stokes.

Figure 6: Spectral-reference and ALRD rank trajectories for the scientific snapshot ensembles retained from the original paper.

Table 7: Complete MNIST comparison over ten paired runs.

Method	Dimension	Test MSE	Accuracy (%)
Fixed AE $r = 4$	4	$.04270 \pm .00068$	63.34 ± 2.73
Fixed AE $r = 8$	8	$.03315 \pm .00067$	79.20 ± 1.04
Fixed AE $r = 16$	16	$.03208 \pm .00047$	83.36 ± 1.05
Fixed AE $r = 32$	32	$.03305 \pm .00043$	$84.13 \pm .79$
VAE $r = 16$	16	$.03411 \pm .00045$	82.96 ± 1.12
Post-train pruning	16.50 ± 1.27	$.03389 \pm .00090$	82.08 ± 1.14
Gated ALRD	16.50 ± 1.27	$.02966 \pm .00134$	81.74 ± 1.37

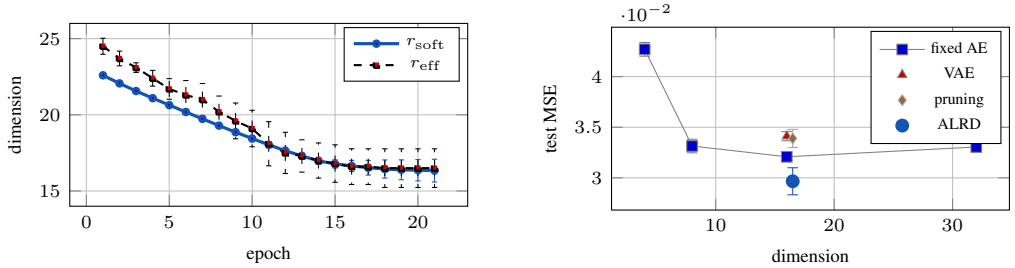
C.3 Nonlinear relaxation and complete MNIST results

Let $h_\phi(x) \in \mathbb{R}^{32}$ be the last encoder layer and

$$a_j = \mathrm{sigmoid}(g_j/\tau), \qquad z = a \odot h_\phi(x) + \xi, \qquad r_{\mathrm{soft}} = \sum_{j=1}^{32} a_j.$$

The training loss is reconstruction error plus γr_{soft}, weight decay, and $32^{-1}\sum_j a_j(1-a_j)$, which encourages nearly binary gates. The gates provide a differentiable relaxation of the exact birth and death process. Theorem 4.2 governs the blockwise Gaussian process defined in Sections 2–4.

Each of ten seeds uses 8,000 training and 2,000 test MNIST images, architecture $784 \to 256 \to 32$, Adam with learning rate 10^{-3}, batch size 256, $\tau = 0.5$, weight decay 10^{-4}, $\gamma = 4 \times 10^{-4}$, and polarization coefficient 5×10^{-3}. Once the threshold count $\#\{j : a_j \geq 0.5\}$ is unchanged for three epoch transitions, gates are hard-thresholded and the retained model is fine-tuned for two epochs at 5×10^{-4}.



(a) Learned dimension. (b) Reconstruction–dimension tradeoff.

Figure 7: The gated nonlinear relaxation on MNIST (mean and one standard deviation over ten paired runs).

Before fine-tuning, gated ALRD has MSE 0.0315 ± 0.0014 and accuracy $81.51 \pm 1.36\%$; after fine-tuning these become the values in the table. The effective dimension changes zero times over the final three epoch transitions for every seed.

Table 8: MNIST rank-price path over ten paired runs.

$10^4\gamma$	Dimension	Fine-tuned MSE	Accuracy (%)
1	$23.9 \pm .74$	$.03055 \pm .00077$	82.72 ± 1.07
2	$21.5 \pm .85$	$.03035 \pm .00073$	82.52 ± 1.01
3	18.8 ± 1.62	$.02982 \pm .00133$	$82.10 \pm .95$
4	16.2 ± 1.23	$.02913 \pm .00094$	$81.79 \pm .86$
6	12.9 ± 2.02	$.02926 \pm .00103$	81.15 ± 1.08
8	9.1 ± 1.79	$.03105 \pm .00280$	78.69 ± 2.42
12	5.2 ± 2.25	$.04183 \pm .00793$	63.40 ± 13.75