OBSERVABLE CLOCKS FOR SEQUENTIAL RECOVERY IN TENSOR PCA

Anonymous authors
Paper under double-blind review

ABSTRACT

Online stochastic gradient descent on multi-spiked tensor PCA recovers the planted spikes one at a time. Long plateaus, during which an overlap grows from its random-initialization scale, are followed by rapid recovery. These stages are defined through overlaps with the unknown spikes, so they cannot be observed directly. We show that they can nevertheless be read off online from a quantity SGD already computes. For each estimator column, the contraction of the fresh sample with the current iterate—a rescaled per-sample loss term—tracks the overlap-defined column signal uniformly over polynomially many steps. A windowed threshold on this signal gives an observable stopping rule. From initializations at the random scale with separated greedy weights, it reports the columns in the latent greedy recovery order, and each report falls within a vanishing fraction of its stage of the corresponding certified recovery time. The proof combines a reconstruction of the known sequential-recovery dynamics with new two-sided bounds on stage durations and threshold margins. We further show that the lower bulk of the transverse Riemannian Hessian carries the same signal as a spectral shift, uniformly along the trajectory under non-Gaussian noise. This spectral readout is consistent, but it is a noisier readout of the same quantity rather than an improvement. Simulations illustrate both monitors at finite size.

1 INTRODUCTION

Training by stochastic gradient descent often proceeds in stages: long plateaus during which the loss barely moves are followed by rapid drops as a new direction is learned (Saxe et al., 2014; Abbe et al., 2023). The stages are defined through the unknown target, so a practitioner sees only the loss curve, and whether that curve reveals which component has been learned, and when, usually cannot be checked. We study this question in a model where the latent stages are known exactly.

Sequential recovery in multi-spiked tensor PCA. We observe fresh tensors

$$Y^t = \sqrt{N} \sum_{i=1}^{r} \lambda_i v_i^{\otimes p} + W^t, \qquad t \geq 1, \tag{1}$$

with fixed order $p \geq 3$ and rank $r \geq 2$, unknown orthonormal spikes $v_1, \ldots, v_r \in \mathbb{R}^N$, known weights $\lambda_1 \geq \cdots \geq \lambda_r > 0$, and independent sub-Gaussian noise (Section 2); this is an online version of the spiked tensor model (Montanari & Richard, 2014). Online SGD on the Stiefel manifold $\mathrm{St}(N, r) = \{X \in \mathbb{R}^{N \times r} : X^\top X = I_r\}$ recovers the spikes one at a time (Ben Arous et al., 2026). All overlaps $m_{ij} = \langle v_i, x_j \rangle$ start at order $N^{-1/2}$; a greedily selected overlap grows through a long plateau, becomes macroscopic, and eliminates the overlaps sharing its row and column, after which the next selected overlap grows. The order is fixed at initialization. The same orthogonal tensor structure underlies online moment methods for latent-variable models, where SGD decomposes whitened moment tensors for community detection and topic modeling (Huang et al., 2015); spiked tensor models also describe hypergraph community detection (Kim et al., 2017).

Can the loss report the stages? A running monitor cannot evaluate overlaps with the unknown spikes. *Can fresh observations report the same stages, with the correct order and time scale, without*

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.

access to those directions? Consider

$$U_{j,t} = N^{-1/2}\langle Y^t, x_{j,t-1}^{\otimes p}\rangle = \sum_{i=1}^{r} \lambda_i m_{ij,t-1}^p + N^{-1/2}\langle W^t, x_{j,t-1}^{\otimes p}\rangle. \tag{2}$$

SGD computes $U_{j,t}$ at no extra cost: it is the column-j per-sample loss scaled by $-(\lambda_j\sqrt{N})^{-1}$, equal to the latent column signal plus fresh noise of size $N^{-1/2}$. We show that short-window thresholding answers our question affirmatively. Informally, from any random-scale initialization with separated greedy weights, with probability tending to one, **columns are reported in the latent greedy recovery order, each within $O(h_N^{-1}\log N)$ steps of its stage certification—recovery to $1-\varepsilon$ and cleanup of competing overlaps—a vanishing fraction of the stage duration.** Here $h_N = \delta_N/\sqrt{N}$, with step size δ_N/N. The challenge is dynamical: the latent signal must stay below threshold until shortly before certification and above it just after. Extending a reconstruction of the sequential-recovery argument, we prove these margins and matching two-sided stage-duration bounds.

What the Hessian sees. Hessian spectra probe high-dimensional landscapes and SGD (Ben Arous et al., 2019; Ros et al., 2019; Piccolo, 2023), with emerging outlier eigenspaces suggested as stopping cues (Ben Arous et al., 2025). Analogously to the energy shift in spherical models (Auffinger et al., 2013), the transverse column-j Hessian is shifted by exactly $p\lambda_j U_{j,t}$. Under sub-Gaussian noise, a centered lower-bulk eigenvalue reads this shift from a single fresh sample, tracking the latent column signal uniformly along the adaptive trajectory. The proof combines conditional matrix universality (Brailovskaya & van Handel, 2024), spectral counting, and finite-rank interlacing. Under Gaussian noise, this readout equals $U_{j,t}$ plus an independent spectral error: consistent, but not more accurate.

1.1 Our Contributions

- **The loss as a stage clock.** A windowed threshold on the per-sample contraction reports the latent greedy order and localizes every certified stage time within $O(h_N^{-1}\log N)$ steps (Theorem 3.2).
- **Two-sided stage estimates.** Along the sequential-recovery dynamics we prove lower and upper bounds of order S_N on every stage duration, together with fixed threshold margins before and after certification (Proposition D.1), complementing the upper bounds of Ben Arous et al. (2026).
- **A spectral readout.** The centered lower-bulk eigenvalue of the transverse Hessian tracks the same signal uniformly along predictable trajectories with sub-Gaussian noise (Theorem 3.1), and we give the exact identity relating it to the contraction in (18).

Our guarantees are uniform over a fixed set of separated initializations at the random scale and make no claim about Haar initialization; the independent-noise and fixed-rank assumptions are discussed in Section 5.

Related work. Contractions serve as stopping or estimation statistics for tensor power iteration on a fixed tensor (Anandkumar et al., 2014, Appendix C) and in spike-strength estimation (Huang et al., 2022); Wu & Zhou (2024) give a stopping rule for rank-one power iteration based on consecutive-iterate correlations. These rules decide when a single run has converged. We instead report several successive stages along one continuing SGD trajectory with fresh samples, building on the analysis of high-dimensional online SGD from weak initial correlations (Ben Arous et al., 2021).

2 Setting

2.1 Model and online dynamics

Fix integers $p \geq 3$, $r \geq 2$, weights $\lambda_1 \geq \cdots \geq \lambda_r > 0$, and orthonormal planted directions $V = (v_1, \ldots, v_r) \in \mathrm{St}(N, r)$, with $N \geq 2r$. Write $\lambda_- := \lambda_r$ and $\lambda_+ := \lambda_1$. The raw observations are

$$\widehat{Y}^t = \widehat{W}^t + \sqrt{N}\sum_{i=1}^{r} \lambda_i v_i^{\otimes p}. \tag{3}$$

The weights are known and V is unknown; the numerical value of ν is needed for analytic Hessian centering. Tensor inner products are Frobenius inner products over ordered coordinates. The raw coordinates $\{\widehat{W}^t_{i_1\cdots i_p}\}_{t,i_1,\dots,i_p}$ are i.i.d. sub-Gaussian, centered, have variance $\nu^2 > 0$, and satisfy $\|\widehat{W}^t_{i_1\cdots i_p}\|_{\psi_2} \le K$, where $\|Z\|_{\psi_2} := \inf\{s > 0 : \mathbb{E}\exp(Z^2/s^2) \le 2\}$. Write $Y^t = \mathrm{Sym}(\widehat{Y}^t)$ and $W^t = \mathrm{Sym}(\widehat{W}^t)$. For $X = (x_1, \dots, x_r) \in \mathrm{St}(N, r)$, define the single-sample loss

$$\mathcal{L}(X; Y) = -\sum_{j=1}^{r} \lambda_j \langle Y, x_j^{\otimes p}\rangle. \tag{4}$$

Since $\langle Y, x^{\otimes p}\rangle = \langle \mathrm{Sym}(Y), x^{\otimes p}\rangle$, symmetrization preserves the loss as a function of X, its gradient and Hessian, and the resulting SGD trajectory. The symmetrized coordinates need not be independent, and the raw and symmetrized tensors need not have the same law.

Equip $\mathrm{St}(N, r)$ with the ambient Frobenius metric. For $\Xi \in T_X \mathrm{St}(N, r)$, the polar retraction and online update are

$$R_X(\Xi) = (X + \Xi)(I_r + \Xi^\top \Xi)^{-1/2}, \quad X_t = R_{X_{t-1}}\left(-\frac{\delta_N}{N}\nabla_{\mathrm{St}}\mathcal{L}(X_{t-1}; Y^t)\right). \tag{5}$$

Here δ_N/N is the step size. Fix $C_\delta > 0$, $a_0 < \infty$ and a deterministic $d_{0,N} > 0$ such that $d_{0,N} \downarrow 0$ and $d_{0,N}^{-1} \le N^{a_0}$. For $p = 3$, also assume $d_{0,N} \log N \to 0$, as in the stage estimates used below. We use the dynamical scale

$$\delta_N = C_\delta d_{0,N} (\log N)^{-1} N^{-(p-3)/2}, \qquad h_N := \frac{\delta_N}{\sqrt{N}}, \qquad S_N := d_{0,N}^{-1} N^{p-2} \log N. \tag{6}$$

The quantity S_N is a deterministic reference scale for stage durations, not a stopping time. Define $m_{ij}(X) = \langle v_i, x_j\rangle$ and $m_{ij,t} = m_{ij}(X_t)$. The initialization is independent of the observations. With $\mathcal{F}_{t,N} = \sigma(X_0, \widehat{Y}^1, \dots, \widehat{Y}^t)$, the state X_{t-1} is predictable and W^t is independent of $\mathcal{F}_{t-1,N}$. Denote the joint law by $\mathbb{P}_N$ and the law with deterministic initialization $X_0 = x$ by $\mathbb{P}_N^x$. The planted directions and model parameters are fixed under these laws. All limits take $N \to \infty$ with p, r, λ_i, ν, K fixed.

2.2 Latent order and certified stages

We use the greedy sequential-elimination mechanism of Ben Arous et al. (2026, Theorem 1.5 and Section 5.1) to label the latent recovery stages. For odd p, work with positive overlaps and set $z_{ij}(X) = m_{ij}(X)$. For even $p \ge 4$, freeze the initial signs and set $z_{ij}(X) = \mathrm{sgn}(m_{ij}(X_0))m_{ij}(X)$, assuming nonzero initial overlaps. Define $G_{ij}(X) = \lambda_i\lambda_j z_{ij}(X)^{p-2}$; for even p this is $\lambda_i\lambda_j|m_{ij}(X)|^{p-2}$.

Fix a nonempty initialization set

$$\mathcal{C}_N := \mathcal{C}_1^{\mathrm{init}}(\overline{\gamma}, \underline{\gamma}) \cap \mathcal{C}_2^{\mathrm{init}}(\overline{\gamma}, \gamma_{\mathrm{sep}}), \qquad \overline{\gamma} > \underline{\gamma} > 0, \quad \gamma_{\mathrm{sep}} > 0. \tag{7}$$

The first set bounds every initial z_{ij} between fixed positive multiples of $N^{-1/2}$; the second imposes fixed relative separation of the initial G_{ij}. For even p, membership is checked using $|m_{ij}(X_0)|$ before the signs are frozen. The initialization inequalities and stage-dependent bounds are given in Appendix B.1 and Appendix B.2; uniform statements below concern this fixed set and make no Haar-probability claim.

Starting from $I_0 = J_0 = [r]$, define the unique greedy matching

$$e_k = (i_k^*, j_k^*) := \arg\max_{(i,j)\in I_{k-1}\times J_{k-1}} G_{ij}(X_0), \quad I_k := I_{k-1} \setminus \{i_k^*\}, \quad J_k := J_{k-1} \setminus \{j_k^*\}, \quad k \le r. \tag{8}$$

The deletion is only an analytical convention; SGD continues to update every column.

Fix $\varepsilon_{\mathrm{rec}} \in (0, 1)$ and let $\mathcal{R}_{\le k}^{[k]} = \bigcap_{a\le k} \mathcal{R}_a^{[k]}$ be the recovered-pair condition defined in Appendix B.2. The certified regions are

$$E_k^\sharp = \mathcal{R}_{\le k}^{[k]} \cap \mathcal{C}_{1,k+1}^{\mathrm{act}} \cap \mathcal{C}_{2,k+1}^{\mathrm{gr}} \quad (k < r), \qquad E_r^\sharp = \mathcal{R}_{\le r}^{[r]}. \tag{9}$$

Here the first k selected overlaps have reached $1-\varepsilon_{\mathrm{rec}}$, their row and column cross-overlaps have been cleaned up, and the surviving overlaps retain the conditions needed for later stages. The complete stage-dependent bounds are in Appendix B.2.

Define the latent certified times and stage durations by

$$\tau_0^B = 0, \qquad \tau_k^B := \inf\{t \geq \tau_{k-1}^B : X_t \in E_k^\sharp\}, \quad k \in [r], \qquad \inf \varnothing = \infty. \tag{10}$$

$$D_{k,N} := \tau_k^B - \tau_{k-1}^B \quad \text{when both times are finite.} \tag{11}$$

These times depend on unknown overlaps and serve only as the comparison for observable reports. The exact stage-duration and score-margin event used by the reporting proof is defined in Appendix D.1. The original stage derivations and the fixed-grid bridge are included in the appendix.

Choose a deterministic integer observation budget Q_N before the run, with fixed $B < \infty$ such that

$$Q_N \leq N^B, \qquad Q_N/S_N \longrightarrow \infty. \tag{12}$$

The algorithms process observations through Q_N and terminate.

2.3 Theoretical statistics

For column j immediately before observation t, the unobservable overlap-dependent quantity is

$$F_{j,t} := \sum_{i=1}^{r} \lambda_i m_{ij,t-1}^p. \tag{13}$$

Its direct tensor-contraction statistic is

$$U_{j,t} := \frac{\langle Y^t, x_{j,t-1}^{\otimes p}\rangle}{\sqrt{N}} = \sum_{i=1}^{r} \lambda_i m_{ij,t-1}^p + \frac{\langle W^t, x_{j,t-1}^{\otimes p}\rangle}{\sqrt{N}}. \tag{14}$$

The second statistic is defined from the Hessian quadratic form on the transverse column subspace $\mathcal{H}_j(X) := \{z\mathrm{e}_j^\top : X^\top z = 0\} \subset T_X \operatorname{St}(N,r)$. Put $d_N = N - r$, and let $Q_X \in \mathbb{R}^{N\times d_N}$ have orthonormal columns spanning $\operatorname{span}(X)^\perp$. Under the identification $u \mapsto (Q_X u)\mathrm{e}_j^\top$, the Hessian compression, divided by $\sqrt{N}$, is

$$\mathscr{K}_{j,t} = \frac{p\lambda_j}{\sqrt{N}}\left[\langle Y^t, x^{\otimes p}\rangle I_{d_N} - (p-1)Q_X^\top Y^t(x^{\otimes(p-2)})Q_X\right], \qquad X = X_{t-1}, \quad x = x_{j,t-1}. \tag{15}$$

Here $[Y(x^{\otimes(p-2)})]_{ab} := \sum_{i_3,\dots,i_p} Y_{abi_3\cdots i_p} x_{i_3}\cdots x_{i_p}$. This is a quadratic-form compression; no invariance of $\mathcal{H}_j(X)$ under the Hessian operator is assumed. The compression decomposes as

$$\mathscr{K}_{j,t} = A_{j,t} + p\lambda_j \sum_{i=1}^{r} \lambda_i m_{ij,t-1}^p I_{d_N} - D_{j,t}, \quad D_{j,t} = p(p-1)\lambda_j \sum_{i=1}^{r} \lambda_i m_{ij,t-1}^{p-2} (Q_X^\top v_i)(Q_X^\top v_i)^\top, \tag{16}$$

where $A_{j,t}$ is obtained from (15) by replacing Y^t with W^t. The latent score contributes a scalar shift, whereas $\operatorname{rank}(D_{j,t}) \leq r$. This motivates an eigenvalue beyond the at most r lower outliers created by the finite-rank perturbation.

Let $\mu_1(\mathscr{K}_{j,t}) \leq \mu_2(\mathscr{K}_{j,t}) \leq \cdots \leq \mu_{d_N}(\mathscr{K}_{j,t})$ denote the eigenvalues of $\mathscr{K}_{j,t}$ in increasing order. Choose deterministic integers $r+1 \leq \ell_N \leq d_N$, $\frac{\ell_N}{d_N} \to 0$.

The inequalities are imposed for all sufficiently large N; in particular they require $d_N \geq r+1$. Define

$$S_{j,t} := S_{j,t}^H := \frac{\mu_{\ell_N}(\mathscr{K}_{j,t}) + \rho_{p,j}}{p\lambda_j}, \qquad \rho_{p,j} := 2\nu\lambda_j\sqrt{p(p-1)}. \tag{17}$$

The eigenvalues, and hence $S_{j,t}$, do not depend on the choice of Q_X. Both $U_{j,t}$ and $S_{j,t}$ use the same fresh pair (X_{t-1}, Y^t); no independence between them is required. In fact, writing $M_{j,t} := Q_X^\top Y^t(x_{j,t-1}^{\otimes(p-2)})Q_X$, the increasing eigenvalue convention gives the exact identity

$$S_{j,t} = U_{j,t} + \frac{\rho_{p,j}}{p\lambda_j} - \frac{p-1}{\sqrt{N}}\mu_{d_N+1-\ell_N}(M_{j,t}). \tag{18}$$

The lower-bulk theorem controls the last two terms jointly.

Fixed-window stopping times for the exact scores. Fix one family $T_{j,t} \in \{U_{j,t}, S_{j,t}\}$ for all columns and all times; this choice is not changed during a run.

$$T_{j,t} \in \{U_{j,t}, S_{j,t}\} \quad \text{with the same choice for all } (j,t). \tag{19}$$

Choose a threshold, frequency, and deterministic integer window length before observing data:

$$0 < \theta < \lambda_-(1-\varepsilon_{\mathrm{rec}})^p, \qquad f \in (0,1], \qquad w_N \in \mathbb{N}, \quad h_N w_N \to 0. \tag{20}$$

Put $q_{\max} = \lfloor Q_N / w_N \rfloor$ and define the fixed windows and their endpoints by

$$\mathcal{W}_q := \{(q-1)w_N + 1, \dots, q w_N\}, \qquad t_q := q w_N, \qquad 1 \le q \le q_{\max}. \tag{21}$$

Starting with $R_0^{(T)} = [r]$, at each endpoint form

$$C_q^{(T)} := \left\{ j \in R_{q-1}^{(T)} : \sum_{t \in \mathcal{W}_q} \mathbf{1}\{T_{j,t} \ge \theta\} \ge \lceil f w_N \rceil \right\}. \tag{22}$$

If $C_q^{(T)}$ is nonempty, put $J_q^{(T)} = \min C_q^{(T)}$ and $R_q^{(T)} = R_{q-1}^{(T)} \setminus \{J_q^{(T)}\}$; otherwise put $J_q^{(T)} = 0$ and $R_q^{(T)} = R_{q-1}^{(T)}$. Only one column can be reported at an endpoint; the sample counts are reset at the next window, whether or not a report occurs. For $k \in [r]$ define

$$\begin{aligned} q_k^{(T)} &:= \inf \left\{ 1 \le q \le q_{\max} : \sum_{a=1}^{q} \mathbf{1}\{J_a^{(T)} \ne 0\} \ge k \right\}, \\ (\widehat{\tau}_k^{(T)}, \widehat{j}_k^{(T)}) &:= (t_{q_k^{(T)}}, J_{q_k^{(T)}}^{(T)}), \text{ if } q_k^{(T)} < \infty. \quad \widehat{\tau}_0^{(T)} := 0. \end{aligned} \tag{23}$$

Here $\inf \varnothing = \infty$ and a missing report is encoded by $(\widehat{\tau}_k^{(T)}, \widehat{j}_k^{(T)}) = (\infty, 0)$. Since each $T_{j,t}$ uses (X_{t-1}, Y^t), the event $\{\widehat{\tau}_k^{(T)} \le t\}$ is determined by the first $\lfloor t/w_N \rfloor$ completed windows and belongs to $\mathcal{F}_{t,N}$. Thus $\widehat{\tau}_k^{(T)}$ is an *observable* stopping time; the certified stage times τ_k^B in (10) use unknown overlaps. When T is clear we abbreviate the reported pair by $(\widehat{\tau}_k^{\mathrm{fw}}, \widehat{j}_k^{\mathrm{fw}})$. Theorem 3.1 controls each score relative to (13); Theorem 3.2 compares these report times with the certified times.

2.4 Numerical statistics

Set $h_{j,t} := \mu_{\ell_N}(\mathscr{K}_{j,t})$, using the same spectral index as in (17), and let $\widehat{h}_{j,t}$ denote a numerical approximation of this indexed eigenvalue. We consider three numerical counterparts of $S_{j,t}$:

$$\widehat{S}_{j,t}^{\rho} := \frac{\widehat{h}_{j,t} + \rho_{p,j}}{p\lambda_j}, \quad \widehat{S}_{j,t}^{\mathrm{null}} := \frac{\widehat{h}_{j,t} - \widehat{\zeta}_{j,N}^{0}(X_{t-1})}{p\lambda_j}, \quad \widehat{S}_{j,t}^{\mathrm{burn}} := \frac{\widehat{h}_{j,t} - \widehat{\zeta}_{j,N}^{\mathrm{burn}}}{p\lambda_j}. \tag{24}$$

Analytic centering. The first statistic uses the null edge $-\rho_{p,j}$ and requires known ν but no auxiliary observations.

Independent null centering. For a query state $X = (x_1, \dots, x_r)$ and a symmetric noise-only tensor W, define

$$\mathscr{K}_j^0(X; W) := \frac{p\lambda_j}{\sqrt{N}} \left[\langle W, x_j^{\otimes p} \rangle I_{d_N} - (p-1) Q_X^\top W(x_j^{\otimes(p-2)}) Q_X \right], \tag{25}$$

and

$$h_{j,N}^0(X; W) := \mu_{\ell_N}\left(\mathscr{K}_j^0(X; W)\right). \tag{26}$$

Let $\widehat{h}_{j,N}^0(X; W)$ denote a numerical approximation of this same indexed eigenvalue.

At each query state X, draw independent noise-only tensors $W^{0,1}, \dots, W^{0,B_N^{\mathrm{cal}}}$ with the law of W^t, independently of SGD, and set

$$\widehat{\zeta}_{j,N}^0(X) := \mathrm{Med}\left\{ \widehat{h}_{j,N}^0(X; W^{0,b}) : 1 \le b \le B_N^{\mathrm{cal}} \right\}. \tag{27}$$

Algorithm 1 Fixed-window reporting with direct contraction U

Input: $X_0 \in \mathrm{St}(N, r)$, fresh raw observations through Q_N; known p, r, λ_i; $\delta_N, w_N, \theta, f, Q_N$ as in (6), (12), and (20). **State:** $R = [r]$, counters $c_j = 0$, ordered list $L = ()$.

for $t = 1, \ldots, Q_N$ **do**
  Receive $\widehat{Y}^t$; set $Y^t = \mathrm{Sym}(\widehat{Y}^t)$ and $X = X_{t-1}$.
  for each $j \in R$ **do**
    Compute $U_{j,t} = N^{-1/2}\langle Y^t, x_j^{\otimes p}\rangle$; set $c_j \leftarrow c_j + \mathbf{1}\{U_{j,t} \ge \theta\}$.
  Update all columns to X_t by (5) using the same Y^t.
  if $t \equiv 0 \pmod{w_N}$ **then**
    Set $C_t = \{j \in R : c_j \ge \lceil f w_N \rceil\}$.
    If $C_t \neq \varnothing$, set $j = \min C_t$, append (t, j) to L, and remove j from R.
    Reset all counters to zero, whether or not a report was emitted.
Output: L; set missing report pairs to $(\infty, 0)$.

The median uses a fresh batch at the actual query state. Uniform consistency follows from Lemma D.4 if $\log(rQ_N \vee 2) = o(B_N^{\mathrm{cal}})$ and the null-replicate eigensolver errors vanish uniformly; polynomial auxiliary sample complexity additionally requires $rQ_N B_N^{\mathrm{cal}} \le N^{B'}$ for a fixed B'. This mode needs a full-law noise simulator or independent signal-free observations; ν alone does not suffice.

One-pass burn-in centering. Choose a deterministic calibration length $B_N^{\mathrm{burn}} \in \mathbb{N}$. For the predetermined reporting-window length w_N, define

$$a_N := w_N \left\lceil \frac{B_N^{\mathrm{burn}}}{w_N} \right\rceil. \tag{28}$$

The prefix ends at B_N^{burn} and monitoring begins after the next grid endpoint a_N. Assume

$$B_N^{\mathrm{burn}} \to \infty, \qquad B_N^{\mathrm{burn}} = o(S_N), \qquad a_N + w_N \le Q_N. \tag{29}$$

During this prefix SGD continues; compute $\widehat{h}_{j,t}$ for all j at every step and then freeze the columnwise centers

$$\widehat{\zeta}_{j,N}^{\mathrm{burn}} := \mathrm{Med}\left\{\widehat{h}_{j,t} : 1 \le t \le B_N^{\mathrm{burn}}\right\}. \tag{30}$$

Only full windows after a_N can report; no center is recomputed after a report. Unlike independent-null calibration, the prefix uses dependent, signal-bearing observations. The microscopic-prefix and numerical error conditions for its consistency appear in Theorem D.5 and (474). For even sample sizes, both medians average the two central values. Each numerical score follows (22)– (23). Enlarge $\mathcal{F}_{t,N}$ by auxiliary null draws and solver randomness revealed through time t; the numerical report times are stopping times for this enlarged filtration. In burn-in mode, set $C_q^{(T)} = \varnothing$ for $t_q \le a_N$.

3 Main Results

3.1 Numerical algorithms

Algorithms 1 and 2 implement (23) with direct contraction or one numerical Hessian score. Each score is computed at X_{t-1} using Y^t before the same observation updates all SGD columns. The fixed grid and the reporting decision never interrupt SGD.

The numerical modes use the same window decisions and report-time convention as (23); burn-in windows with $t_q \le a_N$ are inactive. Null calibration and any solver randomness are included in the observation filtration. The score requires the ℓ_Nth *smallest* eigenvalue of the compression (15) on $X_{t-1}^{\perp}$, rather than an eigenvalue of its ambient-space extension. Solver certificates and error conditions are given in Theorem D.5, including (473).

3.2 Theoretical guarantees

Theorem 3.1 (Uniform observable approximation)**.** *Under the observation model, let $X_{t-1} \in \mathrm{St}(N, r)$ be any predictable trajectory. For every fixed $B < \infty$ and integer $Q_N \le N^B$, some*

Algorithm 2 Fixed-window reporting with a numerical Hessian score

Input: X_0, observations and common parameters of Algorithm 1; ℓ_N, an indexed eigensolver configuration, and one fixed choice $\widehat{S} \in \{\widehat{S}^{\rho}, \widehat{S}^{\mathrm{null}}, \widehat{S}^{\mathrm{burn}}\}$. Analytic mode also needs known ν; null mode needs $B_N^{\mathrm{cal}} \in \mathbb{N}$ and independent same-law null data; burn-in mode needs deterministic $B_N^{\mathrm{burn}} \in \mathbb{N}$ as in (29). **State:** $R = [r]$, $c_j = 0$, $L = ()$; set $a_N = 0$, except for burn-in, where $a_N = w_N \lceil B_N^{\mathrm{burn}} / w_N \rceil$.

for $t = 1, \ldots, Q_N$ **do**
  Receive $\widehat{Y}^t$; set $Y^t = \mathrm{Sym}(\widehat{Y}^t)$ and $X = X_{t-1}$.
  if burn-in mode and $t \le B_N^{\mathrm{burn}}$ **then**
    Compute and store $\widehat{h}_{j,t}$ for all j; at $t = B_N^{\mathrm{burn}}$ freeze (30).
  if $t > a_N$ **then for each** $j \in R$
    Compute the indexed eigenvalue approximation $\widehat{h}_{j,t}$ at (X, Y^t).
    In null mode, draw a fresh null batch at (j, X) and compute (27).
    Evaluate the selected score in (24); set $c_j \leftarrow c_j + \mathbf{1}\{\widehat{S}_{j,t} \ge \theta\}$.
  Update all columns to X_t by (5) using the same Y^t.
  if $t \equiv 0 \pmod{w_N}$ **then**
    If $t > a_N$, form $C_t = \{j \in R : c_j \ge \lceil f w_N \rceil\}$; otherwise set $C_t = \varnothing$.
    If $C_t \neq \varnothing$, set $j = \min C_t$, append (t, j) to L, and remove j from R.
    Reset all counters to zero, whether or not a report was emitted.
Output: L; set missing report pairs to $(\infty, 0)$.

$C_B < \infty$ *satisfies*

$$\mathbb{P}_N \left(\max_{j \le r,\, t \le Q_N} |U_{j,t} - F_{j,t}| \le C_B \sqrt{\frac{\log N}{N}} \right) \longrightarrow 1. \tag{31}$$

Under the spectral index conditions in (17)*, for every fixed* $\epsilon > 0$*,*

$$\mathbb{P}_N \left(\max_{j \le r,\, t \le Q_N} \max\{|S_{j,t} - F_{j,t}|,\ |S_{j,t} - U_{j,t}|\} \le \epsilon \right) \longrightarrow 1. \tag{32}$$

Both conclusions hold uniformly over predictable states and deterministic initializations, without a recovery or SGD step assumption.

These bounds apply to each predictable trajectory, not simultaneously to all states on one noise realization. Fixed admissible spectral indices are covered by the appendix's uniform bulk and lower-edge estimates via rank-r interlacing. Proposition D.1 establishes, with high probability, the stage-duration bounds $D_{k,N} \asymp S_N$ and threshold margins used in the reporting theorem below.

Write $\mathcal{E}_N^{\mathrm{fin}} := \{0 = \tau_0^B < \cdots < \tau_r^B \le Q_N\} \cap \{0 = \widehat{\tau}_0^{\mathrm{fw}} < \cdots < \widehat{\tau}_r^{\mathrm{fw}} \le Q_N\}$.

Theorem 3.2 (Observable order and recovery clocks)**.** *Assume the observation model, fixed* $r \ge 2$*, initialization supported on the fixed set* $\mathcal{C}_N$*, and* (6)*,* (12)*, and* (20)*. Fix* $T \in \{U, S\}$ *for the entire run and apply the grid* (21) *and the reporting rules of Algorithm 1 with* T *in place of* U*. The choice* S *additionally requires the lower-bulk index conditions. For some fixed* $C < \infty$ *depending on the initialization set and fixed parameters,*

$$\mathbb{P}_N \left(\mathcal{E}_N^{\mathrm{fin}} \cap \left\{ \widehat{j}_k^{\mathrm{fw}} = j_k^*, \quad -C h_N^{-1} \log N \le \widehat{\tau}_k^{\mathrm{fw}} - \tau_k^B \le 2 w_N \quad \textit{for all } k \le r \right\} \right) \longrightarrow 1. \tag{33}$$

In particular, the report times agree with their latent stage times on the scale $D_{k,N} \asymp S_N$*, and, for every fixed* $\epsilon > 0$*,*

$$\mathbb{P}_N \left(\mathcal{E}_N^{\mathrm{fin}} \cap \left\{ \max_{k \le r} \left| \frac{\widehat{\tau}_k^{\mathrm{fw}} - \widehat{\tau}_{k-1}^{\mathrm{fw}}}{\tau_k^B - \tau_{k-1}^B} - 1 \right| \le \epsilon \right\} \right) \longrightarrow 1. \tag{34}$$

The conclusions hold uniformly over $\mathbb{P}_N^x$ *for* $x \in \mathcal{C}_N$*.*

Reports identify recovering columns, not planted labels, and may precede certification. Stage conditions use the positive-overlap convention for odd p and the frozen-sign convention of Section 2.2 for even p; the latter preserves observable sums. Guarantees are uniform over the stated initialization set, with no unconditional Haar clock claim.

N	p=3	Ref.	U	S^{ρ}	S^{burn}	S^{null}	p=4	Ref.	U	S^{ρ}	S^{burn}	S^{null}
32		29	29	15	1	27		29	27	15	3	27
64		28	25	18	1	25		28	26	13	12	26
128		29	28	19	7	28		25	25	15	24	25
256		30	30	27	19	30		28	28	22	28	28
512		30	30	25	29	30		30	30	28	30	30
Total		146	142	104	57	140		140	136	93	97	136

Table 1: **Complete column-order agreement.** Each row pools ten seeds at each of $r = 3, 7, 11$: 30 trials per degree, or 150 in Total. Ref. counts complete latent-reference sequences; monitor entries count complete sequences matching that reference exactly. Missing reports count as nonmatches.

Numerical extension. Under Theorem 3.2, Algorithm 2 inherits the order and clock guarantees under the joint law $\overline{\mathbb{P}}_N$ (including calibration and solver randomness) provided indexed-eigenvalue and center errors vanish uniformly over monitored calls in probability. Burn-in additionally requires (29) and uniformly accurate calibration-prefix solves. See Lemma D.4 for exact independent-null center consistency and Theorem D.5 for the burn-in argument, even-p signed-prefix condition, and solver error budgets. Uniformity over $\mathcal{C}_N$ requires uniform numerical error probabilities. Sparse probes, growing rank, and optimized Hessian computation are not covered.

Roles of the four monitors. Among the four monitors, direct contraction U attains the best or tied-best complete column-order match counts in the reported aggregate comparisons of Section 4. The role of the Hessian statistic is a spectral characterization: its lower-bulk eigenvalue tracks $F_{j,t}$ uniformly along predictable trajectories under non-Gaussian noise and finite-rank deformation. Both reporting guarantees use the same deterministic stage-margin argument. The contraction monitor uses (31), whereas the Hessian monitor uses (32); each then applies Proposition D.3 independently. The contraction guarantee requires no spectral approximation or calibration.

4 Numerical Experiments

We test whether observable monitors reproduce the column order and timing of a latent overlap-based reference on the *same* SGD trajectory. The comparisons measure reporting fidelity, not faster optimization or certified full recovery.

Experimental setup. We use $p \in \{3, 4\}$, $N \in \{32, 64, 128, 256, 512\}$ and $r \in \{3, 7, 11\}$, with 10 random seeds per configuration. In (3), draw V from Haar measure and use fresh Gaussian coordinate noise with $\nu^2 = 1$. Gaussian noise is a sub-Gaussian special case; non-Gaussian robustness is not tested. Set planted strengths and column weights to $\lambda_i = 8 - 6(i-1)/(r-1)$ and initialize $X_0 = \mathrm{QR}_+(Z + 0.22\sqrt{N}V)$, where Z is standard Gaussian and QR_+ uses a positive R-diagonal. Run the polar-retracted update (5) for $Q_N = 400$ observations with $\delta_N = 0.15$ (actual step $0.15/N$). Reporting does not alter these updates.

Monitors and reference. All monitors use nonoverlapping windows $w_N = 8$, threshold $\theta = 0.6$ and frequency $f = 0.75$, requiring six scores satisfying $T_{j,t} \geq \theta$. At each endpoint, report the smallest eligible column index, remove it from that monitor, and reset all remaining counts, even when no report occurs. We compare U with the Hessian scores S^{ρ}, S^{burn} and S^{null} defined in (24)–(30). Their centers are, respectively, the analytic lower edge $-\rho_j$, a columnwise median frozen after eight observations, and an average over two independent noise-only replicates per active query. Use $\ell_N = r + \lceil N^{1/4} \rceil$ for the ℓ_Nth smallest eigenvalue of the $(N-r)$-dimensional compression; $r + 1 \leq \ell_N \leq N - r$ throughout the grid. The latent reference applies identical reporting rules to $F_{j,t} = \sum_i \lambda_i m_{ij,t-1}^p$. Its outputs $(\tau_k^{(m)}, j_k^{(m)})$ are evaluation-only quantities, not the certified clocks τ_k^B or planted labels.

Order agreement. Table 1 gives U the highest or tied-highest aggregate match count: $142/150$ for $p = 3$ and $136/150$ for $p = 4$. At $N = 512$, both U and null centering match in all 30 trials per degree, versus 25 and 28 for analytic centering.

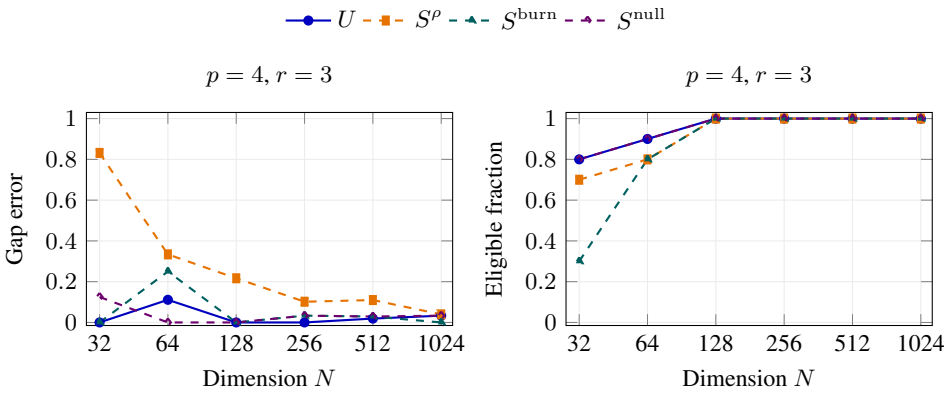


Figure 1: **Timing fidelity and eligibility.** Left: mean $E_s^{(T)}$ conditional on complete identical column orders. Right: eligible fraction among ten attempts. Bands are pointwise 95% bootstrap and Wilson intervals, respectively.

Timing and uncertainty. Figure 1 uses $p = 4, r = 3$ and the same dimension grid, extended to $N = 1024$, with other settings unchanged. The smaller dimensions reuse Table 1's seeds and do not add independent trials. For each complete, exactly order-aligned paired trial s, define

$$R_{k,N}^{(T)} = \frac{\widehat{\tau}_k^{(T)} - \widehat{\tau}_{k-1}^{(T)}}{\tau_k^{(m)} - \tau_{k-1}^{(m)}}, \qquad E_s^{(T)} = \max_{k \le r} |R_{k,N}^{(T)} - 1|, \tag{35}$$

with time-zero times set to zero. We average trial maxima rather than pooled gap medians, which can hide errors under window quantization. Timing error is undefined for ineligible trials; their frequency is reported separately using all ten attempts. Pointwise 95% intervals use 20,000 seed-bootstrap resamples (seed 94621) for mean error and Wilson intervals for eligibility.

Analytic-centering error decreases from 0.832 at $N = 32$ (7/10 eligible) to 0.040 at $N = 1024$ (10/10), but increases from 0.102 to 0.110 between $N = 256$ and 512. At $N = 32$, U and burn-in have zero conditional error but only 8/10 and 3/10 eligible trials. At $N = 1024$, mean $\pm$ sample SD is 0.033 ± 0.105, 0.040 ± 0.105, 0.000 ± 0.000 and 0.033 ± 0.105 for $U, S^{\rho}, S^{\text{burn}}, S^{\text{null}}$, respectively, all with 10/10 eligibility.

5 Conclusion

We establish online reporting clocks for the sequential recovery of spikes in multi-spiked tensor PCA. The per-sample contraction, which SGD already computes, tracks the overlap-based signal of each estimator column. A centered lower-bulk Hessian statistic provides a complementary spectral readout of that same signal. Under the stated initialization, scaling, and stage-separation conditions, uniform observation bounds and stage-duration estimates show that a fixed-window rule reports recovering columns in the latent greedy order and matches certified inter-stage durations to first order. The recovery stages can therefore be monitored without access to the planted spikes or interruption of SGD.

The finite-size Gaussian experiments support this reporting behavior: direct contraction achieves the highest or tied-highest aggregate order-match counts among the evaluated monitors. The rank-50 experiment in Appendix A also shows why the distinction between dominant-span progress, individual recovery, and accurate alarms matters. A report signals progress through a recovery stage; it does not certify a prescribed reconstruction accuracy or accelerate learning. Extending these clocks to growing rank and dependent observations, and reducing the cost of indexed Hessian computation, are natural next steps.

STATEMENT ON THE USE OF AI

Generative AI tools assisted with language and structural editing, literature and citation checking, drafting and critical review of parts of the theoretical exposition and proofs, and implementation and plotting of the numerical experiments. The authors are responsible for the reported results, proofs, and conclusions.

REFERENCES

Emmanuel Abbe, Enric Boix-Adserà, and Theodor Misiakiewicz. SGD learning on neural networks: Leap complexity and saddle-to-saddle dynamics. In *Proceedings of the Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 2552–2623. PMLR, 2023.

Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15 (80):2773–2832, 2014.

Antonio Auffinger, Gérard Ben Arous, and Jiří Černý. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66:165–201, 2013. doi: 10.1002/cpa.21422.

Gérard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72:2282–2330, 2019. doi: 10.1002/cpa.21861.

Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Online stochastic gradient descent on non-convex losses from high-dimensional inference. *Journal of Machine Learning Research*, 22 (106):1–51, 2021.

Gérard Ben Arous, Reza Gheissari, Jiaoyang Huang, and Aukosh Jagannath. Spectral alignment of stochastic gradient descent for high-dimensional classification tasks. *The Annals of Applied Probability*, 35(4):2767–2822, 2025. doi: 10.1214/25-AAP2187.

Gérard Ben Arous, Cédric Gerbelot, and Vanessa Piccolo. Stochastic gradient descent in high dimensions for multi-spiked tensor PCA. *Communications on Pure and Applied Mathematics*, 79 (10):2291–2369, 2026. doi: 10.1002/cpa.70056.

Tatiana Brailovskaya and Ramon van Handel. Universality and sharp matrix concentration inequalities. *Geometric and Functional Analysis*, 34(6):1734–1838, 2024. doi: 10.1007/s00039-024-00692-9.

Furong Huang, U. N. Niranjan, Mohammad Umar Hakeem, and Animashree Anandkumar. Online tensor methods for learning latent variable models. *Journal of Machine Learning Research*, 16 (86):2797–2835, 2015.

Jiaoyang Huang, Daniel Z. Huang, Qing Yang, and Guang Cheng. Power iteration for tensor PCA. *Journal of Machine Learning Research*, 23(128):1–47, 2022.

A. A. Kholopov, A. N. Tikhomirov, and D. A. Timushev. Rate of convergence to the semicircle law for the Gaussian Orthogonal Ensemble. *Theory of Probability & Its Applications*, 52(1):171–177, 2008. doi: 10.1137/S0040585X97982906. Translated from Teoriya Veroyatnostei i ee Primeneniya 52(1), 180–185 (2007).

Chiheon Kim, Afonso S. Bandeira, and Michel X. Goemans. Community detection in hypergraphs, spiked tensor models, and sum-of-squares. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pp. 124–128. IEEE, 2017. doi: 10.1109/SAMPTA.2017.8024470.

Michel Ledoux and Brian Rider. Small deviations for beta ensembles. *Electronic Journal of Probability*, 15(41):1319–1343, 2010. doi: 10.1214/EJP.v15-798.

Andrea Montanari and Emile Richard. A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems 27*, 2014.

Vanessa Piccolo. Topological complexity of spiked random polynomials and finite-rank spherical integrals, 2023.

Valentina Ros, Gérard Ben Arous, Giulio Biroli, and Chiara Cammarota. Complex energy landscapes in spiked-tensor and simple glassy models: Ruggedness, arrangements of local minima, and phase transitions. *Physical Review X*, 9:011003, 2019. doi: 10.1103/PhysRevX.9.011003.

Andrew M. Saxe, James L. McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. In *International Conference on Learning Representations*, 2014.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1–2):1–230, 2015. doi: 10.1561/2200000048.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018. ISBN 978-1-108-41519-4. doi: 10.1017/9781108231596.

Yuchen Wu and Kangjie Zhou. Sharp analysis of power iteration for tensor PCA. *Journal of Machine Learning Research*, 25(195):1–42, 2024.

A Additional Experiments: Dominant-Direction Signals and Monitor Sensitivity

Purpose, settings and provenance. We revisit the rank-50 illustration to distinguish dominant-span coverage from individual recovery and accurate reports. The available later cohort contains five seeds $2027, \ldots, 2031$ per degree $p = 3, 4$, with $N = 200$, $r = 50$, Haar initialization, Gaussian raw noise $\nu = 1$, $\lambda_{1:5} = 8$ and $\lambda_{6:50} = 0.4$ for both planted and column weights. Use $\delta_N = 0.15$, $Q_N = 200$, $w_N = 8$, $\theta = 0.2$ and $f = 0.75$; the update and reporting rules are unchanged. Only U and the latent-window comparator are evaluated, not Hessian monitors. All ten paths were replayed and their saved occupancies, normalized losses, events and score windows checked. Replaying a seed adds diagnostics, not an independent trial.

Three complementary measurements. With $M_t = V^\top X_t$, define $\mathrm{Occ}_5(t) = \sum_{i\le 5, j\le 50} m_{ij,t}^2/5$; the weak occupancy averages the remaining 45 rows. The Haar expectation for either is $r/N = 0.25$. Let $R_\alpha(t)$ be the largest number of distinct strong directions that can be injectively matched to distinct columns with $|m_{ij,t}| \ge \alpha$. This prevents duplicate columns and rotated subspace coverage from being counted as separate recoveries. For each emitted report (t, j), assign its evaluation-only label $i^* = \arg\max_i |m_{ij,t}|$. It is useful at cutoff α if $i^* \le 5$ and $|m_{i^*j,t}| \ge \alpha$. $P_\alpha(t)$ is the fraction of emitted alarms that are useful; $C_\alpha(t)$ is the fraction of the five distinct strong labels usefully reported. Repeated labels do not inflate coverage. Precision before the first alarm is undefined. The cutoffs 0.5 and 0.95 apply to *absolute*, not squared, overlap.

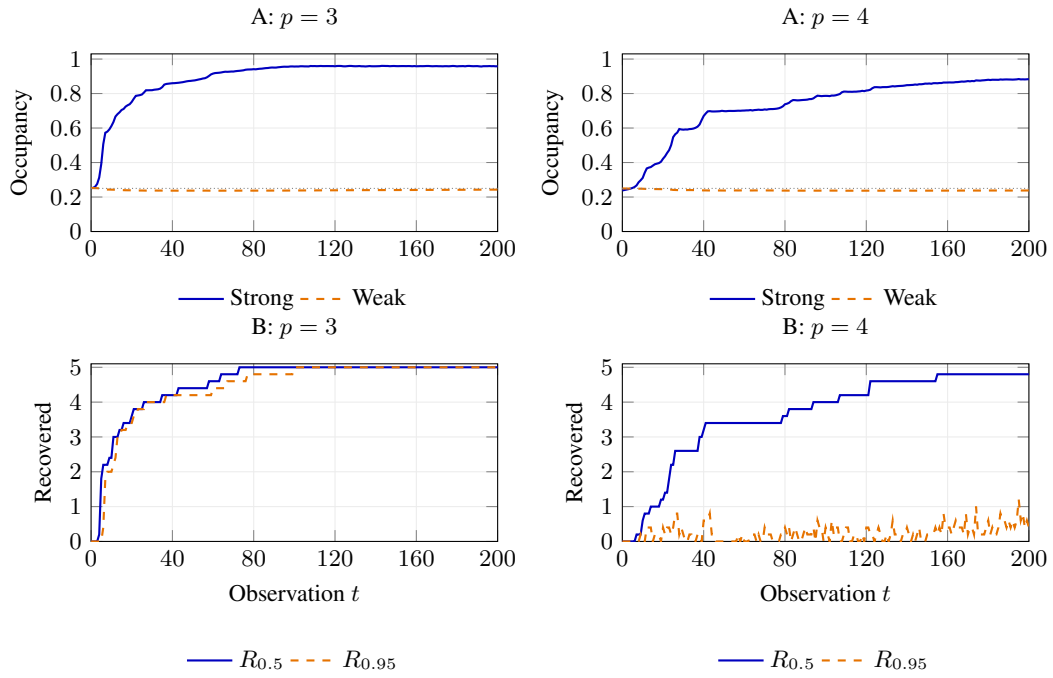


Figure 2: **Coverage is not a recovery or alarm certificate.** Five-seed later cohort, $N = 200, r = 50$, with $p = 3$ (left) and $p = 4$ (right). Row A shows strong/weak span occupancy (dotted Haar reference 0.25); row B shows injectively matched strong components at cutoffs $0.5/0.95$. Scores use X_{t-1} and fresh observation t; recovery diagnostics use X_t. Curves are seed means; shading gives pointwise 95% seed-bootstrap intervals (10,000 resamples, seed 94621); alarm precision and distinct-label coverage, with uncertainty, appear in Table 2. Settings and provenance are specified above.

Executed results and failure case. Table 2 reports cross-seed mean ± sample SD. At $t = 200$, $p = 3$ recovers all five strong components at absolute overlap 0.95, whereas $p = 4$ has occupancy 0.884 but only 0.40 strict matches on average. No baseline $p = 4$ alarm meets the strict cutoff. At the more permissive 0.5 cutoff, final $p = 4$ precision and distinct-label coverage are 0.96 ± 0.09 and 0.92 ± 0.11. These differences are not numerical replay failures: they expose the distinction between span coverage, component alignment and report-time accuracy.

p	t	Occ_5	$R_{0.95}$	$P_{0.95}$	$C_{0.95}$
3	40	0.860 ± 0.060	4.20 ± 0.45	1.00 ± 0.00	0.80 ± 0.00
3	200	0.958 ± 0.004	5.00 ± 0.00	0.82 ± 0.18	0.96 ± 0.09
4	40	0.670 ± 0.065	0.60 ± 0.55	0.00 ± 0.00	0.00 ± 0.00
4	200	0.884 ± 0.056	0.40 ± 0.55	0.00 ± 0.00	0.00 ± 0.00

Table 2: **Strict-cutoff audit.** Mean ± sample SD over five seeds per degree. $R_{0.95}$ is a count out of five; $P_{0.95}$ and $C_{0.95}$ are fractions. At these checkpoints all five seeds have emitted an alarm, so precision has five eligible observations. High occupancy can coexist with few strict component matches and inaccurate strict-cutoff alarms.

An unfavorable example is $p = 4$, seed 2030: the report at $t = 120$ for column 34 has largest planted overlap 0.482 and fails cutoff 0.5. Its U and latent-score windows both have eight qualifying observations; observable-score noise alone does not explain it. The threshold $\theta = 0.2$ is a progress-score threshold, not an accuracy threshold: already $8(0.45)^4 = 0.328 > 0.2$ in a single-strong-direction example.

Executed paired monitor ablation. Holding all ten recorded SGD paths fixed, reprocess their complete pre-update U arrays with $w_N \in \{4, 8, 16\}$ at $\theta = 0.2$, and $\theta \in \{0.1, 0.2, 0.4\}$ at $w_N = 8$, retaining $f = 0.75$, the tie/removal/reset rules and $Q_N = 200$. This post hoc one-factor-at-a-time reporting ablation adds no training or out-of-sample evidence. Only complete windows report: $w_N = 16$ ends at $t = 192$, leaving eight observations without an extra window. Each condition has five paired seeds per degree.

		$p = 3$		$p = 4$	
w_N	θ	$P_{0.5}$	$C_{0.5}$	$P_{0.5}$	$C_{0.5}$
8	0.2	0.86 ± 0.19	0.96 ± 0.09	0.96 ± 0.09	0.92 ± 0.11
4	0.2	0.80 ± 0.22	0.96 ± 0.09	0.92 ± 0.11	0.88 ± 0.11
16	0.2	0.93 ± 0.09	1.00 ± 0.00	1.00 ± 0.00	0.96 ± 0.09
8	0.1	0.81 ± 0.20	0.96 ± 0.09	0.92 ± 0.11	0.88 ± 0.11
8	0.4	0.90 ± 0.14	0.96 ± 0.09	1.00 ± 0.00	0.96 ± 0.09

Table 3: **Sensitivity of reporting, not learning.** Final useful-cutoff precision $P_{0.5}$ and distinct-label coverage $C_{0.5}$, mean ± sample SD over five reused seeds per degree; row one is the unchanged baseline. All conditions share identical SGD paths.

A.1 Random-scale reporting and external baselines

Construction and score coupling. For the new $p = r = 3$ runs, take $V = [I_r, 0]^\top$; rotational invariance of the Gaussian raw tensor makes this equivalent in law to rotating V. Draw C_{ij} independently and uniformly from $[0.28, 1.32]$, set $M = C/\sqrt{N}$, and reject until $\|M^\top M\|_{\mathrm{op}} < 0.8$, the greedily chosen column order is not $(1, 2, 3)$, and the best available initial weight $G_{ij} = 9m_{ij,0}$ exceeds the runner-up by a factor 1.1 at each row-and-column deletion. With an independent orthonormal $Z_\perp \in \mathbb{R}^{(N-r)\times r}$, set $X_0 = \begin{bmatrix} M \\ Z_\perp (I - M^\top M)^{1/2} \end{bmatrix}$. Thus $X_0^\top X_0 = I_r$ and every initial overlap is positive and of order $N^{-1/2}$. This rejection only enforces a selected leader/runner-up margin; it does not verify every separation and stage condition in $\mathcal{C}_N$.

For each fresh Gaussian observation and current column x_j, draw independent $\xi_j \sim N(0,1)$ and $z_j \sim N(0, I_N)$ and set

$$a_j = \sqrt{N} \sum_i \lambda_i m_{ij,t-1}^2 v_i + \xi_j x_j + (I - x_j x_j^\top) z_j / \sqrt{3}, \qquad U_{j,t} = x_j^\top a_j / \sqrt{N}. \tag{36}$$

The a_j have the exact joint conditional law of the order-three contractions $Y^t(\cdot, x_j, x_j)$ for orthonormal X. We use $H_j = -3\lambda_j a_j$, its Stiefel tangent projection, and the polar retraction in (5). This shortcut is used only for Gaussian noise; the Rademacher experiment below samples an explicit raw tensor. Independent seeds are 24100–24111 at both dimensions, $\nu = 1$, all spike and column weights equal to 3, $\delta_N = 0.108/(\log N)^3$, $w = 16$, $\theta = 1.5$, $f = 0.75$ and $Q = 150{,}000$ or $450{,}000$ for $N = 64$ or 128. With $d_{0,N} = 0.108/(\log N)^2$ and $C_\delta = 1$, these fixed caps are about $3.5S_N$ and $3.3S_N$. Runs stop early only after all three planted pairs meet the cleanup diagnostic and all three monitors complete; missing events retain the full horizon. The requested full-step trajectory of seed 24100 is saved alongside per-trial records.

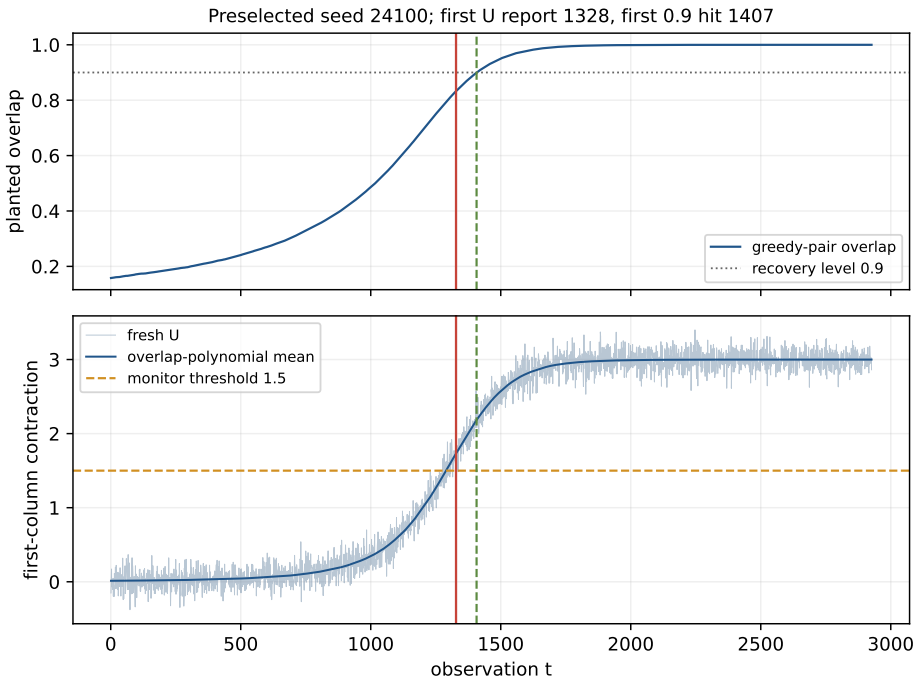


Figure 3: **Plateau and first report from a random-scale start.** The preselected seed 24100 at $N = 64, p = r = 3$ shows the first greedy-pair overlap, the fresh contraction U, its overlap-polynomial mean, and the threshold 1.5. The first U report is at step 1328; that overlap first reaches 0.9 at step 1407. These are finite-size diagnostics, not the certified τ_1^B.

Evaluation clocks and monitor baselines. For $e_k = (i_k^*, j_k^*)$, let

$$\beta_k^{0.9} := \inf\{t \geq 1 : m_{i_k^* j_k^*, t} \geq 0.9\}. \tag{37}$$

Define the cleanup *proxy* $\widetilde{\tau}_k$ as the first time all first k selected overlaps reach 0.9 and every other overlap sharing one of their selected rows or columns has magnitude at most 0.15. It omits the surviving microscopic bands and future greedy conditions of $E_k^\sharp$, and is not τ_k^B. A score at time t uses X_{t-1}, whereas these diagnostics use X_t.

The paper's U rule counts at least 12 of 16 scores above 1.5 and selects the smallest eligible column. We also report (i) a separate random-eligible tie ablation of that same rule; (ii) an exponential moving average $E_{j,0} = 0$, $E_{j,t} = (2/17)U_{j,t} + (15/17)E_{j,t-1}$, eligible when $E_{j,t} \geq 1.5$ at a window endpoint; and (iii) a window-queried Page CUSUM $C_{j,0} = 0$, $C_{j,t} = \max\{0, C_{j,t-1} + U_{j,t} - 0.75\}$,

eligible when $C_{j,t} \geq 9$ at an endpoint. Each monitor has its own remaining column set and reports at most one column per complete window, using the smallest index for ties except in the explicit random-tie ablation. These thresholds were specified before the twelve seeds; they do not equate the methods' false-alarm rates. Counts are reset between fixed windows; EMA and CUSUM states continue, including after a column is reported.

N	monitor	exact order	median $\widehat{\tau} - \beta^{0.9}$	adjacent ratio	alarm overlap
64	U	$11/12$	-86.5	1.004	0.820
64	EMA	$11/12$	-94	1.004	0.810
64	CUSUM	$11/12$	-182	1.017	0.709
128	U	$12/12$	-219	1.002	0.808
128	EMA	$12/12$	-232.5	1.003	0.800
128	CUSUM	$12/12$	-444.5	1.008	0.684

Table 4: **New random-scale cohort, twelve seeds per size.** Medians of raw report-minus-hit steps pool 34 matched events at $N = 64$ and 36 at $N = 128$; adjacent ratios use only the 11 or 12 complete order-aligned runs (22 or 24 gaps). Alarm overlap is for the greedy row matched to the reported column. The $N = 64$ order-mismatched trial contributes one matched $\beta^{0.9}$ pair and two additional alarms. These diagnostics are not the certified clocks τ_k^B.

At $N = 64$, the complete order-aligned U runs have adjacent $\beta^{0.9}$ gaps between 104 and 5923 steps; at $N = 128$ the range is 346–19,241. The corresponding U gap-ratio ranges are $[0.877, 1.385]$ and $[0.878, 1.156]$. The U report-minus-$\beta^{0.9}$ ranges are $[-153, -70]$ and $[-328, -191]$ steps. All 36 U alarms at each dimension are below 0.9 in their matched planted overlap; the erroneous alarm in the order-mismatched $N = 64$ trial has overlap 0.150 with its greedy row; its third alarm has matched overlap near zero. The Wilson 95% interval for $11/12$ is $[0.646, 0.985]$, so these small cohorts do not establish a dimension trend. No completed window has multiple eligible columns for U, and its random-tie ablation gives identical outputs. The numerical scales $h_N^{-1} \log N$ are 22,181 and 58,114, while S_N is 42,627 and 135,381. The two median U leads divided by $h_N^{-1} \log N$ are -0.00390 and -0.00377; divided by S_N they are -0.00203 and -0.00162. Median adjacent $\beta^{0.9}$ gaps are 802.5 and 2481.5 steps, or $0.0188 S_N$ and $0.0183 S_N$. These are finite-size diagnostics against $\beta^{0.9}$, not a measurement of $D_{k,N}$ or its asymptotic law.

Unconditional Haar diagnostic. A separate $N = 64$, $r = p = 3$ cohort of twelve unconditioned Haar X_0 with the same $Q = 150{,}000$ cap and other parameters (seeds 25100–25111) has $11/12$ full $\beta^{0.9}$ recoveries, $11/12$ complete U reports, and $10/12$ exact complete column orders relative to the exploratory positive-weight greedy matching. For this signed cohort the exploratory ranking uses $9 \max\{m_{ij,0}, 0\}$ with a deterministic numerical tie convention; it is not the theorem's greedy order for negative overlaps. Only $4/12$ trials have positive initial overlaps on all three selected greedy pairs. A negative selected overlap can change sign over this long finite run, but for odd p the initial negative overlaps fall outside the paper's positive-branch stage analysis.

A.2 Non-Gaussian raw noise and spectral finite-size effects

Explicit tensor paths. To check the actual raw-coordinate model without a Gaussian shortcut, we draw one N^3 tensor of independent standard Gaussian or symmetric Rademacher entries per SGD step, symmetrize over all six permutations, and contract the *same* sample for U and the gradient. At each $N = 16, 24$, forty paired initial states have $r = 2, p = 3$, $\lambda = (4, 3)$, positive $m_{ij,0} \in [0.5, 1.5]/\sqrt{N}$, and initial leader/runner-up ratio at least 1.1; the two noise laws use independent paths from the same initial V, X_0. Set $\delta_N = 0.1$, $Q = 800$, $(w, \theta, f) = (8, 0.6, 0.75)$.

N	raw noise	complete greedy order	95% Wilson interval	RMS $\left(U_{j,t} - \sum_i \lambda_i m_{ij,t-1}^3\right)$
16	Gaussian	$37/40$	$[0.801, 0.974]$	0.2486
16	Rademacher	$33/40$	$[0.681, 0.913]$	0.2477
24	Gaussian	$34/40$	$[0.709, 0.929]$	0.2044
24	Rademacher	$34/40$	$[0.709, 0.929]$	0.2043

Table 5: **Explicit raw tensor paths.** All four groups complete both reports in $40/40$ trials. The order column compares with the positive-overlap greedy matching at initialization; the RMS pools all steps and columns within a trial, then averages over trials.

An $N = 8, p = 3$ explicit-Gaussian check with 2500 independent tensors finds at most 6.2% relative error in the ten checked conditional score/gradient/Hessian variances and largest off-diagonal sample covariance 0.029. A finite-difference check of the raw-tensor Euclidean gradient and one polar step is included in the saved validation record.

A non-Gaussian Hessian check at one fixed state. Holding the same microscopic $N = 24, r = 2, p = 3$ state fixed, sample 160 independent raw tensors for each law and compute both U and the indexed transverse compression from that tensor (320 pooled column-samples per law). The latent column signals are 0.036 and 0.081 at the selected state. For $\ell = 3 = r + 1$, the empirical mean $\pm$ sample SD of $S^\rho - \sum_i \lambda_i m_{ij}^3$ is 0.453 ± 0.229 under Gaussian and 0.447 ± 0.231 under Rademacher entries. At $\ell = 5 = r + \lceil N^{1/4} \rceil$, these become 0.777 ± 0.220 and 0.775 ± 0.216. The corresponding single-score crossings of $\theta = 0.6$ are 118/320 and 119/320 at $\ell = 3$, and 278/320 and 280/320 at $\ell = 5$; U crosses once in 320 scores for each noise law.

Gaussian spectral-index sweep. In the exact conditional Gaussian fixed-state model with $r = 11$, $\lambda_j = 5$, $\nu = 1$, we sample 100 independent GOE fluctuations per dimension and degree and keep the same draw across indices. The table reports the *null-state* mean $\pm$ sample SD of $S^\rho - U$; an aligned overlap 0.2 with its rank-one compression perturbation is also saved, and retains the large finite-index bias.

		$p = 3$: ℓ choice			$p = 4$: default
N	default ℓ/d	$r + 1$	$r + \lceil N^{1/4} \rceil$	$r + \lceil N^{1/2} \rceil$	$r + \lceil N^{1/4} \rceil$
32	$14/21$	1.724 ± 0.072	1.940 ± 0.071	2.269 ± 0.081	2.057 ± 0.076
64	$14/53$	0.941 ± 0.035	1.036 ± 0.033	1.273 ± 0.033	1.098 ± 0.035
128	$15/117$	0.550 ± 0.021	0.635 ± 0.021	0.841 ± 0.019	0.673 ± 0.022
256	$15/245$	0.331 ± 0.014	0.383 ± 0.013	0.563 ± 0.012	0.406 ± 0.014
512	$16/501$	0.201 ± 0.009	0.244 ± 0.008	0.399 ± 0.007	0.258 ± 0.008

Table 6: **Finite spectral-index bias, fixed-state Gaussian law.** Each cell is the empirical mean $\pm$ sample SD of $S^\rho - U$ from 100 independent draws. At $N = 32, r = 11$, the old default index is 14 out of $d = 21$, deep inside the finite bulk.

At null state, if q_α is the lower semicircle quantile of mass α, the corresponding leading finite-d prediction is

$$\mathbb{E}[S^\rho - U] \;\simeq\; \frac{2\nu\lambda_j\sqrt{p(p-1)} + \nu\lambda_j\sqrt{p(p-1)d/N}\, q_{\ell/d}}{p\lambda_j}, \qquad d = N - r. \tag{38}$$

The saved quantile predictions are within about 0.02 of the observed means in this sweep. For $\ell/d \ll 1$ the quantile's lower-edge displacement scales as $(\ell/d)^{2/3}$; the finite d/N factor and finite-rank perturbations also matter at the smallest sizes.

Incremental computation time. With one BLAS thread, five repetitions and $p = r = 3$, the median time per fixed-state dense Gaussian proxy query is:

N	U scalar transform	GOE construction	Indexed eigensolve and shift
256	0.000022	0.78	1.72
512	0.000022	3.65	8.64
1024	0.000022	15.29	54.08

Table 7: **Kernel timing (milliseconds/query).** The U column times only scalar rescaling of an already available contraction (10,000 repetitions within each timing run); the Hessian columns separately time dense GOE construction and indexed eigensolve plus scaling. These are CPU kernel costs, not end-to-end SGD or GPU wall times.

At $N = 1024$, constructing the matrix takes a median 15.29 ms and the eigensolve with shift 54.08 ms per column query; the direct score is already computed by the SGD loss.

B RECONSTRUCTION OF SEQUENTIAL RECOVERY

B.1 MODEL AND SEQUENTIAL-ELIMINATION BACKGROUND

Fix $p \geq 3$ and $r \geq 2$, with p and r independent of N. Let $v_1, \dots, v_r \in \mathbb{R}^N$ be orthonormal and let $\lambda_1 \geq \cdots \geq \lambda_r > 0$ be fixed. We observe fresh tensors

$$Y^t = \sqrt{N} \sum_{i=1}^{r} \lambda_i v_i^{\otimes p} + W^t, \qquad W^t = \mathrm{Sym}(\widehat{W}^t), \qquad t \geq 1, \tag{39}$$

where the raw coordinates of $\widehat{W}^t$ are independent, centered, have variance ν^2, and have uniformly bounded sub-Gaussian norms. The raw tensors are independent across t. For $X = (x_1, \dots, x_r) \in \mathrm{St}(N, r)$, set

$$\mathcal{L}(X; Y) = -\sum_{j=1}^{r} \lambda_j \langle Y, x_j^{\otimes p} \rangle, \qquad m_{ij}(X) = \langle v_i, x_j \rangle. \tag{40}$$

The online iterate uses the Stiefel gradient and polar retraction:

$$X_t = R_{X_{t-1}}\left(-\frac{\delta_N}{N} \nabla_{\mathrm{St}} \mathcal{L}(X_{t-1}; Y^t) \right), \qquad R_X(H) = (X + H)(I_r + H^\top H)^{-1/2}. \tag{41}$$

Write $m_{ij,t} = m_{ij}(X_t)$.

We recall the sequential-elimination rule of Ben Arous et al. (2026, Definition 1.4 and Eq. (1.9)). In the positive-overlap regime, its initial pair weights are

$$G_{ij}(X) = \lambda_i \lambda_j m_{ij}(X)^{p-2}. \tag{42}$$

Starting with $I_0 = J_0 = [r]$ and $A_0 = I_0 \times J_0$, choose the pairs once from X_0, for $k = 1, \dots, r$, by

$$e_k = (i_k^*, j_k^*) := \operatorname*{arg\,max}_{(i,j) \in A_{k-1}} G_{ij}(X_0), \quad I_k := I_{k-1} \setminus \{i_k^*\}, \quad J_k := J_{k-1} \setminus \{j_k^*\}, A_k := I_k \times J_k. \tag{43}$$

Thus each selected pair removes one row and one column. The ordering in (43) is determined at time zero.

The deterministic sets appearing in Ben Arous et al. (2026, Definitions 3.1 and 3.3) are, in this notation,

$$\mathcal{C}_1^{\mathrm{init}} := \left\{ X \in \mathrm{St}(N, r) : \frac{\underline{\gamma}}{\sqrt{N}} \leq m_{ij}(X) < \frac{\overline{\gamma}}{\sqrt{N}} \quad \text{for all } i, j \right\}, \tag{44}$$

$$\mathcal{C}_2^{\mathrm{init}} := \left\{ X \in \mathrm{St}(N, r) : \left| \frac{G_{ij}(X)}{G_{ab}(X)} - 1 \right| > \frac{\gamma_{\mathrm{sep}}}{\overline{\gamma}} \quad \text{for all } (i,j) \neq (a,b) \right\}. \tag{45}$$

Here $\overline{\gamma} > \max\{\underline{\gamma}, \gamma_{\mathrm{sep}}\} > 0$, and ratios are evaluated where their denominators are nonzero. Put $\mathcal{C}_N := \mathcal{C}_1^{\mathrm{init}} \cap \mathcal{C}_2^{\mathrm{init}}$. On $\mathcal{C}_N$ every initial overlap is positive and each maximizer in (43) is unique.

B.2 RECONSTRUCTED SEQUENTIAL RECOVERY

We reconstruct the overlap-stage argument for the model (39)–(41). We present the argument for positive overlaps; for even p, we work with overlap magnitudes under the frozen-sign convention of Section 2.2, with the corresponding sign adjustments in the proof.

Fix $\varepsilon_{\mathrm{rec}} \in (0, 1)$ and write $\chi_N = (\log N)^{-1/2} N^{-(p-1)/4}$. The construction gives positive, N-independent stage constants $\gamma_{2,k} < \gamma_{1,k}$, $C_{\chi,k}$, and $\Theta_k > 1$ for $k = 1, \dots, r$, with $\gamma_{2,1} = \underline{\gamma}$, $\gamma_{1,1} = \overline{\gamma}$, and $\Theta_1 = 1 + \gamma_{\mathrm{sep}}/\overline{\gamma}$. For $1 \leq k \leq r$, let

$$\mathcal{C}_{1,k}^{\mathrm{act}} := \left\{ X : \frac{\gamma_{2,k}}{\sqrt{N}} \leq m_{ij}(X) \leq \frac{\gamma_{1,k}}{\sqrt{N}} \quad \text{for all } (i,j) \in A_{k-1} \right\}, \tag{46}$$

$$\mathcal{C}_{2,k}^{\mathrm{gr}} := \bigcap_{\ell=k}^{r} \bigcap_{f \in A_{\ell-1} \setminus \{e_\ell\}} \left\{ X : \frac{G_{e_\ell}(X)}{G_f(X)} \geq \Theta_k \right\}. \tag{47}$$

The ratios in (47) are used on the positive active band. For $1 \le a \le s \le r$, put

$$\mathcal{R}_a^{[s]} := \left\{ X : \begin{array}{c} m_{i_a^* j_a^*}(X) \ge 1 - \varepsilon_{\mathrm{rec}}, \\ |m_{i_a^*, j}(X)| \le C_{\chi,s}\chi_N \quad (j \ne j_a^*), \\ |m_{i, j_a^*}(X)| \le C_{\chi,s}\chi_N \quad (i \ne i_a^*) \end{array} \right\}. \tag{48}$$

The certified stage sets and their successive entrance times are

$$\begin{aligned} E_s^\sharp &:= \left(\bigcap_{a=1}^{s} \mathcal{R}_a^{[s]} \right) \cap \mathcal{C}_{1,s+1}^{\mathrm{act}} \cap \mathcal{C}_{2,s+1}^{\mathrm{gr}}, \qquad 1 \le s < r, \\ E_r^\sharp &:= \bigcap_{a=1}^{r} \mathcal{R}_a^{[r]}, \\ \tau_0^B &:= 0, \qquad \tau_k^B := \inf\{ t \ge \tau_{k-1}^B : X_t \in E_k^\sharp \}, \quad 1 \le k \le r. \end{aligned} \tag{49}$$

An empty intersection is the whole state space; an empty hitting-time set has infimum ∞.

Theorem B.1 (Reconstructed sequential recovery)**.** *Suppose $\mathcal{C}_N$ is nonempty for all sufficiently large N, $X_0 \in \mathcal{C}_N$, and*

$$\delta_N = \frac{C_\delta d_{0,N}}{\log N} N^{-(p-3)/2}, \qquad d_{0,N} \downarrow 0, \qquad d_{0,N}^{-1} \le N^{a_0}, \tag{50}$$

for fixed $C_\delta > 0$ and $a_0 < \infty$; when $p = 3$, also assume $d_{0,N} \log N \to 0$. Set $S_N = d_{0,N}^{-1} (\log N) N^{p-2}$. The stage constants above can be chosen so that there are $C_1, \dots, C_r < \infty$ and $\eta_N \to 0$ satisfying

$$\begin{gathered} \mathcal{G}_N^{\mathrm{rec}} := \{ 0 = \tau_0^B < \tau_1^B < \cdots < \tau_r^B < \infty \} \cap \bigcap_{k=1}^{r} \{ X_{\tau_k^B} \in E_k^\sharp,\ \tau_k^B - \tau_{k-1}^B \le C_k S_N \}, \\ \inf_{X_0 \in \mathcal{C}_N} \mathbb{P}_{X_0}(\mathcal{G}_N^{\mathrm{rec}}) \ge 1 - \eta_N. \end{gathered} \tag{51}$$

In particular, the fixed greedy pairs $e_1, \dots, e_r$ are certified in their selected order at $\tau_1^B, \dots, \tau_r^B$. If $V = (v_1, \dots, v_r)$ and P_π is the permutation matrix defined by $\pi(j_k^*) = i_k^*$, then, on the same event,

$$\tau_r^B \le \left(\sum_{k=1}^{r} C_k \right) S_N, \qquad \left\| X_{\tau_r^B} - V P_\pi \right\|_F^2 \le 2r\varepsilon_{\mathrm{rec}}. \tag{52}$$

B.3 Detailed reconstruction argument

The following estimates spell out the concentration, polar retraction, microscopic race, macroscopic growth, cleanup, and stage induction used for Theorem B.1. We give deterministic parameter choices and conditional failure probabilities so the same calculations apply at the random certified restart times.

Provenance of the reconstruction. The sequential-elimination argument follows Ben Arous et al. (2026). The specific result numbers in the comparisons below refer to its arXiv v2 version (arXiv:2410.18162v2). The cross-overlap scale χ_N is already present in the recovery target of their Theorem 3.5. Lemma B.3 records gradient, Gram, polar-retraction, and martingale estimates corresponding to their Lemmas 4.1, 4.2, 4.4, and 4.5. The scalar comparison in Lemma B.10 follows the discrete-comparison approach of their Appendix A. We give self-contained proofs in the present normalization, recording the constants, candidate-exit bounds, and conditional failure probabilities needed by Proposition D.1. The stage-dependent tube and block estimates below specify the quantitative bounds used at each restart.

Proof constants and stage notation

Use the model and update in (39)– (41), the weights and greedy matching in (42)– (43), and the initialization sets (44)–(45). Throughout the proof set $d := p - 2$, $X_{\mathrm{in}} := X_0$, $Y_t := Y^t$, $W_t := W^t$,

and $\varepsilon_{\mathrm{rec}} := \varepsilon_{\mathrm{rec}}$; X_{in} always denotes the global initialization, whereas a later stage starts at a stopping-time state X_s. For a pair $f = (i, j)$ write $m_f := m_{ij}$ and $G_f := G_{ij}$, and put $H_t(X) := \mathcal{L}(X; W_t)$. Set $\lambda_- := \min_a \lambda_a$ and $\lambda_+ := \max_a \lambda_a$. Choose a fixed K such that $\|\widehat{W}^t_\alpha\|_{\psi_2} \le K$ for every t and $\alpha \in [N]^p$, where $\|Z\|_{\psi_2} := \inf\{s > 0 : \mathbb{E}e^{Z^2/s^2} \le 2\}$. The step size is the one in (50). The following constants make all subsequent probability and drift bounds uniform over the specified initial and restart states.

$$\Lambda_2 := \left(\sum_{a=1}^{r} \lambda_a^2\right)^{1/2}, \qquad L_A := p\Lambda_2, \qquad C_\Phi := p\Lambda_2^2, \tag{53}$$

$$\alpha_1 := 2pK\lambda_+(1 + \sqrt{r}), \tag{54}$$

$$T_G := \sqrt{8}\, K L_A \sqrt{r \log 5 + 3}, \qquad B_G := 4T_G^2 = 32K^2 L_A^2 (r \log 5 + 3), \tag{55}$$

$$\alpha_2 := 2C_\Phi^2 + 2B_G, \qquad c_R := 2. \tag{56}$$

Set also

$$C_G := r^{1/4}\sqrt{\alpha_2}; \tag{57}$$

on the Gram event, $\|\nabla_{\mathrm{St}}\mathcal{L}\|_F \le C_G\sqrt{N}$. Define the explicit polynomial coefficient bounds

$$C_D := p\lambda_+^2(1 + r^2), \tag{58}$$

$$C_{D,1} := p\lambda_+^2[(p-1) + r^2(p+1)], \tag{59}$$

$$C_{D,2} := p\lambda_+^2[(p-1)(p-2) + r^2 p(p+1)]. \tag{60}$$

For the active/old monomial split, use

$$C_{\mathrm{act}} := pr^2\lambda_+^2, \qquad C_{\mathrm{old}} := 2pr^2\lambda_+^2. \tag{61}$$

Define also

$$c_p^{(1)} := \frac{1}{2} - \frac{p-1}{2p}(2p)^{-1/(p-1)}, \tag{62}$$

$$c_p^{(2)} := 1 - \frac{1}{2p}\left(\frac{p-1}{p}\right)^{p-1}, \tag{63}$$

$$c_p := \min\{c_p^{(1)}, c_p^{(2)}\}, \qquad K_\lambda := 1 + \left(\frac{\lambda_+}{\lambda_-}\right)^{2/p}. \tag{64}$$

Set

$$C_{\mathrm{coup}} := K_\lambda^2 C_{D,2}. \tag{65}$$

The constants $C_D, C_{D,1}, C_{D,2}$ bound, respectively, $|D_{ij}(X)|/\sqrt{N}$ and the entrywise ℓ^∞ operator norms of its first and second derivatives on $[-1, 1]^{r\times r}$.

The greedy labels $e_k = (i_k^*, j_k^*)$ and active rectangles A_k are fixed once at X_{in} by (43). The all-pairs condition (45) is needed only at this initialization; subsequent stages retain the directed comparisons between each prescribed future leader and its competitors. Nonleaders may tie after the first recovery.

Separation budget and the five gamma families Define

$$B_0 := \left(\frac{1 + \bar{c}}{1 - \bar{c}}\right)^{p-1}, \qquad \Theta_1 := 1 + \frac{\gamma_{\mathrm{sep}}}{\overline{\gamma}}, \tag{66}$$

where, with $q_0 := \Theta_1^{1/[2r(p-1)]}$, take explicitly

$$\bar{c} := \min\left\{\frac{1}{4}, \frac{q_0 - 1}{q_0 + 1}\right\}. \tag{67}$$

Then $0 < \bar{c} < 1/2$ and

$$\Theta_1 > B_0^r. \tag{68}$$

This is not an extra random assumption on the realized initialization. Indeed, if

$$R_{\mathrm{in}}(X_{\mathrm{in}}) := \min_{1\le k\le r-1} \min_{f\in A_{k-1}\setminus\{e_k\}} \frac{G_{e_k}(X_{\mathrm{in}})}{G_f(X_{\mathrm{in}})}, \tag{69}$$

then (45) and maximality of e_k imply

$$R_{\mathrm{in}}(X_{\mathrm{in}}) \geq \Theta_1 > B_0^r. \tag{70}$$

Thus the proof uses the deterministic lower budget Θ_1, not a new hypothesis $R_{\mathrm{in}} > B_0^r$. For $1 \leq k \leq r$, define

$$\Theta_k := \Theta_1 B_0^{-(k-1)}. \tag{71}$$

Then $\Theta_k > B_0 > 1$ for every $k \leq r$. Set

$$\gamma_{1,1} := \overline{\gamma}, \qquad \gamma_{2,1} := \underline{\gamma}, \qquad \gamma_1 := \gamma_{\mathrm{sep}}, \tag{72}$$

and, at every stage, encode the directed gap by

$$\gamma_k := (\Theta_k - 1)\gamma_{1,k}, \qquad 1 + \frac{\gamma_k}{\gamma_{1,k}} = \Theta_k. \tag{73}$$

There is no need to impose $\gamma_k < \gamma_{1,k}$; that restriction would be inconsistent whenever $\Theta_k \geq 2$. For $c \in (0,1)$, put

$$A(c) := \frac{1-c}{1+c}, \qquad B(c) := \left(\frac{1+c}{1-c}\right)^{p-1}. \tag{74}$$

Define

$$q_{\mathrm{ord}} := \min\left\{B_0^{1/(4d)}, \Theta_r^{1/[2(2p-3)]}, \frac{1+\bar{c}}{1-\bar{c}}\right\}, \qquad c_{\mathrm{ord}} := \frac{q_{\mathrm{ord}}-1}{q_{\mathrm{ord}}+1}. \tag{75}$$

Then $c_{\mathrm{ord}} \in (0, \bar{c}]$ and

$$A(c_{\mathrm{ord}})^{2d} > B_0^{-1}, \qquad A(c_{\mathrm{ord}})^d \Theta_r > B(c_{\mathrm{ord}}). \tag{76}$$

Indeed $q_{\mathrm{ord}} > 1$, while the half-power margins in (75) imply both strict inequalities directly. Only now choose the macroscopic threshold, explicitly, by

$$\varepsilon_{\mathrm{mac}} := \min\left\{\frac{1-\varepsilon_{\mathrm{rec}}}{4}, \frac{1}{8}, \sqrt{\frac{c_{\mathrm{ord}} p \lambda_-^2}{128 C_D}}\right\}. \tag{77}$$

This order avoids circularity: c_{ord} depends only on the initial separation budget, whereas every later drift margin may depend on $\varepsilon_{\mathrm{mac}}$. Put

$$C_\lambda := \frac{\max_{i,j} \lambda_i \lambda_j}{\min_{i,j} \lambda_i \lambda_j} < \infty. \tag{78}$$

Before choosing ρ_k, fix the two constants that control the entire Regime-II duration:

$$\kappa_{II,k} := \frac{1}{4} p \lambda_-^2 (1 - \varepsilon_{\mathrm{mac}}^2) \left(\frac{16}{17}\right)^{p-1}, \tag{79}$$

$$C_{T,k} := \frac{8(16/15)^{p-2}}{\kappa_{II,k}(1 - 2^{-(p-2)})}. \tag{80}$$

They depend only on constants already fixed; in particular neither depends on ρ_k or $\gamma_{+,k}$. Suppose $\gamma_{1,k}, \gamma_{2,k}, \Theta_k$ are fixed. Choose

$$\rho_k := \min\left\{\frac{1}{4}, (2C_\lambda B_0)^{-1/d}, \frac{c_{\mathrm{ord}} p \lambda_-^2}{64 C_D}, \frac{1}{4}\left(\frac{1}{48 C_D C_{T,k}}\right)^{1/(p-2)}\right\}. \tag{81}$$

This is made before $\gamma_{+,k}$ and gives the explicit strict Regime-II absorption margins used below. Put

$$\gamma_{-,k} := \frac{1-\bar{c}}{2}\gamma_{2,k}. \tag{82}$$

Let $\mathfrak{R}_k := \Theta_k / B_0 > 1$ and $[x]_+ := \max\{x, 0\}$. If $k < r$, set

$$\gamma_{+,k} := 1 + \max\left\{\left[\frac{\mathfrak{R}_k (1+\bar{c})^d \rho_k^{-d} \gamma_{1,k}^d - (1-\bar{c})^d \gamma_{2,k}^d}{\mathfrak{R}_k - 1}\right]_+^{1/d}, \frac{(1+\bar{c})\gamma_{1,k}}{\rho_k}\right\}. \tag{83}$$

For the singleton final stage set

$$\gamma_{+,r} := 1 + (1 + \bar{c})\gamma_{1,r}. \tag{84}$$

After $\gamma_{+,k}$ has been fixed, put

$$c_8 := \max\left\{8, \frac{2}{1-\bar{c}}\right\}. \tag{85}$$

Then define

$$S_{1,k} := \frac{2(16/15)^{p-1}\varepsilon_{\mathrm{mac}}}{\kappa_{II,k}}, \qquad S_{2,k} := \varepsilon_{\mathrm{mac}} S_{1,k}, \tag{86}$$

$$C_{0,k} := c_8 K_\lambda \exp(C_D S_{2,k})(1 + C_D S_{1,k})\left(1 + \rho_k + \frac{\gamma_{+,k}}{\gamma_{-,k}}\right). \tag{87}$$

Thus no later enlargement of $C_{0,k}$ is permitted. Define

$$L_k^{\mathrm{in}} := C_{0,k}\gamma_{+,k}, \qquad L_k^{\mathrm{out}} := 2K_\lambda L_k^{\mathrm{in}} + 8. \tag{88}$$

For $k < r$, define the next stage by

$$\gamma_{2,k+1} := \frac{\gamma_{-,k}}{4C_{0,k}}, \qquad \gamma_{1,k+1} := 2C_{0,k}\gamma_{+,k}, \qquad \gamma_{k+1} := (\Theta_{k+1} - 1)\gamma_{1,k+1}. \tag{89}$$

All five families are now determined stage by stage. Since r is fixed, all of them remain positive, finite, and independent of N.

Stage-local reserves Use the active band and directed set defined in (46)– (47), and the scale χ_N in (48). Empty intersections impose no condition. Only the initial set (45) requires separation between every pair. Set $C_{\chi,0} := 2$.

$$\kappa_0 := \frac{1}{2} p c_p \lambda_-^2 \min\{(1-\varepsilon_{\mathrm{rec}})^p, \min\{1-\varepsilon_{\mathrm{rec}}, \varepsilon_{\mathrm{mac}}/2\}^p\}, \tag{90}$$

$$B_\chi := 1 + \frac{3K_\lambda \alpha_2 C_\delta}{2\kappa_0}\sqrt{1 + \alpha_2 C_\delta^2}. \tag{91}$$

Recursively, for $1 \le s \le r$, put

$$C_{\chi,s}^{\mathrm{in}} := 8K_\lambda^2 C_{\chi,s-1} + 8B_\chi, \qquad C_{\chi,s} := 2K_\lambda C_{\chi,s}^{\mathrm{in}} + 8B_\chi. \tag{92}$$

These formulas provide the inner-to-outer slack for stability. After the recursion has fixed all cross radii, put

$$U_k^{\mathrm{mic}} := 2 + L_k^{\mathrm{out}} + C_{\chi,k}, \qquad U_k^{\mathrm{cl}} := 2 + L_k^{\mathrm{out}} + C_{\chi,k} + \mathbf{1}_{\{k<r\}}\gamma_{1,k+1}. \tag{93}$$

Certified restart notation The recovery tubes and sharp stage sets are (48) and (49). For $0 \le q \le s$, write $\mathcal{R}_{\le q}^{[s]} := \bigcap_{a=1}^{q} \mathcal{R}_a^{[s]}$, with the empty intersection interpreted as the whole state space. When a dependence on $\varepsilon_{\mathrm{rec}}$ is displayed, it refers to the same tubes and sets. Put $E_0^\sharp := \mathcal{C}_{1,1}^{\mathrm{act}} \cap \mathcal{C}_{2,1}^{\mathrm{gr}}$ and, for notational uniformity only, $\mathcal{C}_{1,r+1}^{\mathrm{act}} = \mathcal{C}_{2,r+1}^{\mathrm{gr}} = \mathrm{St}(N,r)$; no stage-$r+1$ constants are introduced. For a measurable set A and a stopping time s, use

Symbol	Role	Update / defining relation
$\gamma_{2,k}$	Active lower amplitude	$\gamma_{2,k+1} = \gamma_{-,k}/(4C_{0,k})$
$\gamma_{1,k}$	Active upper amplitude	$\gamma_{1,k+1} = 2C_{0,k}\gamma_{+,k}$
γ_k	Directed greedy gap	$1 + \gamma_k/\gamma_{1,k} = \Theta_k$
$\gamma_{-,k}$	Microscopic lower guard	$\gamma_{-,k} = (1-\bar{c})\gamma_{2,k}/2$
$\gamma_{+,k}$	Leader selection amplitude	Explicit clock margin (83)

Table 8: The five stage-indexed gamma families. In particular, $\gamma_{+,k}$ is a leader selection threshold, not the survivor upper endpoint.

$$T_A^{[s]} := \inf\{t \ge s : X_t \in A\}, \qquad T_A := T_A^{[0]}. \tag{94}$$

The symbol $\tau_k^\sharp$ below is an alias for the certified hitting time τ_k^B in (49), including at random restart times. The individual threshold time

$$\beta_k^{\mathrm{rec}} := \inf\{t \geq \tau_{k-1}^\sharp : m_{e_k}(X_t) \geq 1 - \varepsilon_{\mathrm{rec}}\} \tag{95}$$

is different. Always $\beta_k^{\mathrm{rec}} \leq \tau_k^\sharp$, but equality need not hold: at β_k^{rec}, the new row and column may not yet be cleaned and the next survivor band and ordering may not yet be certified. Here, *recovery of spike* k means certified recovery at $\tau_k^\sharp$.

Transition inputs stated here and proved below

Stopped microscopic comparison. Let σ be a bounded stopping time, let Q_N be deterministic, and suppose every relevant positive overlap at σ is at least $g/\sqrt{N}$, where $g > 0$ is fixed. Put $h_N := \delta_N/\sqrt{N}$. For each fixed $c \in (0, 1/2)$, uniformly over the restart state, there is an event of conditional probability at least $1 - \eta_{\mathrm{cmp,N}}$, with the stage-specific value given in (181), on which, before the prescribed bootstrap exit,

$$m_f(X_{\sigma+n}) \geq (1-c)m_f(X_\sigma)[1 - pdh_N(1-c)^{p-1}G_f(X_\sigma)n]^{-1/d}, \tag{96}$$

$$m_f(X_{\sigma+n}) \leq (1+c)m_f(X_\sigma)[1 - pdh_N(1+c)^{p-1}G_f(X_\sigma)n]^{-1/d}. \tag{97}$$

The denominators being used are required to be positive. In particular, if $G_e(X_\sigma)/G_f(X_\sigma) > B(c)$, division gives

$$\frac{G_e(X_{\sigma+n})}{G_f(X_{\sigma+n})} \geq A(c)^d \frac{G_e(X_\sigma)}{G_f(X_\sigma)}. \tag{98}$$

Exact overlap recursion. On the regularity event,

$$m_{ij}(X_{t+1}) = m_{ij}(X_t) + \frac{\delta_N}{N} D_{ij}(X_t) + \xi_{ij}(t+1) + r_{ij}(t+1), \tag{99}$$

where

$$D_{ij}(X) = \sqrt{N}p\lambda_i\lambda_j m_{ij}^{p-1} - \frac{\sqrt{N}p}{2} \sum_{a,b=1}^{r} \lambda_a m_{ib} m_{aj} m_{ab} (\lambda_j m_{aj}^{p-2} + \lambda_b m_{ab}^{p-2}). \tag{100}$$

Conditionally on $\mathcal{F}_t$, $\xi_{ij}(t+1)$ is centered and obeys $\mathbb{E}[e^{\theta\xi_{ij}(t+1)} \mid \mathcal{F}_t] \leq \exp(\theta^2\alpha_1^2\delta_N^2/(2N^2))$. The retraction remainder has the explicit bound proved below. Consequently, for predictable indicators I_t,

$$\mathbb{P}\left(\sup_{n \leq H} \left| \sum_{u=1}^{n} I_{\sigma+u-1}\xi_{ij}(\sigma+u) \right| > a \,\middle|\, \mathcal{F}_\sigma \right) \leq 2\exp\left(-\frac{a^2N^2}{2\alpha_1^2\delta_N^2 H} \right). \tag{101}$$

First-stage cleanup estimates. On the stopped first-stage macroscopic tube, with $q = m_{e_1}$, selected row/column coordinates $z^{\mathrm{row}}, z^{\mathrm{col}}$, and survivors u_f, direct expansion of (100) gives

$$D_q = \sqrt{N}p\lambda_{i_1^*}\lambda_{j_1^*}q^{p-1}(1-q^2) + E_q, \qquad |E_q| \leq \tfrac{1}{2}C_{D,2}(U_1^{\mathrm{cl}})^2 N^{-1/2}, \tag{102}$$

$$D_j^{\mathrm{row}} = -\sqrt{N}\kappa_j^{\mathrm{row}} z_j^{\mathrm{row}} + R_j^{\mathrm{row}}, \qquad \kappa_j^{\mathrm{row}} \geq \kappa_{\mathrm{cl}} > 0, \tag{103}$$

$$D_i^{\mathrm{col}} = -\sqrt{N}\kappa_i^{\mathrm{col}} z_i^{\mathrm{col}} + G_i^{\times} + R_i^{\mathrm{col}}, \qquad |G_i^{\times}| \leq C_{D,1} Z_{\mathrm{row}}. \tag{104}$$

The coupled maximum-norm recursion, blockwise noise estimates, and (101) imply, for

$$H_{1,N}^{\mathrm{cl}} := \left\lceil A_1^{\mathrm{cl}} \frac{\sqrt{N}}{\delta_N} \log N \right\rceil, \tag{105}$$

that the leader is recovered, its row and column contract below $C_{\chi,1}^{\mathrm{out}}\chi_N$, and the deterministic quantity $\omega_{1,N}$ in (341) tends to zero such that

$$\sup_{f \in A_1} \sup_{n \leq H_{1,N}^{\mathrm{cl}}} |m_f(X_{\sigma+n}) - m_f(X_\sigma)| \leq \frac{\omega_{1,N}}{\sqrt{N}}. \tag{106}$$

All these bounds hold uniformly conditionally at a capped macroscopic restart, with failure probability at most $\eta_{1,N}^{\mathrm{cl}}$.

Proof of the stochastic and deterministic input estimates This section proves every input used by the transition argument. Write $a_N^{\mathrm{step}} := \delta_N/N$, $h_N := \delta_N/\sqrt{N}$, and $M(X) := V^{\mathsf{T}}X = (m_{ij}(X))$. Since both V and X have orthonormal columns,

$$\|M(X)\|_{\mathrm{op}} \le 1, \qquad |m_{ij}(X)| \le 1. \tag{107}$$

Fresh-noise concentration and the exact polar step

Lemma B.2 (The scalar moment-generating bound with the stated convention)**.** *Let Z be centered and satisfy $\mathbb{E}\exp(Z^2/K^2) \le 2$. Then, for every $\theta \in \mathbb{R}$,*

$$\mathbb{E}e^{\theta Z} \le e^{2K^2\theta^2}. \tag{108}$$

Consequently, for independent centered variables Z_α satisfying the same bound and deterministic real coefficients a_α,

$$\mathbb{E}\exp\left(\theta\sum_\alpha a_\alpha Z_\alpha\right) \le \exp\left(2K^2\theta^2\sum_\alpha a_\alpha^2\right). \tag{109}$$

Proof. The elementary inequality $e^u - u \le e^{u^2}$ holds for every real u. Indeed, $f(u) = e^{u^2} - e^u + u$ satisfies $f(0) = f'(0) = 0$ and

$$f''(u) = (2+4u^2)e^{u^2} - e^u \ge e^u(2e^{-1/4} - 1) > 0.$$

Put $s = |\theta|K$. If $s \le 1$, the preceding inequality $e^u - u \le e^{u^2}$, centering, and Jensen's inequality for the concave power $x \mapsto x^{s^2}$ give

$$\mathbb{E}e^{\theta Z} \le \mathbb{E}e^{\theta^2 Z^2} \le \left(\mathbb{E}e^{Z^2/K^2}\right)^{s^2} \le 2^{s^2} \le e^{2s^2}.$$

If $s > 1$, the inequality $\theta Z \le Z^2/K^2 + s^2/4$ gives

$$\mathbb{E}e^{\theta Z} \le 2e^{s^2/4} \le e^{2s^2}.$$

This proves (108); multiplying the individual moment-generating bounds proves (109). If the hypothesis is initially stated as $\|Z\|_{\psi_2} \le K$, the displayed exponential-square bound follows by applying the definition at $K + \epsilon$ and then using Fatou's lemma as $\epsilon \downarrow 0$. □

Role in the reconstruction. The linear-combination estimate is applied in Lemma B.3 to the symmetrized tensor noise. It supplies the uniform conditional noise bound needed at stopping-time restarts.

Lemma B.3. ***Coefficient-explicit one-step stochastic estimates***

For every predictable $X_t \in \mathrm{St}(N,r)$, every $i,j \le r$, and every $\theta \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\{\theta\langle v_i, (\nabla_{\mathrm{St}}H_{t+1}(X_t))_j\rangle\} \mid \mathcal{F}_t\right] \le \exp(\theta^2\alpha_1^2/2). \tag{110}$$

Moreover, for every deterministic T,

$$\mathbb{P}_N\left(\max_{1\le s\le T}\|G(X_{s-1};Y_s)\|_F > \alpha_2 N\right) \le Te^{-c_R N}, \tag{111}$$

where $G(X;Y) = \nabla_{\mathrm{St}}\mathcal{L}(X;Y)^{\mathsf{T}}\nabla_{\mathrm{St}}\mathcal{L}(X;Y)$. On the complementary Gram event in (111)*, the exact update obeys*

$$m_{ij}(X_{t+1}) = m_{ij}(X_t) + a_N^{\mathrm{step}}D_{ij}(X_t) + \xi_{ij}(t+1) + r_{ij}(t+1), \tag{112}$$

where

$$\mathbb{E}[e^{\theta\xi_{ij}(t+1)} \mid \mathcal{F}_t] \le \exp\left(\frac{\theta^2\alpha_1^2\delta_N^2}{2N^2}\right), \tag{113}$$

$$|r_{ij}(t+1)| \le b_N^{\mathrm{pol}} := \frac{\alpha_2\delta_N^2}{2N}\sqrt{1 + \frac{\alpha_2\delta_N^2}{N}}. \tag{114}$$

The population drift is exactly (100)*, and*

$$\max_{i,j}|D_{ij}(X)| \le C_D\sqrt{N}. \tag{115}$$

Proof. For $X \in \mathrm{St}(N,r)$ and an $N \times r$ matrix A, write

$$\Pi_X A = A - X\,\mathrm{Sym}(X^\mathsf{T} A).$$

This is the orthogonal projection for the Frobenius inner product onto the embedded Stiefel tangent space. It is self-adjoint and contractive. Fix a realized past, so that the predictable state $X = X_t$ is fixed, and let

$$A_j(W) = -p\lambda_j W(\,\cdot\,, x_j^{\otimes(p-1)}), \qquad Z_X(W) = \Pi_X A(W) = \nabla_{\mathrm{St}} H_{t+1}(X).$$

For every deterministic test matrix U of Frobenius norm one,

$$\langle U, Z_X(W)\rangle_F = \langle \Pi_X U, A(W)\rangle_F = -p\left\langle \widehat{W}, \sum_{j=1}^r \lambda_j\,\mathrm{Sym}\Big((\Pi_X U)_j \otimes x_j^{\otimes(p-1)}\Big)\right\rangle_F .$$

Here $W = \mathrm{Sym}(\widehat{W})$, and tensor symmetrization is the orthogonal projection onto the symmetric tensors. The deterministic coefficient tensor in the last line has Frobenius norm at most

$$p\sum_{j=1}^r \lambda_j \|(\Pi_X U)_j\|_2 \le p\Lambda_2 \|\Pi_X U\|_F \le L_A.$$

The coordinates of $\widehat{W}$ are conditionally independent and obey the stated scalar assumptions. Lemma B.2 therefore gives, uniformly in the realized predictable state,

$$\mathbb{E}\Big[e^{\theta\langle U, Z_X(W_{t+1})\rangle_F} \mid \mathcal{F}_t\Big] \le e^{2K^2L_A^2\theta^2}. \tag{116}$$

For $U = v_i e_j^\mathsf{T}$, its norm is one and $L_A = p\Lambda_2 \le p\lambda_+\sqrt{r} \le p\lambda_+(1+\sqrt{r})$. Thus $2K^2L_A^2 \le \alpha_1^2/2$, proving (110).

For the Gram bound, take a deterministic $1/2$-net $\mathcal{N}$ of the unit sphere in $\mathbb{R}^{N\times r}$, with $|\mathcal{N}| \le 5^{Nr}$. Such a net follows by the usual disjoint-ball volume argument applied to a maximal $1/2$-separated set. Every matrix Z satisfies $\|Z\|_F \le 2\max_{U\in\mathcal{N}} |\langle U, Z\rangle_F|$. The two-sided Chernoff bound from (116) is

$$\mathbb{P}\big(|\langle U, Z_X(W_{t+1})\rangle_F| > a \mid \mathcal{F}_t\big) \le 2\exp\left(-\frac{a^2}{8K^2L_A^2}\right).$$

Consequently,

$$\begin{aligned}\mathbb{P}\big(\|Z_X(W_{t+1})\|_F > 2T_G\sqrt{N} \mid \mathcal{F}_t\big) &\le 2\,5^{Nr}\exp\left(-\frac{T_G^2 N}{8K^2L_A^2}\right)\\ &= 2e^{-3N} \le e^{-2N} \qquad (N \ge 1).\end{aligned}$$

The deterministic population gradient satisfies

$$(\nabla\Phi(X))_j = -p\sqrt{N}\lambda_j \sum_{i=1}^r \lambda_i m_{ij}^{p-1} v_i.$$

By orthonormality of the v_i and $|m_{ij}| \le 1$,

$$\|\nabla_{\mathrm{St}}\Phi(X)\|_F^2 \le \|\nabla\Phi(X)\|_F^2 = p^2 N \sum_{i,j} \lambda_i^2\lambda_j^2 m_{ij}^{2p-2} \le C_\Phi^2 N.$$

Outside the preceding noise-gradient exceptional event, put $g = \nabla_{\mathrm{St}}\mathcal{L}(X; Y_{t+1})$. Then

$$\|g\|_F^2 \le 2C_\Phi^2 N + 2B_G N = \alpha_2 N, \qquad \|g^\mathsf{T} g\|_F \le \mathrm{tr}(g^\mathsf{T} g) = \|g\|_F^2.$$

For each time the conditional failure probability of the stated Gram bound is therefore at most e^{-2N}. Taking expectations and then a union bound over $s \le T$ proves (111) with $c_R = 2$. Independence of the events at different times is not used. On the Gram event itself, even without referring to the smaller event constructed above,

$$\|g\|_F^2 = \mathrm{tr}(g^\mathsf{T} g) \le \sqrt{r}\,\|g^\mathsf{T} g\|_F \le \sqrt{r}\,\alpha_2 N,$$

which is exactly the stated $C_G = r^{1/4}\sqrt{\alpha_2}$ bound.

Let $\eta = \delta_N/N$ and $\mathsf{G} = g^\mathsf{T} g$. Tangency gives $X^\mathsf{T} g + g^\mathsf{T} X = 0$, so the polar step is exactly

$$X_{t+1} = (X - \eta g)(I_r + \eta^2 \mathsf{G})^{-1/2}.$$

Define

$$D_{ij}(X) = -\langle v_i, (\nabla_{\mathrm{St}} \Phi(X))_j \rangle, \qquad \xi_{ij}(t+1) = -\eta \langle v_i, (Z_X(W_{t+1}))_j \rangle.$$

The remaining overlap increment is the (i,j) contraction of

$$(X - \eta g)\big((I_r + \eta^2 \mathsf{G})^{-1/2} - I_r\big).$$

Since G is positive semidefinite and $1 - (1+x)^{-1/2} \le x/2$ for $x \ge 0$,

$$\|(I_r + \eta^2 \mathsf{G})^{-1/2} - I_r\|_{\mathrm{op}} \le \frac{\eta^2}{2} \|\mathsf{G}\|_{\mathrm{op}},$$
$$\|X - \eta g\|_{\mathrm{op}}^2 = \|I_r + \eta^2 \mathsf{G}\|_{\mathrm{op}} = 1 + \eta^2 \|\mathsf{G}\|_{\mathrm{op}}.$$

On the Gram event $\|\mathsf{G}\|_{\mathrm{op}} \le \alpha_2 N$. Therefore every retraction remainder obeys

$$|r_{ij}(t+1)| \le \frac{\alpha_2 \delta_N^2}{2N} \sqrt{1 + \frac{\alpha_2 \delta_N^2}{N}},$$

which proves (112) and (114). Equation (110) with θ replaced by $-\eta\theta$ proves (113); its conditional mean is zero by the centering of the fresh raw tensor.

Finally, insert the displayed population gradient into Π_X. The resulting coordinate identity is exactly (100). Its direct monomial has modulus at most $p\lambda_+^2\sqrt{N}$. There are $2r^2$ projection monomials, each bounded by $(p/2)\lambda_+^2\sqrt{N}$. Hence

$$|D_{ij}(X)| \le p\lambda_+^2(1 + r^2)\sqrt{N} = C_D\sqrt{N},$$

proving (115).

Stopped noise under the model law. For completeness, if I_t is a predictable zero–one indicator, (113) makes the exponential process for $\sum_{u=1}^n I_{\sigma+u-1}\xi_{ij}(\sigma+u)$, compensated by $n\theta^2\alpha_1^2\delta_N^2/(2N^2)$, a conditional nonnegative supermartingale after a bounded stopping time σ. The maximal inequality and optimization at $\theta = aN^2/(H\alpha_1^2\delta_N^2)$ give the upper-tail probability $\exp(-a^2N^2/(2H\alpha_1^2\delta_N^2))$. Applying the same argument to the negative sum and adding proves (101). This argument is not conditioned on the Gram or any later success event. In particular, this is the stopped estimate (101). □

Role in the reconstruction. The conditional mgf and Gram bounds control the stopped microscopic comparison in Lemma B.11, the recovered-block stability in Lemma B.12, and the cleanup forcing in Lemma B.20. The exact drift and polar remainder identify the discrete overlap dynamics used throughout all three regimes.

Two deterministic summation lemmas

Lemma B.4. ***Abel bound for predictable contraction weights*** *Let $0 \le \beta_t \le 1$, $P_{s,n} := \prod_{u=s}^{n-1}(1-\beta_u)$, and $S_u := \sum_{s=1}^u \zeta_s$. Then*

$$\left|\sum_{s=1}^n P_{s,n}\zeta_s\right| \le 2 \max_{u \le n} |S_u|. \tag{117}$$

Proof. Since $P_{s,n}$ is nondecreasing in s, summation by parts gives

$$\sum_{s=1}^n P_{s,n}\zeta_s = P_{n,n}S_n - \sum_{s=1}^{n-1}(P_{s+1,n} - P_{s,n})S_s.$$

Here $P_{n,n} = 1$, every difference is nonnegative, and their sum is $1 - P_{1,n} \le 1$. Taking absolute values proves (117). □

Role in the reconstruction. This summation-by-parts estimate proves the scalar stable-recursion bound in Lemma B.5; its same cumulative-noise control is used for the surviving overlaps during Regime II, where the contraction weights depend on earlier updates.

Lemma B.5. ***Explicit stable-recursion estimate***

Suppose

$$y_{n+1} = (1 - \kappa_n h)y_n + f_n + \zeta_{n+1}, \qquad 0 < \kappa_* \le \kappa_n \le \kappa^*, \qquad h\kappa^* \le \tfrac{1}{2}.$$

Then, for every n,

$$|y_n| \le e^{-\kappa_* hn}|y_0| + \frac{\max_{s<n}|f_s|}{\kappa_* h} + 2\max_{u\le n}\left|\sum_{s=1}^{u}\zeta_s\right|. \tag{118}$$

Proof. Put $P_{s,n} := \prod_{u=s}^{n-1}(1-\kappa_u h)$, $P_{n,n} := 1$. Repeated substitution gives the identity

$$y_n = P_{0,n}y_0 + \sum_{s=0}^{n-1} P_{s+1,n}f_s + \sum_{s=1}^{n} P_{s,n}\zeta_s. \tag{119}$$

Because $0 \le 1-\kappa_u h \le 1-\kappa_* h$,

$$0 \le P_{0,n} \le (1-\kappa_* h)^n \le e^{-\kappa_* hn}, \tag{120}$$

$$\sum_{s=0}^{n-1} P_{s+1,n} \le \sum_{j=0}^{n-1}(1-\kappa_* h)^j \le \frac{1}{\kappa_* h}. \tag{121}$$

For fixed n, $P_{s,n}$ is nondecreasing in s, belongs to $[0,1]$, and $P_{n,n} = 1$. Lemma B.4 therefore gives

$$\left|\sum_{s=1}^{n} P_{s,n}\zeta_s\right| \le 2\max_{u\le n}\left|\sum_{s=1}^{u}\zeta_s\right|. \tag{122}$$

Taking absolute values in (119) and substituting (120)–(122) proves (118). □

Role in the reconstruction. This scalar estimate explains why a contracting overlap follows its deterministic decay up to a forcing floor. The cross coordinates actually used in Lemmas B.12 and B.20 require the coupled block estimate below; no later claim relies on this scalar lemma as a substitute for that estimate.

Exact diagonal linearization of the Stiefel drift Relabel rows and columns independently by the fixed greedy matching. Define

$$M_{ab} := m_{i_a^*, j_b^*}, \quad \mu_a := \lambda_{i_a^*}, \quad \nu_b := \lambda_{j_b^*}, \quad q_a := M_{aa}, \quad L_a := \mu_a\nu_a. \tag{123}$$

Thus the weight of the direct drift of M_{ab} is $\mu_a\nu_b$, not $\lambda_a\lambda_b$ using the labels before relabeling. Put $\overline{M} = \operatorname{diag}(q_1,\dots,q_r)$, $E = M - \overline{M}$, and $e = \|E\|_\infty$. Throughout this subsection, $p \ge 3$, $\lambda_- \le \mu_a, \nu_a \le \lambda_+$, and $|M_{ab}| \le 1$. Set

$$R_\lambda := \left(\frac{\lambda_+}{\lambda_-}\right)^{2/p}, \quad K_\lambda := 1 + R_\lambda, \quad \alpha_{ab} := \left(\frac{L_b}{L_a}\right)^{1/p},$$
$$\|z\|_{ab} := \frac{|z_1| + \alpha_{ab}|z_2|}{\min\{1, \alpha_{ab}\}}. \tag{124}$$

The division by $\min\{1,\alpha_{ab}\}$ is deliberate. It does not change any induced operator norm on a fixed block, and it gives

$$\max\{|z_1|,|z_2|\} \le \|z\|_{ab} \le K_\lambda \max\{|z_1|,|z_2|\}. \tag{125}$$

Indeed, the two coefficients in this norm are at least one, and their sum is $1 + \max\{\alpha_{ab}, \alpha_{ab}^{-1}\} \le K_\lambda$. We use this normalized norm everywhere below. The normalization prevents an unrecorded conversion factor when a block bound is turned into a bound on an individual overlap.

Lemma B.6. ***Complete two-coordinate block calculation*** *For $a < b$, let $Z_{ab} = (M_{ab}, M_{ba})^\mathsf{T}$, $A_a = L_a q_a^p$, and $B_b = L_b q_b^p$. At $\overline{M}$, the exact Jacobian of $-D/\sqrt{N}$ on this two-coordinate block is*

$$A_{ab} = p \begin{pmatrix} A_a/2 + B_b & L_b q_a q_b^{p-1}/2 \\ L_a q_b q_a^{p-1}/2 & A_a + B_b/2 \end{pmatrix}. \tag{126}$$

Its eigenvalues are $p(A_a + B_b)$ and $p(A_a + B_b)/2$. Define

$$c_p^{(1)} := \frac{1}{2} - \frac{p-1}{2p}(2p)^{-1/(p-1)}, \qquad c_p^{(2)} := 1 - \frac{1}{2p}\left(\frac{p-1}{p}\right)^{p-1}, \tag{127}$$
$$c_p := \min\{c_p^{(1)}, c_p^{(2)}\} > 0, \qquad L_{\mathrm{cross}} := 3pK_\lambda \lambda_+^2.$$

If $0 \le q_a, q_b \le 1$, then

$$\|A_{ab}\|_{ab\to ab} \le L_{\mathrm{cross}}, \qquad \|(I - hA_{ab})z\|_{ab} \le (1 - hpc_p A_a)\|z\|_{ab} \tag{128}$$

whenever $hL_{\mathrm{cross}} \le 1/2$. For general M as above,

$$\left\| \begin{pmatrix} D_{ab}(M) \\ D_{ba}(M) \end{pmatrix} + \sqrt{N} A_{ab} Z_{ab} \right\|_{ab} \le \tfrac{1}{2} K_\lambda C_{D,2} \sqrt{N} e^2, \tag{129}$$

$$|D_{aa}(M) - \sqrt{N} p L_a q_a^{p-1}(1 - q_a^2)| \le \tfrac{1}{2} C_{D,2} \sqrt{N} e^2, \tag{130}$$

where $C_{D,2} = p\lambda_+^2[(p-1)(p-2) + r^2 p(p+1)]$.

Proof. Exact differentiation and the factor p. With $\mathscr{G}(M) = \Lambda_\mu M^{\circ(p-1)} \Lambda_\nu$, the projection formula in (100) gives, exactly,

$$\frac{D_{ab}(M)}{\sqrt{N}} = p\mu_a \nu_b M_{ab}^{p-1} - \frac{p}{2} \sum_{c,d=1}^{r} \left(\mu_c \nu_b M_{ad} M_{cd} M_{cb}^{p-1} + \mu_c \nu_d M_{ad} M_{cb} M_{cd}^{p-1} \right). \tag{131}$$

Equivalently, $D(M)/\sqrt{N} = p\{\mathscr{G}(M) - M\,\mathrm{Sym}(M^\mathsf{T}\mathscr{G}(M))\}$. Thus both the direct and projected terms carry the same factor p; dropping it from the direct term would give an incorrect diagonal logistic drift. For the diagonal $\overline{M}$, differentiation in a matrix direction H, with $\dot{\mathscr{G}} = \mathscr{G}'(\overline{M})[H]$, gives

$$\frac{D'(\overline{M})[H]}{\sqrt{N}} = p\dot{\mathscr{G}} - pH\,\mathrm{Sym}(\overline{M}^\mathsf{T}\mathscr{G}(\overline{M})) - p\overline{M}\,\mathrm{Sym}(H^\mathsf{T}\mathscr{G}(\overline{M}) + \overline{M}^\mathsf{T}\dot{\mathscr{G}}).$$

When $p \ge 3$, $\dot{\mathscr{G}}_{ab} = 0$ for $a \ne b$. The entries of $-D'(\overline{M})/\sqrt{N}$ on $Z_{ab} = (M_{ab}, M_{ba})^\mathsf{T}$ are consequently the four entries in (126); the diagonal drift is $D_{aa}(\overline{M})/\sqrt{N} = pL_a q_a^{p-1}(1 - q_a^2)$. For completeness, setting $t = (L_b/L_a)^{1/p} q_b/q_a$ (and passing by continuity when $q_a = 0$), the two weighted-column contraction margins, in units of $pL_a q_a^p$, are

$$\tfrac{1}{2} + t^p - \tfrac{t}{2}, \qquad 1 + \tfrac{1}{2}(t^p - t^{p-1}).$$

For $A_a = L_a q_a^p > 0$ and $\alpha_{ab} = (L_b/L_a)^{1/p}$, these are respectively $(A_{ab,11} - \alpha_{ab} A_{ab,21})/(pA_a)$ and $(A_{ab,22} - \alpha_{ab}^{-1} A_{ab,12})/(pA_a)$. Their minima over $t \ge 0$ equal $c_p^{(1)}$ and $c_p^{(2)}$ in (127). The diagonal entries of $I - hA_{ab}$ remain nonnegative if $hL_{\mathrm{cross}} \le 1/2$, and the weighted ℓ^1 column sums give precisely (128). Entrywise bounds give $\|A_{ab}\|_{ab\to ab} \le 3pK_\lambda \lambda_+^2$.

For a general matrix M, write

$$M = \overline{M} + E, \qquad E := M - \overline{M},$$

where

$$\overline{M} := \mathrm{diag}(M_{11}, \ldots, M_{rr})$$

is the diagonal part of M. Consider the line segment

$$M(t) := \overline{M} + tE, \qquad t \in [0, 1].$$

Let

$$G(M) := \frac{D(M)}{\sqrt{N}}.$$

Applying Taylor's theorem with integral remainder along the path $t \mapsto M(t)$, we obtain, for every i, j,

$$G_{ij}(M) = G_{ij}(\overline{M}) + G'_{ij}(\overline{M})[E] + \int_0^1 (1-t)\, G''_{ij}(M(t))[E,E]\, dt. \tag{132}$$

Equivalently,

$$G_{ij}(M) - G_{ij}(\overline{M}) - G'_{ij}(\overline{M})[E] = \int_0^1 (1-t)\, G''_{ij}(M(t))[E,E]\, dt.$$

For $a \neq b$, we have

$$G_{ab}(\overline{M}) = G_{ba}(\overline{M}) = 0$$

and

$$\begin{pmatrix} G'_{ab}(\overline{M})[E] \\ G'_{ba}(\overline{M})[E] \end{pmatrix} = -A_{ab} Z_{ab}, \qquad Z_{ab} := \begin{pmatrix} E_{ab} \\ E_{ba} \end{pmatrix}.$$

Therefore,

$$\frac{1}{\sqrt{N}} \begin{pmatrix} D_{ab}(M) \\ D_{ba}(M) \end{pmatrix} + A_{ab} Z_{ab} = \int_0^1 (1-t) \begin{pmatrix} G''_{ab}(M(t))[E,E] \\ G''_{ba}(M(t))[E,E] \end{pmatrix} dt.$$

Let

$$e := \|E\|_\infty = \max_{u,v} |E_{uv}|.$$

From the form of $G(M)$,

$$\left|G''_{ij}(M(t))[E,E]\right| \le C_{D,2} e^2 \qquad \text{for all } t \in [0,1],$$

then

$$\left|G_{ij}(M) - G_{ij}(\overline{M}) - G'_{ij}(\overline{M})[E]\right| \le C_{D,2} e^2 \int_0^1 (1-t)\, dt = \frac{1}{2} C_{D,2} e^2.$$

Consequently,

$$\left\| \begin{pmatrix} D_{ab}(M) \\ D_{ba}(M) \end{pmatrix} + \sqrt{N}\, A_{ab} Z_{ab} \right\|_{ab} \le \frac{1}{2} K_\lambda C_{D,2} \sqrt{N}\, e^2.$$

Since $D(M) = \sqrt{N}\, G(M)$, with the same method, multiplying the preceding inequality by $\sqrt{N}$ yields

$$\left| D_{aa}(M) - \sqrt{N}\, p L_a q_a^{p-1} (1 - q_a^2) \right| \le \frac{1}{2} C_{D,2} \sqrt{N}\, e^2. \tag{133}$$

□

Role in the reconstruction. The weighted Jacobian gives contraction of selected row–column crosses and the inward logistic drift of a recovered leader. Both ingredients are used in the coupled estimate below, Lemma B.12, and the multi-stage cleanup in Lemma B.20.

Lemma B.7. ***Cross-factor census and coupled remainder*** *Fix $0 \le k \le r$, put $D_k := \{1, \dots, k\}$, and let $\mathcal{I}_k := \{(a,b) : 1 \le a < b \le r,\ a \le k\}$. Define*

$$Z_k^{\max}(M) := \max_{(a,b) \in \mathcal{I}_k} \|Z_{ab}\|_{ab}, \qquad \max \varnothing := 0. \tag{134}$$

For $(a,b) \in \mathcal{I}_k$, set

$$R_{ab}(M) := \begin{pmatrix} D_{ab}(M) \\ D_{ba}(M) \end{pmatrix} + \sqrt{N} A_{ab} Z_{ab}. \tag{135}$$

Then, with $C_D = p\lambda_+^2(1+r^2)$ and the common enlarged constant $C_{\mathrm{coup}} = K_\lambda^2 C_{D,2}$,

$$\max_{(a,b) \in \mathcal{I}_k} \|R_{ab}(M)\|_{ab} \le K_\lambda C_D \sqrt{N} e Z_k^{\max}(M) \le C_{\mathrm{coup}} \sqrt{N} e Z_k^{\max}(M). \tag{136}$$

In particular, the estimate applies to $k-1$ and k separately; the small size of the old block need not be replaced by the larger size of the new block.

Proof. If $\mathcal{I}_k$ is empty, both sides vanish. Otherwise, use(131), we have two components,

$$\mu_c \nu_b M_{ad} M_{cd} M_{cb}^{p-1}, \quad \mu_c \nu_d M_{ad} M_{cb} M_{cd}^{p-1}, \quad (c,d) \in [r]^2. \tag{137}$$

Fix $a < b$ with $a \le k$; no condition $b > k$ is imposed. For either monomial at output ab, exactly one of the following three cases holds:

$$d \neq a: \quad M_{ad} \text{ is a touching off-diagonal factor,} \tag{138}$$

$$d = a,\ c \neq a: \quad M_{ca} \text{ is a touching off-diagonal factor,} \tag{139}$$

$$d = a,\ c = a: \quad M_{ab} \text{ is a touching off-diagonal factor.} \tag{140}$$

Here *touching* means that at least one coordinate index belongs to D_k. Such a coordinate belongs to a pair in $\mathcal{I}_k$, whether its orientation is upper or lower triangular. And, monomial at output ba has the same situation.

The direct monomials must also be included. They are $\sqrt{N} p \mu_a \nu_b M_{ab}^{p-1}$ and $\sqrt{N} p \mu_b \nu_a M_{ba}^{p-1}$. They contain a touching factor and, because $p - 1 \ge 2$, at least one additional off-diagonal factor. At fixed indices a coordinate of M is either exactly its diagonal value q_i or exactly an entry of E. Thus a monomial has a well-defined off-diagonal degree; no binomial expansion is needed. Every surviving monomial has degree at least two, one touching factor, and one other off-diagonal factor. By (134), the touching factor is bounded by $Z_k^{\max}$ with coefficient one; the other factor is bounded by e; every remaining factor has magnitude at most one. This applies also when a coordinate appears with multiplicity: use two of its factor occurrences separately.

The total absolute coefficient of each component of $D/\sqrt{N}$ is bounded by

$$p\lambda_+^2 + 2r^2(p\lambda_+^2/2) = C_D.$$

Finally, via the same method as (132), $(p-1)(p-2) \ge 2$, $p(p+1) \ge 2$, and $K_\lambda \ge 1$ imply $C_{D,2} \ge 2C_D$ and $K_\lambda C_D \le K_\lambda^2 C_{D,2} = C_{\text{coup}}$. Now, we prove the common bound without suppressing any norm-conversion factor. □

Role in the reconstruction. The extra touching-cross factor lets the nonlinear remainder be absorbed by the linear block contraction. Lemma B.8 uses this bound for the old blocks, and Lemmas B.12 and B.20 use it to keep recovered rows and columns small.

Coupled matrix recursion, accumulated noise, and block resets The next lemma supplies the vector estimate that a scalar Abel inequality does not supply. For a fixed $\mathcal{I}_k$, stack the vectors Z_{ab} and use the product norm $\|z\|_{\mathcal{I}_k} := \max_{(a,b)\in\mathcal{I}_k} \|z_{ab}\|_{ab}$. The argument also works for any fixed finite collection of normed blocks.

Lemma B.8. ***Coupled contraction with cumulative noise*** *Let $\mathcal{I}$ be a fixed finite block collection, with the product norm just defined. Suppose for $0 \le n < H$ that*

$$z_{n+1} = (I - hA_n)z_n + hr_n + \xi_{n+1} + \pi_{n+1}, \tag{141}$$

where A_n is block diagonal and, pathwise,

$$\begin{gathered}\|I - hA_n\| \le 1 - 2\kappa h, \qquad \|A_n\| \le L_{\text{cross}}, \qquad \|r_n\| \le \varepsilon \|z_n\|, \\ 0 \le \varepsilon \le \kappa, \qquad 0 < 2\kappa h \le 1.\end{gathered} \tag{142}$$

Put $S_0 = 0$, $S_n = \sum_{s=1}^n \xi_s$, and assume

$$\max_{n \le H} \|S_n\| \le B, \qquad \max_{1 \le n \le H} \|\pi_n\| \le P. \tag{143}$$

Define the explicit error floor

$$F(\kappa, B, P) := \left(2 + \frac{L_{\text{cross}}}{\kappa}\right) B + \frac{P}{\kappa h}. \tag{144}$$

Then for every $0 \le n \le H$,

$$\|z_n\| \le (1 - \kappa h)^n \|z_0\| + F \le e^{-\kappa h n} \|z_0\| + F. \tag{145}$$

For $H \geq 1$, the accumulated squared size satisfies

$$h \sum_{n=0}^{H-1} \|z_n\|^2 \leq \frac{2h\|z_0\|^2}{1-(1-\kappa h)^2} + 2HhF^2 \leq \frac{2\|z_0\|^2}{\kappa} + 2HhF^2. \tag{146}$$

For the Stiefel specialization, let $U > 0$ be a deterministic localization envelope such that, throughout the time interval under consideration,

$$e(M_n) := \|M_n - \operatorname{diag} M_n\|_\infty \leq \frac{U}{\sqrt{N}}.$$

Assume additionally that $0 \leq q_b \leq 1$ for every b. Sufficient deterministic conditions are then

$$\begin{aligned} q_a \geq q_* > 0 \quad (a \leq k), \qquad e(M_n) \leq \frac{U}{\sqrt{N}}, \qquad \kappa := \frac{1}{2} p c_p \lambda_-^2 q_*^p, \\ N \geq \left(\frac{C_{\mathrm{coup}} U}{\kappa}\right)^2, \qquad hL_{\mathrm{cross}} \leq \frac{1}{2}. \end{aligned} \tag{147}$$

In the stage-k microscopic application below, we take

$$U = U_k^{\mathrm{mic}},$$

where U_k^{mic} is defined in (93).

Proof. All the following steps are deterministic on a path satisfying the hypotheses. The coefficients and the remainder may depend on the past; no independence between them and earlier noise is used. Define $y_n = z_n - S_n$. Since $S_{n+1} = S_n + \xi_{n+1}$, subtracting S_{n+1} from (141) gives the exact identity

$$y_{n+1} = (I - hA_n)y_n + hr_n - hA_nS_n + \pi_{n+1}. \tag{148}$$

Use the triangle inequality, $\|z_n\| \leq \|y_n\| + B$, and each of the hypotheses, in order:

$$\begin{aligned} \|y_{n+1}\| &\leq (1-2\kappa h)\|y_n\| + h\varepsilon\|z_n\| + hL_{\mathrm{cross}}B + P \\ &\leq (1-(2\kappa-\varepsilon)h)\|y_n\| + h(L_{\mathrm{cross}}+\varepsilon)B + P \\ &\leq (1-\kappa h)\|y_n\| + h(L_{\mathrm{cross}}+\varepsilon)B + P. \end{aligned} \tag{149}$$

Let $a = 1 - \kappa h \in [1/2, 1)$. Repeated substitution, starting from $y_0 = z_0$, yields

$$\begin{aligned} \|y_n\| &\leq a^n\|z_0\| + [h(L_{\mathrm{cross}}+\varepsilon)B + P]\sum_{j=0}^{n-1} a^j \\ &= a^n\|z_0\| + \left(\frac{L_{\mathrm{cross}}+\varepsilon}{\kappa}B + \frac{P}{\kappa h}\right)(1-a^n). \end{aligned} \tag{150}$$

Add $\|S_n\| \leq B$, bound $1 - a^n \leq 1$, and use $\varepsilon/\kappa \leq 1$. The result is (145). Squaring its first bound, using $(u+v)^2 \leq 2u^2 + 2v^2$, and summing gives

$$h\sum_{n=0}^{H-1}\|z_n\|^2 \leq 2h\|z_0\|^2 \sum_{n=0}^{H-1} a^{2n} + 2HhF^2.$$

Now $\sum a^{2n} \leq (1-a^2)^{-1}$, and $1 - a^2 = \kappa h(2-\kappa h) \geq \kappa h$, proving (146).

For the Stiefel application, Lemma B.6 gives block contraction at least $pc_p\lambda_-^2 q_*^p = 2\kappa$. The block diagonal operator has norm equal to the maximum of its block norms. Lemma B.7 gives $\|r_n\| \leq \varepsilon\|z_n\|$ with $\varepsilon = C_{\mathrm{coup}}U/\sqrt{N} \leq \kappa$. The bound on N in (147) is precisely the squared form of this last inequality. Each polar coordinate is at most b_N^{pol}, and (125) gives the stated P. This proves every required input. □

Role in the reconstruction. The common maximum bound handles coupled cross pairs while retaining cancellation of their accumulated noise. Lemma B.9 restarts it on time blocks; Lemma B.12 and Lemma B.20 then apply it on the full stage horizons.

The estimate is a coupled maximum estimate.

Lemma B.9. ***Explicit block resets and stopped sub-Gaussian probability***

Suppose the hypotheses of Lemma B.8 hold up to a common stopping time on a horizon H. Put

$$\ell := \left\lceil \frac{1}{\kappa h} \right\rceil, \qquad J := \lceil H/\ell \rceil, \qquad q := e^{-1}, \qquad C_{\mathrm{reset}} := 1 + \frac{1}{1-e^{-1}} < 3. \tag{151}$$

Assume that within every block of length at most ℓ, cumulative noise measured from that block's beginning is bounded by B, and each step's polar error is at most P. Let $F = F(\kappa, B, P)$. At every full block endpoint,

$$\|z_{j\ell}\| \le q^j \|z_0\| + \frac{1-q^j}{1-q} F, \tag{152}$$

and at all times before the common stopping time,

$$\|z_n\| \le \|z_0\| + C_{\mathrm{reset}} F. \tag{153}$$

Consequently a sufficient explicit tube budget is

$$\|z_0\| \le R_{\mathrm{in}}, \qquad F \le \frac{R_{\mathrm{out}} - R_{\mathrm{in}}}{C_{\mathrm{reset}}}, \qquad 0 < R_{\mathrm{in}} < R_{\mathrm{out}}. \tag{154}$$

A sufficient split of this budget, for any desired allowance $f_* > 0$, is

$$B \le \frac{f_*}{2(2 + L_{\mathrm{cross}}/\kappa)}, \qquad P \le \frac{\kappa h f_*}{2} \quad \Longrightarrow \quad F \le f_*. \tag{155}$$

For a fixed stage and a block collection with $n_{\mathrm{sc}} = 2|\mathcal{I}| \ge 1$ scalar coordinates, let σ be any finite stopping-time restart. Suppose each scalar noise increment, conditionally on its past, is centered sub-Gaussian with variance proxy $v_N^2 := \alpha_1^2 \delta_N^2 / N^2$. For predictable stopping indicators, the conditional probability that any scalar partial noise sum within any of the J blocks exceeds $b > 0$ in absolute value is at most

$$2 n_{\mathrm{sc}} J \exp\left(-\frac{b^2}{2\ell v_N^2} \right). \tag{156}$$

On its complement one may take $B = K_\lambda b$. For a desired failure probability $0 < \beta < 1$, an explicit choice is

$$b = v_N \sqrt{2\ell \log(2 n_{\mathrm{sc}} J/\beta)}, \qquad B = K_\lambda v_N \sqrt{2\ell \log(2 n_{\mathrm{sc}} J/\beta)}. \tag{157}$$

The additional conditional probability of a Gram failure in these H steps is at most $H e^{-c_R N}$. If $\mathcal{I}$ is empty, all cross-block statements are void and no cross-noise failure is charged.

Proof. **Deterministic resets.** For a block beginning at $j\ell$, apply Lemma B.8 after resetting its noise sum to zero. For $0 \le t \le \ell$, as long as the hypotheses hold, this gives

$$\|z_{j\ell+t}\| \le (1-\kappa h)^t \|z_{j\ell}\| + F.$$

At a full endpoint, $(1-\kappa h)^\ell \le e^{-\kappa h \ell} \le e^{-1} = q$. Writing $U_j = \|z_{j\ell}\|$, we therefore have $U_{j+1} \le q U_j + F$. Induction gives

$$U_j \le q^j U_0 + F \sum_{i=0}^{j-1} q^i = q^j U_0 + F(1-q^j)/(1-q),$$

which proves (152). At an intermediate time the bound is at most $U_j + F$, hence at most $U_0 + [1 + (1-q)^{-1}]F$, proving (153). The last, shorter block uses the same within-block estimate and needs no endpoint contraction. Substituting (154) proves the tube bound. The two inequalities in (155) bound the two summands of F separately by $f_*/2$, so their sum is at most f_*. Every coefficient in this allocation is explicit.

The stopped maximal inequality. Let σ be the restart time, and let $\tau \ge \sigma$ be the common stopping time, expressed in the global time index. We first fix one scalar coordinate and one time block. For some $0 \le j < J$, write

$$\rho := \sigma + j\ell, \qquad m := \min\{\ell, H - j\ell\}$$

for the beginning and length of this block. In the argument below, $\xi_{\rho+s}$ denotes the noise increment in the chosen scalar coordinate.

For $1 \le s \le m$, define

$$I_{s-1} := \mathbf{1}_{\{\rho+s-1<\tau\}}, \qquad \widetilde{\xi}_s := I_{s-1}\xi_{\rho+s},$$

and let

$$S_t := \sum_{s=1}^{t} \widetilde{\xi}_s, \qquad S_0 := 0, \qquad 0 \le t \le m.$$

Thus $I_{s-1} = 1$ whenever the process has not stopped before the update from time $\rho + s - 1$ to time $\rho + s$. Once the process has stopped, subsequent increments in S_t are set to zero.

We first explain why controlling S_t gives the noise bound needed in the deterministic argument. If $\rho + t \le \tau$, then, for every $1 \le s \le t$,

$$\rho + s - 1 < \tau.$$

Hence all the corresponding indicators equal one, and

$$S_t = \sum_{s=1}^{t} I_{s-1}\xi_{\rho+s} = \sum_{s=1}^{t} \xi_{\rho+s}.$$

Consequently,

$$\max_{0\le t\le m} |S_t| \le b$$

implies

$$\left|\sum_{s=1}^{t} \xi_{\rho+s}\right| \le b \qquad \text{whenever } 0 \le t \le m \text{ and } \rho + t \le \tau.$$

This includes the update that reaches the first exit time. No bound on the unstopped noise sum after τ is needed.

We now estimate the probability that S_t exceeds b. Introduce the local filtration

$$\mathcal{G}_t := \mathcal{F}_{\rho+t}, \qquad 0 \le t \le m.$$

Since τ is a stopping time, I_{s-1} is $\mathcal{G}_{s-1}$-measurable. On $\{I_{s-1} = 0\}$, the stopped increment is zero. On $\{I_{s-1} = 1\}$, the conditional sub-Gaussian hypothesis applies to the next noise increment. Therefore, for every $\theta \in \mathbb{R}$,

$$\mathbb{E}\Big[e^{\theta\widetilde{\xi}_s} \,\Big|\, \mathcal{G}_{s-1}\Big] \le e^{\theta^2 v_N^2/2}.$$

For a fixed $\theta > 0$, define

$$Y_t := \exp\left(\theta S_t - \frac{\theta^2 v_N^2 t}{2}\right).$$

Then $Y_0 = 1$, and

$$\begin{aligned}\mathbb{E}[Y_t \mid \mathcal{G}_{t-1}] &= Y_{t-1}\mathbb{E}\left[\exp\left(\theta\widetilde{\xi}_t - \frac{\theta^2 v_N^2}{2}\right) \,\middle|\, \mathcal{G}_{t-1}\right] \\ &\le Y_{t-1}.\end{aligned}$$

Thus $(Y_t)_{0\le t\le m}$ is a nonnegative supermartingale.

Let

$$\eta_b := \inf\{0 \le t \le m : S_t \ge b\}, \qquad \inf \varnothing := \infty,$$

and define the bounded stopping time

$$T := \eta_b \wedge m.$$

If a crossing occurs, then

$$T = \eta_b, \qquad S_T \ge b, \qquad T \le m \le \ell.$$

On this event,

$$Y_T \ge \exp\left(\theta b - \frac{\theta^2 v_N^2 \ell}{2}\right).$$

If no crossing occurs, Y_T is still nonnegative. Consequently, optional stopping gives

$$1 \geq \mathbb{E}[Y_T \mid \mathcal{G}_0] \geq \exp\left(\theta b - \frac{\theta^2 v_N^2 \ell}{2}\right) \mathbb{P}(\eta_b \leq m \mid \mathcal{G}_0).$$

Since

$$\{\eta_b \leq m\} = \left\{\max_{0\leq t\leq m} S_t \geq b\right\},$$

we obtain

$$\mathbb{P}\left(\max_{0\leq t\leq m} S_t \geq b \,\middle|\, \mathcal{F}_\rho\right) \leq \exp\left(-\theta b + \frac{\theta^2 v_N^2 \ell}{2}\right).$$

Choosing

$$\theta = \frac{b}{\ell v_N^2}$$

yields

$$\mathbb{P}\left(\max_{0\leq t\leq m} S_t \geq b \,\middle|\, \mathcal{F}_\rho\right) \leq \exp\left(-\frac{b^2}{2\ell v_N^2}\right).$$

The same argument applied to $-S_t$, followed by a union bound over the two signs, gives

$$\mathbb{P}\left(\max_{0\leq t\leq m} |S_t| \geq b \,\middle|\, \mathcal{F}_\rho\right) \leq 2\exp\left(-\frac{b^2}{2\ell v_N^2}\right).$$

We next combine these bounds across coordinates and blocks. Let $A_{j,c}$ be the event that the stopped partial noise sum for scalar coordinate c in block j reaches magnitude b. Because $\rho_j := \sigma + j\ell$ is a stopping time and $\mathcal{F}_\sigma \subseteq \mathcal{F}_{\rho_j}$, the tower property gives

$$\mathbb{P}(A_{j,c} \mid \mathcal{F}_\sigma) = \mathbb{E}\big[\mathbb{P}(A_{j,c} \mid \mathcal{F}_{\rho_j}) \,\big|\, \mathcal{F}_\sigma\big] \leq 2\exp\left(-\frac{b^2}{2\ell v_N^2}\right).$$

Taking a union bound over the n_{sc} scalar coordinates and the J time blocks proves

$$\mathbb{P}\left(\bigcup_{j=0}^{J-1}\bigcup_{c=1}^{n_{\mathrm{sc}}} A_{j,c} \,\middle|\, \mathcal{F}_\sigma\right) \leq 2n_{\mathrm{sc}} J \exp\left(-\frac{b^2}{2\ell v_N^2}\right),$$

which is (156). No independence between coordinates or between these events is required.

On the complementary event, every scalar partial noise sum has magnitude at most b. By (125), every two-coordinate block sum therefore has block norm at most $K_\lambda b$. Taking the maximum over the block collection shows that the deterministic cumulative-noise bound holds with

$$B = K_\lambda b.$$

In particular, the choice

$$b = v_N\sqrt{2\ell \log\left(\frac{2n_{\mathrm{sc}}J}{\beta}\right)}$$

makes the preceding upper bound equal to β. Thus the required noise bounds hold simultaneously with conditional probability at least $1-\beta$.

The Gram events are handled separately. The conditional one-step Gram failure bound, followed by the tower property and a union bound over the at most H relevant updates, gives an additional failure probability at most

$$He^{-c_R N}.$$

The indicator I_{s-1} may depend on information available before the current update, including earlier Gram outcomes, but it must not depend on the current fresh sample. In particular, we do not multiply the current noise increment by the indicator of the current Gram good event.

Finally, on the joint noise and Gram good event, the deterministic estimates apply through a first exit: the update that reaches the exit time uses hypotheses evaluated at the preceding state. If the first exit is a tube exit at local time

$$n_* := \tau - \sigma \leq H,$$

the deterministic bound therefore gives

$$\|z_{n_*}\| \le \|z_0\| + C_{\mathrm{reset}}F \le R_{\mathrm{in}} + C_{\mathrm{reset}}F.$$

With the strict budget

$$R_{\mathrm{in}} + C_{\mathrm{reset}}F < R_{\mathrm{out}},$$

this contradicts a tube exit satisfying $\|z_{n_*}\| \ge R_{\mathrm{out}}$. Hence the tube condition cannot be the first stopping condition to fail on this good event. □

Role in the reconstruction. The block reset allows contraction to persist for the polynomial horizons of Lemma B.12 and Lemma B.20. Its stopped maximal inequality also justifies the noise event used in Lemma B.11, including the update into a candidate exit.

Stopped microscopic comparison with a computable error budget

Lemma B.10. ***Discrete reciprocal-power comparison*** *Let $d \ge 1$, $a > 0$, and let $y_n > 0$ satisfy, up to a stopping time,*

$$y_{n+1} = y_n + a y_n^{d+1} + e_{n+1}, \qquad \max_{u \le n} \left| \sum_{s=1}^{u} e_s \right| \le b. \tag{158}$$

Assume $b \le c y_0/[32(d+1)]$ and $a y_n^d \le c/[64(d+1)]$ throughout the stopped interval, where $0 < c < 1/2$. Then

$$y_n \ge (1-c) y_0 [1 - da(1-c)^{d+1} y_0^d n]^{-1/d}, \tag{159}$$

$$y_n \le (1+c) y_0 [1 - da(1+c)^{d+1} y_0^d n]^{-1/d}, \tag{160}$$

as long as the displayed denominators being used are positive.

Proof. Let

$$E_n := \sum_{s=1}^{n} e_s, \qquad \widetilde{y}_n := y_n - E_n, \qquad \varepsilon := \frac{c}{32(d+1)}. \tag{161}$$

Since $\widetilde{y}_0 = y_0$ and $\widetilde{y}_{n+1} - \widetilde{y}_n = a y_n^{d+1} > 0$, induction gives $\widetilde{y}_n \ge y_0$. Hence

$$|E_n| \le b \le \varepsilon y_0 \le \varepsilon \widetilde{y}_n, \qquad (1-\varepsilon)\widetilde{y}_n \le y_n \le (1+\varepsilon)\widetilde{y}_n. \tag{162}$$

Put $a_- := a(1-\varepsilon)^{d+1}$ and $a_+ := a(1+\varepsilon)^{d+1}$. Then

$$a_- \widetilde{y}_n^{d+1} \le \widetilde{y}_{n+1} - \widetilde{y}_n \le a_+ \widetilde{y}_n^{d+1}. \tag{163}$$

Moreover,

$$a_+ \widetilde{y}_n^d \le a y_n^d \frac{(1+\varepsilon)^{d+1}}{(1-\varepsilon)^d} \le \frac{c}{64(d+1)} \exp\left((d+1)\varepsilon + \frac{d\varepsilon}{1-\varepsilon} \right) \le \frac{c}{32(d+1)}. \tag{164}$$

The last inequality follows from $\varepsilon \le 1/[64(d+1)]$, whence the exponent is at most $(2d+1)/[63(d+1)] < 1/16 < \log 2$. For every $x \ge 0$, convexity and Taylor's formula with integral remainder give

$$1 - dx \le (1+x)^{-d} \le 1 - dx + \frac{d(d+1)}{2} x^2. \tag{165}$$

Apply the left inequality to the upper increment in (163); this gives

$$\widetilde{y}_{n+1}^{-d} \ge \widetilde{y}_n^{-d} - d a_+. \tag{166}$$

Apply the right inequality to the lower increment and use (164); this gives

$$\widetilde{y}_{n+1}^{-d} \le \widetilde{y}_n^{-d} - d a_- \left(1 - \frac{d+1}{2} a_+ \widetilde{y}_n^d \right) \le \widetilde{y}_n^{-d} - d a_- (1 - c/64). \tag{167}$$

Now

$$a_+ \le a(1+c)^{d+1}, \tag{168}$$

$$a_-(1-c/64) \ge a(1-(d+1)\varepsilon - c/64) \ge a(1-c) \ge a(1-c)^{d+1}. \tag{169}$$

Summing (166)–(167), using $\widetilde{y}_0 = y_0$, inverting the positive denominators, and finally using $1-\varepsilon \geq 1-c$ and $1+\varepsilon \leq 1+c$ in (162) proves (159)–(160).

First-exit endpoint. Let $\tau \leq H$ be an integer first-exit time. Suppose the hypotheses on y_j and ay_j^d hold at the pre-update states $j < \tau$, and the cumulative error bound holds through τ. For $n \leq \tau$ set

$$E_n = \sum_{j<n} e_{j+1}, \qquad \widetilde{y}_n = y_0 + a\sum_{j<n} y_j^{d+1}, \qquad y_n = \widetilde{y}_n + E_n. \tag{170}$$

All summands in the middle expression have positive pre-update bases. Thus

$$\widetilde{y}_n \geq y_0, \qquad y_n \geq y_0 - b > 0, \qquad |E_n| \leq b \leq \frac{c}{32(d+1)}\widetilde{y}_n \quad (n \leq \tau). \tag{171}$$

In particular this proves positivity at a candidate lower-exit update, before invoking any negative power there. Summing (166)–(169) for $j = 0, \ldots, n-1$ uses only $j < \tau$ and gives

$$\begin{aligned} \widetilde{y}_n^{-d} &\geq y_0^{-d} - da(1+c)^{d+1}n, \\ \widetilde{y}_n^{-d} &\leq y_0^{-d} - da(1-c)^{d+1}n. \end{aligned} \tag{172}$$

The function $x \mapsto x^{-1/d}$ is decreasing on $(0,\infty)$. Invert the first inequality only when its right side is positive to obtain the upper envelope; invert the second only when its right side is positive to obtain the lower envelope. Each one-sided envelope needs its own denominator to be positive. □

Role in the reconstruction. The reciprocal-power envelopes turn microscopic overlap growth into explicit hitting-time clocks. Lemma B.11 applies them simultaneously to the active rectangle; Lemmas B.13, B.14, and B.19 use those clocks to select and grow each leader.

For a stage k, let $U_k^{\mathrm{mic}} > 0$ be the largest fixed coefficient in its microscopic bootstrap, so that every active overlap is at most $U_k^{\mathrm{mic}}/\sqrt{N}$. Define

$$A_k^{\mathrm{sel}} := 2 + \frac{4}{p(p-2)C_\delta(1-c_{\mathrm{ord}})^{p-1}\lambda_-^2\gamma_{2,k}^{p-2}}, \tag{173}$$

$$A_k^{\mathrm{mac}} := 2 + \frac{8(16/15)^{p-2}\gamma_{+,k}^{-(p-2)}}{C_\delta\kappa_{II,k}(1-2^{-(p-2)})}, \tag{174}$$

$$A_k^{\mathrm{mic}} := A_k^{\mathrm{sel}} + A_k^{\mathrm{mac}}, \tag{175}$$

$$Q_{k,N}^{\mathrm{mic}} := \left\lceil A_k^{\mathrm{mic}} d_{0,N}^{-1}(\log N)N^{p-2} \right\rceil, \tag{176}$$

$$a_{k,N}^{\mathrm{mic}} := \frac{\alpha_1\delta_N}{N}\sqrt{2Q_{k,N}^{\mathrm{mic}}\log(4r^2N^3)}, \tag{177}$$

$$b_{k,N}^{\mathrm{pol,mic}} := \frac{\alpha_2\delta_N^2}{2N^{3/2}}(\sqrt{r}U_k^{\mathrm{mic}} + C_G\delta_N), \tag{178}$$

$$\begin{aligned} b_{k,N}^{\mathrm{mic}} := {} & Q_{k,N}^{\mathrm{mic}} b_{k,N}^{\mathrm{pol,mic}} + a_{k,N}^{\mathrm{mic}} + C_{\mathrm{act}}(U_k^{\mathrm{mic}})^{p+1}\frac{\delta_N Q_{k,N}^{\mathrm{mic}}}{N^{(p+2)/2}} \\ & + \mathbf{1}_{\{k\geq 2\}} C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2\chi_N^2\frac{\delta_N Q_{k,N}^{\mathrm{mic}}}{\sqrt{N}}. \end{aligned} \tag{179}$$

The refined polar bound (178) follows because an active spike row has Euclidean norm at most $\sqrt{r}U_k^{\mathrm{mic}}/\sqrt{N}$, while the Euler gradient changes it by at most $C_G\delta_N/\sqrt{N}$. The old-block term is omitted at stage one. The monomial census is exact: an all-active correction contains two additional microscopic factors, whereas a term meeting a recovered row or column contains two old-cross

factors. Let N_k^{mic} be the least integer such that, for every $N \ge N_k^{\mathrm{mic}}$,

$$N \ge \max\{3, N_k^{\mathrm{race}}\}, \qquad d_{0,N} \le 1,$$
$$b_{k,N}^{\mathrm{mic}} \le \frac{c_{\mathrm{ord}}\gamma_{2,k}}{32(p-1)\sqrt{N}}, \qquad b_{k,N}^{\mathrm{mic}} \le \frac{\gamma_{-,k}}{16\sqrt{N}}, \qquad b_{k,N}^{\mathrm{mic}} + h_N C_D \le \varepsilon_{\mathrm{mac}},$$
$$h_N C_{D,1} (U_k^{\mathrm{mic}})^{p-2} N^{-(p-2)/2} \le \frac{c_{\mathrm{ord}}}{8},$$
$$h_N p \lambda_+^2 \left(\frac{U_k^{\mathrm{mic}}}{\sqrt{N}}\right)^{p-2} \le \frac{c_{\mathrm{ord}}}{64(p-1)},$$
$$\frac{C_{\mathrm{act}}(U_k^{\mathrm{mic}})^{p+1}}{p\lambda_-^2 \gamma_{2,k}^{p-1} N} + \mathbf{1}_{\{k\ge 2\}} \frac{C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2}{p\lambda_-^2 \gamma_{2,k}^{p-1} \log N} \le \frac{c_{\mathrm{ord}}}{8},$$
$$\frac{C_D \varepsilon_{\mathrm{mac}}^2}{p\lambda_-^2} + \mathbf{1}_{\{k\ge 2\}} \frac{C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2}{p\lambda_-^2 \gamma_{-,k}^{p-1} \log N} \le \frac{c_{\mathrm{ord}}}{16}.$$

The limits in (50) make this integer finite.

Lemma B.11. ***Stopped comparison, with no hidden constant*** *For $N \ge N_k^{\mathrm{mic}}$, at every capped restart satisfying the stage-k microscopic band, the envelopes* (96)–(97) *hold with $c = c_{\mathrm{ord}}$ up to the common bootstrap exit, except on an event of conditional probability at most $\eta_{k,N}^{\mathrm{cmp}} := Q_{k,N}^{\mathrm{mic}} e^{-c_R N} + N^{-3}$, as in* (181) *below. Consequently* (98) *holds at the same time.*

Proof. Fix a stage k, a capped stopping-time restart s, and $N \ge N_k^{\mathrm{mic}}$. For brevity, write

$$I := I_{k-1}, \qquad J := J_{k-1}, \qquad A := I \times J, \qquad Q := Q_{k,N}^{\mathrm{mic}}, \qquad U := U_k^{\mathrm{mic}}.$$

Let $\tau \le Q$ be the common microscopic stopping time, measured from s and adapted to the shifted filtration $(\mathcal{F}_{s+n})_{n\ge 0}$.

The initial microscopic band gives

$$m_f(X_s) \ge \frac{\gamma_{2,k}}{\sqrt{N}}, \qquad f \in A.$$

At every pre-update time $n < \tau$, the positive microscopic band gives

$$0 < m_f(X_{s+n}) \le \frac{U}{\sqrt{N}}, \qquad f \in A.$$

When $k \ge 2$, the common stop also includes the old-cross guard

$$|m_g(X_{s+n})| \le C_{\chi,k}^{\mathrm{in}} \chi_N$$

for every coordinate g in a previously recovered row or column, other than the recovered pairs themselves. These are pre-update assumptions; no such assumptions are imposed on the state at $n = \tau$.

1. Define the active and old projection corrections.

Fix $f = (i,j) \in A$. Define

$$Q_f^{\mathrm{act}}(X) := \frac{\sqrt{N}p}{2} \sum_{\substack{u\in I\\ v\in J}} \lambda_u m_{iv}(X) m_{uj}(X) m_{uv}(X) \times \left(\lambda_j m_{uj}(X)^{p-2} + \lambda_v m_{uv}(X)^{p-2}\right),$$
$$Q_f^{\mathrm{old}}(X) := \frac{\sqrt{N}p}{2} \sum_{\substack{u,v\in[r]\\ (u,v)\notin I\times J}} \lambda_u m_{iv}(X) m_{uj}(X) m_{uv}(X) \times \left(\lambda_j m_{uj}(X)^{p-2} + \lambda_v m_{uv}(X)^{p-2}\right).$$

The dependence on the fixed stage k is suppressed in this notation. At stage $k = 1$, the second sum is empty, so $Q_f^{\mathrm{old}} = 0$.

Thus (100) is exactly

$$D_f(X) = \sqrt{N} p \lambda_i \lambda_j m_f(X)^{p-1} - Q_f^{\mathrm{act}}(X) - Q_f^{\mathrm{old}}(X).$$

In particular, both corrections include the prefactor $\sqrt{N}p/2$.

2. Construct one event controlling all active noises.

Write

$$\xi_f(t) := \xi_{ij}(t), \qquad r_f(t) := r_{ij}(t).$$

Let $\mathcal{G}_{\mathrm{Gram}}$ be the event that the Gram bounds hold for all Q updates following s, and define

$$\mathcal{G}_{k,N}^{\mathrm{cmp}}(s,\tau) := \mathcal{G}_{\mathrm{Gram}} \cap \bigcap_{f \in A} \left\{ \max_{0 \le u \le Q} \left| \sum_{n=0}^{u-1} \mathbf{1}_{\{n<\tau\}} \xi_f(s+n+1) \right| \le a_{k,N}^{\mathrm{mic}} \right\}. \tag{180}$$

Since τ is a stopping time for $(\mathcal{F}_{s+n})_{n \ge 0}$, the indicator $\mathbf{1}_{\{n<\tau\}}$ is $\mathcal{F}_{s+n}$-measurable. It is therefore predictable for the fresh increment $\xi_f(s+n+1)$.

Apply (101) with

$$H = Q, \qquad a = a_{k,N}^{\mathrm{mic}} = \frac{\alpha_1 \delta_N}{N} \sqrt{2Q \log(4r^2 N^3)}.$$

For each $f \in A$, the resulting failure probability is at most

$$2 \exp\left(-\frac{(a_{k,N}^{\mathrm{mic}})^2 N^2}{2\alpha_1^2 \delta_N^2 Q} \right) = 2 \exp\bigl(-\log(4r^2 N^3) \bigr).$$

There are at most r^2 active coordinates. Combining their bounds with the conditional Gram bound gives

$$\begin{aligned} \mathbb{P}\left((\mathcal{G}_{k,N}^{\mathrm{cmp}}(s,\tau))^c \,\middle|\, \mathcal{F}_s \right) &\le Q e^{-c_R N} + 2r^2 \exp\bigl(-\log(4r^2 N^3) \bigr) \\ &= Q e^{-c_R N} + \frac{1}{2N^3} \\ &\le Q e^{-c_R N} + N^{-3} = \eta_{k,N}^{\mathrm{cmp}}. \end{aligned} \tag{181}$$

The Gram event is handled separately; its current-step indicator is not inserted into the noise increments.

For the rest of the proof, work on $\mathcal{G}_{k,N}^{\mathrm{cmp}}(s,\tau)$. For every $u \le \tau$, all indicators in the corresponding partial sum equal one. Hence

$$\sum_{n=0}^{u-1} \mathbf{1}_{\{n<\tau\}} \xi_f(s+n+1) = \sum_{n=0}^{u-1} \xi_f(s+n+1),$$

and consequently

$$\max_{0 \le u \le \tau} \left| \sum_{n=0}^{u-1} \xi_f(s+n+1) \right| \le a_{k,N}^{\mathrm{mic}}.$$

This includes $u = \tau$, because the increment entering the exit state is multiplied by $\mathbf{1}_{\{\tau-1<\tau\}} = 1$.

3. Rewrite the exact recursion in the form of Lemma B.10.

For the fixed coordinate $f = (i,j) \in A$, set

$$y_n := m_f(X_{s+n}), \qquad 0 \le n \le \tau.$$

Substituting the drift decomposition from Step 1 into (99), and using $h_N = \delta_N / \sqrt{N}$, gives, for $n < \tau$,

$$\begin{aligned} y_{n+1} = y_n &+ h_N p \lambda_i \lambda_j y_n^{p-1} \\ &- \frac{\delta_N}{N} \bigl(Q_f^{\mathrm{act}}(X_{s+n}) + Q_f^{\mathrm{old}}(X_{s+n}) \bigr) + \xi_f(s+n+1) + r_f(s+n+1). \end{aligned}$$

Define

$$a := h_N p \lambda_i \lambda_j, \qquad d := p-2, \qquad b := b^{\mathrm{mic}}_{k,N}, \qquad c := c_{\mathrm{ord}},$$
$$e_{n+1} := -\frac{\delta_N}{N}\left(Q_f^{\mathrm{act}}(X_{s+n}) + Q_f^{\mathrm{old}}(X_{s+n})\right) + \xi_f(s+n+1) + r_f(s+n+1). \tag{182}$$

Since $d+1 = p-1$, the recursion becomes exactly

$$y_{n+1} = y_n + a y_n^{d+1} + e_{n+1}, \qquad n < \tau.$$

This is the recursion in Lemma B.10. We next verify its cumulative-error and small-step assumptions.

4. Bound each contribution to the cumulative error.

First consider Q_f^{act}. Expanding the parentheses in its definition gives the two monomials

$$\lambda_u \lambda_j m_{iv} m_{uj}^{p-1} m_{uv}, \qquad \lambda_u \lambda_v m_{iv} m_{uj} m_{uv}^{p-1}.$$

When $u \in I$ and $v \in J$, every overlap in these monomials is active. Each monomial contains $p+1$ overlap factors, counted with multiplicity. Thus, at $n < \tau$,

$$\begin{aligned} |Q_f^{\mathrm{act}}(X_{s+n})| &\le \frac{\sqrt{N}p}{2}\, 2r^2 \lambda_+^2 \left(\frac{U}{\sqrt{N}}\right)^{p+1} \\ &= C_{\mathrm{act}} U^{p+1} N^{-p/2}. \end{aligned}$$

It follows that, for every $u \le \tau$,

$$\frac{\delta_N}{N}\sum_{n=0}^{u-1} |Q_f^{\mathrm{act}}(X_{s+n})| \le C_{\mathrm{act}} U^{p+1} \frac{\delta_N Q}{N^{(p+2)/2}}.$$

Next consider Q_f^{old}. The complement of $I \times J$ consists of three disjoint cases. In each case, both monomials contain the following two old-cross factors:

Summation indices	Two old-cross factors
$u \notin I,\ v \in J$	$m_{uj},\ m_{uv}$
$u \in I,\ v \notin J$	$m_{iv},\ m_{uv}$
$u \notin I,\ v \notin J$	$m_{iv},\ m_{uj}.$

Each of these two factors is bounded in absolute value by $C^{\mathrm{in}}_{\chi,k}\chi_N$. Every remaining overlap factor has absolute value at most one, by (107). Therefore

$$|Q_f^{\mathrm{old}}(X_{s+n})| \le C_{\mathrm{old}} (C^{\mathrm{in}}_{\chi,k})^2 \sqrt{N}\, \chi_N^2.$$

Since this correction vanishes when $k = 1$, we obtain

$$\frac{\delta_N}{N}\sum_{n=0}^{u-1} |Q_f^{\mathrm{old}}(X_{s+n})| \le \mathbf{1}_{\{k\ge 2\}} C_{\mathrm{old}} (C^{\mathrm{in}}_{\chi,k})^2 \chi_N^2 \frac{\delta_N Q}{\sqrt{N}}.$$

We also verify the refined polar estimate used here. For an active row $i \in I$, entries in active columns satisfy $|m_{iv}| \le U/\sqrt{N}$. Entries in recovered columns are old crosses. Since $p \ge 3$ and $N \ge 3$,

$$\chi_N = \frac{N^{-(p-1)/4}}{\sqrt{\log N}} \le \frac{1}{\sqrt{N}}.$$

Moreover, the definitions (92)–(93) give $C^{\mathrm{in}}_{\chi,k} \le U$. Thus entries in recovered columns also satisfy

$$|m_{iv}| \le C^{\mathrm{in}}_{\chi,k}\chi_N \le \frac{U}{\sqrt{N}}.$$

Consequently, the entire active spike row obeys

$$\|(m_{iv}(X_{s+n}))_{v=1}^{r}\|_2 \le \frac{\sqrt{r}U}{\sqrt{N}}.$$

On the Gram event, the Euler step changes this row by at most $C_G\delta_N/\sqrt{N}$, and the polar correction matrix differs from the identity in operator norm by at most $\alpha_2\delta_N^2/(2N)$. Hence

$$\begin{aligned}|r_f(s+n+1)| &\le \frac{\alpha_2\delta_N^2}{2N}\left(\frac{\sqrt{r}U}{\sqrt{N}}+\frac{C_G\delta_N}{\sqrt{N}}\right)\\ &= b_{k,N}^{\mathrm{pol,mic}},\end{aligned}$$

which is (178). Therefore

$$\sum_{n=0}^{u-1}|r_f(s+n+1)| \le Qb_{k,N}^{\mathrm{pol,mic}}.$$

Finally, Step 2 gives

$$\left|\sum_{n=0}^{u-1}\xi_f(s+n+1)\right| \le a_{k,N}^{\mathrm{mic}}, \qquad u\le\tau.$$

Combining these four bounds with the definition of e_{n+1} yields

$$\begin{aligned}\left|\sum_{n=0}^{u-1}e_{n+1}\right| &\le \frac{\delta_N}{N}\sum_{n=0}^{u-1}|Q_f^{\mathrm{act}}(X_{s+n})|\\ &\quad+\frac{\delta_N}{N}\sum_{n=0}^{u-1}|Q_f^{\mathrm{old}}(X_{s+n})|\\ &\quad+\left|\sum_{n=0}^{u-1}\xi_f(s+n+1)\right|+\sum_{n=0}^{u-1}|r_f(s+n+1)|\\ &\le C_{\mathrm{act}}U^{p+1}\frac{\delta_N Q}{N^{(p+2)/2}}\\ &\quad+\mathbf{1}_{\{k\ge2\}}C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2\chi_N^2\frac{\delta_N Q}{\sqrt{N}}\\ &\quad+a_{k,N}^{\mathrm{mic}}+Qb_{k,N}^{\mathrm{pol,mic}}\\ &= b_{k,N}^{\mathrm{mic}}.\end{aligned}$$

The last equality is precisely (179). In particular,

$$\max_{0\le u\le\tau}\left|\sum_{n=0}^{u-1}e_{n+1}\right| \le b.$$

Only the noise contribution is estimated using cancellation in its partial sums. The other three contributions are bounded by sums of absolute values. No centeredness assumption is needed for e_{n+1}.

5. Verify the two smallness conditions in Lemma B.10, including the exit endpoint.

Because $N\ge N_k^{\mathrm{mic}}$, the defining inequalities for N_k^{mic} give

$$\begin{aligned}b = b_{k,N}^{\mathrm{mic}} &\le \frac{c_{\mathrm{ord}}\gamma_{2,k}}{32(p-1)\sqrt{N}} \le \frac{cy_0}{32(d+1)},\\ ay_n^d &= h_N p\lambda_i\lambda_j y_n^{p-2}\\ &\le h_N p\lambda_+^2\left(\frac{U}{\sqrt{N}}\right)^{p-2} \le \frac{c_{\mathrm{ord}}}{64(p-1)} = \frac{c}{64(d+1)}, \qquad n<\tau.\end{aligned} \tag{183}$$

Here we used $y_0\ge\gamma_{2,k}/\sqrt{N}$ and $d+1=p-1$. Also, $a>0$, $d\ge1$, and $0<c<1/2$.

To justify application through the update entering τ, put

$$E_u := \sum_{n=0}^{u-1}e_{n+1}, \qquad \widetilde{y}_u := y_0 + a\sum_{n=0}^{u-1}y_n^{d+1}, \qquad 0\le u\le\tau.$$

Summing the exact recursion gives

$$y_u = \widetilde{y}_u + E_u.$$

Every summand defining $\widetilde{y}_u$ uses a positive pre-update state. Hence

$$\widetilde{y}_u \geq y_0, \qquad y_u \geq y_0 - b > 0, \qquad 0 \leq u \leq \tau.$$

Thus positivity also holds at the exit endpoint. Together with the cumulative-error estimate through τ and the small-step estimate for $n < \tau$, this verifies all hypotheses of Lemma B.10 with its endpoint convention.

6. Substitute the parameters into the comparison envelopes.

Lemma B.10 now gives

$$y_n \geq (1-c)y_0 \left[1 - da(1-c)^{d+1} y_0^d n\right]^{-1/d},$$
$$y_n \leq (1+c)y_0 \left[1 - da(1+c)^{d+1} y_0^d n\right]^{-1/d}.$$

Each inequality holds for $0 \leq n \leq \tau$ whenever its own displayed denominator is positive.

The substitutions into these denominators are

$$d + 1 = p - 1, \qquad y_0 = m_f(X_s), \qquad G_f(X_s) = \lambda_i \lambda_j y_0^d,$$

and therefore

$$da\, y_0^d = dh_N p \lambda_i \lambda_j y_0^d = pdh_N G_f(X_s).$$

Consequently,

$$m_f(X_{s+n}) \geq (1-c)m_f(X_s) \left[1 - pdh_N(1-c)^{p-1} G_f(X_s) n\right]^{-1/d},$$
$$m_f(X_{s+n}) \leq (1+c)m_f(X_s) \left[1 - pdh_N(1+c)^{p-1} G_f(X_s) n\right]^{-1/d}.$$

These are exactly (96)–(97), with $c = c_{\mathrm{ord}}$.

The event constructed in Step 2 controls every $f \in A$ simultaneously, so the comparison holds simultaneously for all active coordinates. Furthermore, since $c_{\mathrm{ord}} \leq \bar{c} < 1/2$, the same smallness estimates also verify the hypotheses of Lemma B.10 with $c = \bar{c}$. This uses the same event and requires no additional failure probability.

7. Derive the ratio estimate.

Take $c = c_{\mathrm{ord}}$ and let $e, f \in A$ satisfy

$$G_e(X_s) > B(c) G_f(X_s), \qquad B(c) = \left(\frac{1+c}{1-c}\right)^{p-1}.$$

Define the two denominators needed for this comparison:

$$D_e^-(n) := 1 - pdh_N(1-c)^{p-1} G_e(X_s) n,$$
$$D_f^+(n) := 1 - pdh_N(1+c)^{p-1} G_f(X_s) n.$$

The initial gap implies

$$(1-c)^{p-1} G_e(X_s) > (1+c)^{p-1} G_f(X_s),$$

and hence

$$D_f^+(n) \geq D_e^-(n).$$

In particular, if $D_e^-(n) > 0$, then the denominator $D_f^+(n)$ is also positive.

Apply the lower envelope to m_e and the upper envelope to m_f. Taking their ratio to the power d and including the weights in the definition of G gives

$$\frac{G_e(X_{s+n})}{G_f(X_{s+n})} \geq \left(\frac{1-c}{1+c}\right)^d \frac{G_e(X_s)}{G_f(X_s)} \frac{D_f^+(n)}{D_e^-(n)}$$
$$\geq A(c)^d \frac{G_e(X_s)}{G_f(X_s)},$$

where $A(c) = (1-c)/(1+c)$. This proves (98) on its stated denominator domain.

All conclusions hold through the common exit, on an event of conditional probability at least $1 - \eta_{k,N}^{\mathrm{cmp}}$. No conclusion beyond τ is asserted. □

Role in the reconstruction. This is the stage-uniform microscopic race input for Lemma B.13 and the later-stage selection in Lemma B.19. The ratio estimate preserves each future greedy leader until it reaches the next macroscopic regime.

Recovered-block stability with separate maxima and time-block restarts Only pairs touching an already recovered position are included here. Thus the maximum is $W_{k-1} = Z_{k-1}^{\max}$, not $Z_k^{\max}$. All block norms are the normalized norms of Lemma B.6; in particular each scalar cross is bounded by W_{k-1}.

$$b_* := 1 - \varepsilon_{\mathrm{rec}}, \qquad c_* := 1 - \varepsilon_{\mathrm{rec}}/2,$$

$$\kappa_{\mathrm{old}} := \tfrac{1}{2} p c_p \lambda_-^2 b_*^p, \tag{184}$$

$$\ell_{\mathrm{old}} := \tfrac{1}{2} p \lambda_-^2 \min\{b_*^{p-1}(1-b_*^2), c_*^{p-1}(1-c_*^2)\}, \tag{185}$$

$$D_{\mathrm{old}} := 2 + \frac{3pK_\lambda \lambda_+^2}{\kappa_{\mathrm{old}}}, \tag{186}$$

$$L_{k,N}^{\mathrm{old}} := \lceil (\kappa_{\mathrm{old}} h_N)^{-1} \rceil, \qquad B_{k,N}^{\mathrm{old}} := \lceil Q_{k,N}^{\mathrm{mic}} / L_{k,N}^{\mathrm{old}} \rceil, \tag{187}$$

$$b_{k,N}^{\mathrm{old}} := \frac{\chi_N}{192 K_\lambda D_{\mathrm{old}}}, \tag{188}$$

$$\varepsilon_{k,N}^{\mathrm{coup,old}} := \frac{C_{\mathrm{coup}} U_k^{\mathrm{mic}}}{\sqrt{N}}, \tag{189}$$

$$F_{k,N}^{\mathrm{old}} := D_{\mathrm{old}} K_\lambda b_{k,N}^{\mathrm{old}} + \frac{K_\lambda b_N^{\mathrm{pol}}}{\kappa_{\mathrm{old}} h_N},$$

$$\boldsymbol{b}_{k,N}^{\mathrm{old}} := 3F_{k,N}^{\mathrm{old}} = \frac{\chi_N}{64} + \frac{3K_\lambda b_N^{\mathrm{pol}}}{\kappa_{\mathrm{old}} h_N}. \tag{190}$$

The factor 3 is a specified numerical reserve: $1 + (1 - e^{-1})^{-1} < 3$. The boundary drift is minimized over the entire strip $[b_*, c_*]$, not evaluated only at its left endpoint. Let N_k^{old} be the least integer from which all the following inequalities hold for every N:

$$\begin{gathered} N \ge 3, \qquad d_{0,N} \le 1, \\ K_\lambda C_{\chi,k-1} \chi_N + \boldsymbol{b}_{k,N}^{\mathrm{old}} < C_{\chi,k}^{\mathrm{in}} \chi_N, \\ \varepsilon_{k,N}^{\mathrm{coup,old}} \le \kappa_{\mathrm{old}}, \qquad 3pK_\lambda \lambda_+^2 h_N \le \tfrac{1}{2}, \\ \frac{C_{D,2}(U_k^{\mathrm{mic}})^2}{2N} \le \ell_{\mathrm{old}}/8, \qquad b_N^{\mathrm{pol}} \le h_N \ell_{\mathrm{old}}/8, \\ h_N(C_D + \ell_{\mathrm{old}}/8) + b_N^{\mathrm{pol}} \le \varepsilon_{\mathrm{rec}}/2. \end{gathered} \tag{191}$$

The cross-radius reserve B_χ defined earlier includes the factor 3 in the new polar floor, so the first condition has fixed strict slack. All other left sides tend to zero. These conditions therefore define a finite threshold for every specified admissible sequence $d_{0,N}$. Define

$$\begin{aligned} \eta_{k,N}^{\mathrm{old}} := Q_{k,N}^{\mathrm{mic}} e^{-c_R N} + 4r^2 B_{k,N}^{\mathrm{old}} \exp\left[-\frac{(b_{k,N}^{\mathrm{old}})^2 N^2}{2\alpha_1^2 \delta_N^2 L_{k,N}^{\mathrm{old}}} \right] \\ + 2(k-1) Q_{k,N}^{\mathrm{mic}} \exp\left[-\frac{\ell_{\mathrm{old}}^2 N}{128 \alpha_1^2} \right]. \end{aligned} \tag{192}$$

Lemma B.12. ***Long-horizon recovered-block stability*** *Let $k \ge 2$, $N \ge N_k^{\mathrm{old}}$, and let s be a capped stopping time with $X_s \in E_{k-1}^\sharp$. Stop at the active-geometry exit, which includes failure of $0 \le q_b \le 1$ or $\|M - \operatorname{diag} M\|_\infty \le U_k^{\mathrm{mic}}/\sqrt{N}$. Up to this exit and throughout $0 \le n \le Q_{k,N}^{\mathrm{mic}}$, with conditional failure probability at most $\eta_{k,N}^{\mathrm{old}}$,*

$$m_{e_a}(X_{s+n}) \ge 1 - \varepsilon_{\mathrm{rec}} \qquad (a < k), \tag{193}$$

$$\max_{f \in \mathcal{Z}_{k-1}} |m_f(X_{s+n})| \le C_{\chi,k}^{\mathrm{in}} \chi_N. \tag{194}$$

Here $\mathcal{Z}_{k-1}$ denotes off-diagonal coordinates touching one of the first $k-1$ recovered positions, as defined explicitly in (305)*. For an active $f \in A_{k-1}$, the old contribution satisfies* (196) *below throughout this interval. The all-active and relative estimates require, in addition, the fully microscopic band $|m_{ij}| \le U_k^{\mathrm{mic}}/\sqrt{N}$ for every $(i,j) \in A_{k-1}$:*

$$|Q_f^{\mathrm{act}}| \le C_{\mathrm{act}}(U_k^{\mathrm{mic}})^{p+1} N^{-p/2}, \tag{195}$$

$$|Q_f^{\mathrm{old}}| \le C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2 \sqrt{N}\chi_N^2, \tag{196}$$

$$\frac{|Q_f^{\mathrm{act}}| + |Q_f^{\mathrm{old}}|}{\sqrt{N} p\lambda_-^2 (\gamma_{2,k}/\sqrt{N})^{p-1}} \le \frac{C_{\mathrm{act}}(U_k^{\mathrm{mic}})^{p+1}}{p\lambda_-^2 \gamma_{2,k}^{p-1} N} + \frac{C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2}{p\lambda_-^2 \gamma_{2,k}^{p-1} \log N}, \tag{197}$$

$$a_N^{\mathrm{step}} \sum_{n < Q_{k,N}^{\mathrm{mic}}} |Q_f^{\mathrm{old}}| \le \frac{C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2 C_\delta (A_k^{\mathrm{mic}} + 1)}{\sqrt{N}\log N}. \tag{198}$$

The sum is stopped at the indicated geometry exit. The last display uses $N \ge e$ and $d_{0,N} \le 1$.

Proof. The common stop. Intersect the active-geometry stop with the first old-diagonal or old-cross exit. Every estimate below is proved before that common stop and includes the single update into a proposed exit. Before it, each block $a < b$, $a < k$, has $q_a \ge b_*$. Lemma B.6 supplies linear contraction $1 - 2\kappa_{\mathrm{old}} h_N$. Lemma B.7 supplies the simultaneous remainder bound $\|r\| \le \varepsilon_{k,N}^{\mathrm{coup,old}} W_{k-1}$. Notice the index $k-1$ in both statements.

Local noise events and their probabilities. Divide the deterministic horizon into intervals starting at $mL_{k,N}^{\mathrm{old}}$. On each interval stop the coordinate noise with the common predictable indicator. For every scalar coordinate and every interval, require

$$\max_{0 \le u \le L_{k,N}^{\mathrm{old}}} \left| \sum_{v=1}^{u} \xi_f(s + mL_{k,N}^{\mathrm{old}} + v) \mathbf{1}_{\{mL_{k,N}^{\mathrm{old}} + v - 1 < \tau\}} \right| \le b_{k,N}^{\mathrm{old}}. \tag{199}$$

The conditional maximal inequality gives failure at most $2\exp[-(b_{k,N}^{\mathrm{old}})^2 N^2/(2\alpha_1^2 \delta_N^2 L_{k,N}^{\mathrm{old}})]$. There are at most $r^2 B_{k,N}^{\mathrm{old}}$ tests. The second term of (192) is a conservative union bound. Because a normalized block norm is at most K_λ times its largest component, the cumulative vector noise is at most $K_\lambda b_{k,N}^{\mathrm{old}}$. The vector polar forcing is at most $K_\lambda b_N^{\mathrm{pol}}$ per step.

Matrix recursion and time-block iteration. Apply Lemma B.8 to the maximum over all old pairs. On one interval, at offset u,

$$W_{k-1}(mL + u) \le e^{-\kappa_{\mathrm{old}} h_N u} W_{k-1}(mL) + F_{k,N}^{\mathrm{old}}.$$

For a full interval $\kappa_{\mathrm{old}} h_N L \ge 1$. Induction on m gives

$$W_{k-1}(mL) \le e^{-\kappa_{\mathrm{old}} h_N mL} W_{k-1}(0) + F_{k,N}^{\mathrm{old}} \sum_{j=0}^{m-1} e^{-j}. \tag{200}$$

At an interior point add the local floor once. Since $1 + \sum_{j \ge 0} e^{-j} < 3$, this proves

$$W_{k-1}(n) \le e^{-\kappa_{\mathrm{old}} h_N n} W_{k-1}(0) + \boldsymbol{b}_{k,N}^{\mathrm{old}}. \tag{201}$$

This argument never replaces coupled vector weights by scalar Abel weights. At entry $W_{k-1}(0) \le K_\lambda C_{\chi,k-1}\chi_N$; hence (191) prevents the old-cross exit with strict slack.

An entire inward strip. The derivative of $g(q) = q^{p-1}(1-q^2)$ on $(0,1)$ is $q^{p-2}[(p-1) - (p+1)q^2]$. It has one zero, a maximum. Therefore the minimum on $[b_*, c_*]$ is attained at an endpoint, and the normalized logistic drift is at least $2\ell_{\mathrm{old}}$ throughout the strip. The diagonal remainder is at most $\ell_{\mathrm{old}}/8$. Require at each step, for each old diagonal, $|\xi_{aa}| \le h_N \ell_{\mathrm{old}}/8$. The sub-Gaussian tail and a union bound give precisely the last term of (192). Inside the strip the full increment is at least

$$h_N(2\ell_{\mathrm{old}} - \ell_{\mathrm{old}}/8 - \ell_{\mathrm{old}}/8) - b_N^{\mathrm{pol}} > 0.$$

From above c_* the maximum downward displacement is at most $h_N(C_D + \ell_{\mathrm{old}}/8) + b_N^{\mathrm{pol}} \le c_* - b_*$. Both cases exclude a crossing below b_*, including from an entry point exactly at b_*. The Gram union bound is the first failure term. Thus neither old exit occurs before the active-geometry exit.

The exact active/old census. Relabel the target as $i, j \ge k$. In either projected monomial, if $u, v \ge k$, all its $p+1$ factors are active. Under the additional microscopic band its absolute value is at most $(\sqrt{N}p\lambda_+^2/2)(U_k^{\mathrm{mic}}/\sqrt{N})^{p+1}$. Summing the two branches over at most r^2 pairs proves (195). If $u < k, v \ge k$, the factors M_{uj} and M_{uv} are old crosses, with positive exponents in both branches. If $v < k, u \ge k$, the factors M_{iv} and M_{uv} are old crosses. If $u, v < k$, the factors M_{iv} and M_{uj} are old crosses. These three disjoint cases exhaust the complement of $u, v \ge k$. All remaining factors have modulus at most one. Summing proves (196) with the stated conservative C_{old}. Division by the displayed positive denominator gives (197). Finally

$$\delta_N Q_{k,N}^{\mathrm{mic}} \le C_\delta (A_k^{\mathrm{mic}} + 1) N^{(p-1)/2}, \qquad \chi_N^2 = (\log N)^{-1} N^{-(p-1)/2},$$

which proves (198) using $a_N^{\mathrm{step}} = \delta_N/N$. No fully microscopic bound is claimed for a leader that has become macroscopic.

The failure bound tends to zero. The horizon is polynomial because $d_{0,N}^{-1} \le N^{a_0}$. For large N, $L_{k,N}^{\mathrm{old}} \le 2/(\kappa_{\mathrm{old}} h_N)$, and

$$\frac{(b_{k,N}^{\mathrm{old}})^2 N^2}{2\alpha_1^2 \delta_N^2 L_{k,N}^{\mathrm{old}}} \ge \frac{\kappa_{\mathrm{old}} \sqrt{N}}{4\alpha_1^2 (192 K_\lambda D_{\mathrm{old}})^2 C_\delta d_{0,N}}.$$

This exponent dominates every logarithm of the number of intervals. The other exponents are fixed positive multiples of N. Conditional freshness at s justifies all the same estimates uniformly over the entry state. This proves the lemma.

Stopping-time and first-exit check. At the capped restart s, define

$$\begin{aligned}
T_{\mathrm{geom}} &= \inf\{n \ge 0 : \exists b : q_b(n) \notin [0,1] \text{ or } \|M(n) - \operatorname{diag} M(n)\|_\infty > U_k^{\mathrm{mic}}/\sqrt{N}\},\\
T_{\mathrm{diag}} &= \inf\{n \ge 0 : \min_{a<k} q_a(n) < 1 - \varepsilon_{\mathrm{rec}}\},\\
T_{\mathrm{cross}} &= \inf\{n \ge 0 : \max_{f \in \mathcal{Z}_{k-1}} |m_f(X_{s+n})| > C_{\chi,k}^{\mathrm{in}} \chi_N\},\\
\tau &= T_{\mathrm{geom}} \wedge T_{\mathrm{diag}} \wedge T_{\mathrm{cross}} \wedge Q_{k,N}^{\mathrm{mic}}.
\end{aligned} \tag{202}$$

Use $I_n = \mathbf{1}_{\{n<\tau\}}$. For block length $L = L_{k,N}^{\mathrm{old}}$ and $0 \le m < B_{k,N}^{\mathrm{old}}$, its start $s + mL$ is a stopping time. If F_m is the stopped scalar-noise block failure in (199), Lemma B.9 with scalar variance proxy $\alpha_1^2 \delta_N^2 / N^2$ gives

$$\mathbb{P}(F_m \mid \mathcal{F}_{s+mL}) \le 2r^2 \exp\{-(b_{k,N}^{\mathrm{old}})^2 N^2/(2\alpha_1^2 \delta_N^2 L)\} =: p_{\mathrm{block}}.$$

Consequently $\mathbb{P}(F_m \mid \mathcal{F}_s) = \mathbb{E}[\mathbb{P}(F_m \mid \mathcal{F}_{s+mL}) \mid \mathcal{F}_s] \le p_{\mathrm{block}}$, and the sum over $B_{k,N}^{\mathrm{old}}$ blocks is at most the second allowance in (192). The last shorter block uses the same bound. Gram and single-step diagonal failures are the separate first and third allowances. No success event involving future blocks is inserted into this conditional expectation.

The matrix and diagonal recursions above use only pre-update $n < \tau$, so their resulting bounds hold at $n = \tau$ as well. At a candidate $T_{\mathrm{cross}} = \tau \le T_{\mathrm{geom}} \wedge Q_{k,N}^{\mathrm{mic}}$, (191) and (201) give a strict bound below the crossed line. At a candidate $T_{\mathrm{diag}} = \tau$, the entire-strip inequality gives a positive update if the previous state is in $[b_*, c_*]$, and its jump bound gives a value at least b_* if the previous state is above c_*. Both contradict that candidate exit. This includes ties with geometry exit or the deterministic endpoint; no estimate beyond the geometry exit is claimed. □

Role in the reconstruction. This lemma prevents previously recovered spikes from losing their overlap during the next microscopic race. Lemma B.18 uses its old-cross tube to bound the forcing from recovered rows and columns, and Lemma B.19 uses that bound at every later restart.

EXPLICIT APPLICATION CALCULATIONS

Stopped scalar recursion and the four error budgets For $f = (i, j) \in A_{k-1}$ at a capped restart s, use local time n and absolute time $s + n$:

$$y_f(n) = m_f(X_{s+n}), \quad \Lambda_f = \lambda_i \lambda_j, \quad h_N = \delta_N / \sqrt{N}. \tag{203}$$

The global initialization is always X_{in}; it is not X_s unless $s = 0$.

$$y_f(n+1) = y_f(n) + h_N p \Lambda_f y_f(n)^{p-1} + e_f^{\mathrm{act}}(n) + e_f^{\mathrm{old}}(n) + \xi_f(s+n+1) + r_f(s+n+1), \tag{204}$$

$$e_f^{\mathrm{act}}(n) = -\frac{\delta_N}{N} Q_f^{\mathrm{act}}(X_{s+n}), \qquad e_f^{\mathrm{old}}(n) = -\frac{\delta_N}{N} Q_f^{\mathrm{old}}(X_{s+n}). \tag{205}$$

Proof. Let $I_n = \mathbf{1}_{\{n<\tau\}}$, where τ includes all bootstrap exits and $Q_{k,N}^{\mathrm{mic}}$. This is $\mathcal{F}_{s+n}$-measurable. Lemma B.3, with predictable state X_{s+n}, implies the conditional mgf bound (113) also for $I_n \xi_f(s + n + 1)$: if $I_n = 0$ the mgf is 1, otherwise it is bounded by (113). Lemma B.9, or the already proved scalar form (101), with one time block of length $Q_{k,N}^{\mathrm{mic}}$, gives precisely the simultaneous event (180). On its intersection with the Gram event, at every $u \le \tau$,

$$\begin{aligned}
\left| \sum_{n<u} r_f(s+n+1) \right| &\le Q_{k,N}^{\mathrm{mic}} b_{k,N}^{\mathrm{pol,mic}}, \\
\left| \sum_{n<u} \xi_f(s+n+1) \right| &\le a_{k,N}^{\mathrm{mic}}, \\
\left| \sum_{n<u} e_f^{\mathrm{act}}(n) \right| &\le C_{\mathrm{act}} (U_k^{\mathrm{mic}})^{p+1} \frac{\delta_N Q_{k,N}^{\mathrm{mic}}}{N^{(p+2)/2}}, \\
\left| \sum_{n<u} e_f^{\mathrm{old}}(n) \right| &\le \mathbf{1}_{\{k\ge 2\}} C_{\mathrm{old}} (C_{\chi,k}^{\mathrm{in}})^2 \chi_N^2 \frac{\delta_N Q_{k,N}^{\mathrm{mic}}}{\sqrt{N}}.
\end{aligned} \tag{206}$$

The first bound uses the whole active row bound in (178); the third requires every active entry, including the current leader, to be microscopic. The last is Lemma B.12's old-cross census on its stated interval, independently of whether the leader is microscopic. Adding these four nonnegative bounds gives

$$\max_{u \le \tau} \left| \sum_{n<u} (e_f^{\mathrm{act}}(n) + e_f^{\mathrm{old}}(n) + \xi_f(s+n+1) + r_f(s+n+1)) \right| \le b_{k,N}^{\mathrm{mic}}. \tag{207}$$

This is the additive-error input of Lemma B.10. The separate step-size input is (183), not an extra summand in $b_{k,N}^{\mathrm{mic}}$. All estimates include $u = \tau$, since each summand uses the state at $n < \tau$. □

Clock algebra and the admissible set For $k < r$, put $c = \bar{c}$, $g = \gamma_{+,k}$ and $a_f = \sqrt{N} m_f(X_s)$. The admissible pairs are

$$\mathcal{D}_{k,e,f} = \{(a_e, a_f) \in [\gamma_{2,k}, \gamma_{1,k}]^2 : \Lambda_e a_e^d \ge \Theta_k \Lambda_f a_f^d\}. \tag{208}$$

At stage one this restriction follows from (45), (70); at later stages it follows from $\mathcal{C}_{2,k}^{\mathrm{gr}}$ in (47). It is not an additional assumption.

$$t_e^-(c) = \frac{N^{(p-1)/2}}{pd\delta_N} \frac{1 - [(1-c)a_e/g]^d}{(1-c)^{d+1} \Lambda_e a_e^d}, \tag{209}$$

$$t_f^+(c) = \frac{N^{(p-1)/2}}{pd\delta_N} \frac{1 - [(1+c)a_f/(\rho_k g)]^d}{(1+c)^{d+1} \Lambda_f a_f^d}. \tag{210}$$

These are the lower leader envelope's crossing time and the upper competitor envelope's crossing time; until exits are excluded, they are envelope clocks, not global hitting-time theorems.

Proof. For example the lower envelope equation is

$$\frac{g}{\sqrt{N}} = \frac{(1-c)a_e}{\sqrt{N}} \left(1 - pdh_N(1-c)^{p-1}\Lambda_e a_e^d N^{-d/2} t\right)^{-1/d}.$$

Taking the $-d$ power of the positive ratio gives $1 - pdh_N(1-c)^{p-1}\Lambda_e a_e^d N^{-d/2} t = [(1-c)a_e/g]^d$. Solving for t gives (209); replace $(1-c), g, e$ by $(1+c), \rho_k g, f$ to obtain (210). The removed common factor $K_N = N^{(p-1)/2}/(pd\delta_N)$ is positive because $N > 0, p \geq 3, d = p-2 \geq 1, \delta_N > 0$.

Let $R_k = \Theta_k/B_0 > 1$ and let $L_e = t_e^-(c)/K_N, U_f = t_f^+(c)/K_N$. The leader numerator is positive by (83). Multiplication by the positive number $(1+c)^{d+1}\Lambda_f a_f^d$ and the admissible growth constraint yield

$$\begin{aligned}(1+c)^{d+1}\Lambda_f a_f^d L_e &\leq \frac{B_0}{\Theta_k}\left(1 - \frac{(1-c)^d a_e^d}{g^d}\right) \leq \frac{1}{R_k}\left(1 - \frac{(1-c)^d \gamma_{2,k}^d}{g^d}\right), \\ (1+c)^{d+1}\Lambda_f a_f^d U_f &= 1 - \frac{(1+c)^d a_f^d}{\rho_k^d g^d} \geq 1 - \frac{(1+c)^d \gamma_{1,k}^d}{\rho_k^d g^d}. \end{aligned} \tag{211}$$

Multiplying the strict comparison of the last two expressions by $R_k g^d > 0$ and moving terms gives

$$\begin{aligned} g^d - (1-c)^d \gamma_{2,k}^d &< R_k g^d - R_k (1+c)^d \rho_k^{-d} \gamma_{1,k}^d, \\ (R_k - 1) g^d &> R_k (1+c)^d \rho_k^{-d} \gamma_{1,k}^d - (1-c)^d \gamma_{2,k}^d. \end{aligned} \tag{212}$$

Formula (83) is $g = 1 + \max\{[D/(R_k-1)]_+^{1/d}, (1+c)\gamma_{1,k}/\rho_k\}$ with the right side numerator D in (212); hence $g^d > [D/(R_k-1)]_+$ and (212) follows. Its second entry also gives $(1+c)a_f < \rho_k g$, so the competitor clock numerator is positive. Define the actual positive reserve

$$\Delta_k^{\mathrm{race}} = 1 - \frac{(1+c)^d \rho_k^{-d} \gamma_{1,k}^d}{g^d} - \frac{B_0}{\Theta_k}\left(1 - \frac{(1-c)^d \gamma_{2,k}^d}{g^d}\right) > 0. \tag{213}$$

Thus, uniformly on $\mathcal{D}_{k,e,f}$,

$$t_f^+(c) - t_e^-(c) \geq K_N \frac{\Delta_k^{\mathrm{race}}}{(1+c)^{p-1}\lambda_+^2 \gamma_{1,k}^d}, \tag{214}$$

$$t_e^{\mathrm{pole}}(c) - t_e^-(c) = \frac{K_N}{(1-c)\Lambda_e g^d} \geq \frac{K_N}{(1-c)\lambda_+^2 g^d}. \tag{215}$$

The threshold N_k^{race} requires both lower bounds to be at least two (only (215) if $k = r$). The displayed formulas supply the required margins. □

Regime II: the actual crossing and distinct error budgets Fix a stage k and a capped selection restart s. Relabel the matching independently as in (123); write $\mu_i = \lambda_{i_i^*}$, $\nu_j = \lambda_{j_j^*}$ and $\Lambda_{ij} = \mu_i \nu_j$. Thus $q_n = M_{kk}(X_{s+n})$ has direct coefficient $L_k = \mu_k \nu_k$; relabeling does not identify the row and column weights. Put

$$g = \gamma_{+,k}, \quad \rho = \rho_k, \quad c = c_{\mathrm{ord}}, \quad \kappa = \kappa_{II,k},$$

$$C_T = \frac{8(16/15)^d}{\kappa(1-2^{-d})}, \quad U = 4\rho g, \quad C_* = 3C_D, \tag{216}$$

$$T_N = C_T g^{-d} N^{d/2}, \quad \rho \leq \tfrac{1}{4}(48 C_D C_T)^{-1/d}, \tag{217}$$

$$C_* C_T (U/g)^d = 3 C_D C_T (4\rho)^d \leq 1/16, \tag{218}$$

$$v_N = C_{\chi,k}^{\mathrm{in}} \chi_N \mathbf{1}_{\{k \geq 2\}}, \qquad a_N^{\mathrm{II}} = a_{k,N}^{\mathrm{mic}}. \tag{219}$$

Here a_N^{II} abbreviates $a_{k,N}^{\mathrm{mic}}$ within the Regime-II calculation. The symbol T_N is a bound on elapsed time multiplied by h_N; it is not itself an integer hitting time. Assume the entrance bounds $q_0 \geq g/\sqrt{N}$

and $\gamma_{-,k}/\sqrt{N} < m_f(X_s) < \rho g/\sqrt{N}$ for active nonleaders. At $k = r$ these latter collections are empty. For $q_0 < \varepsilon_{\mathrm{mac}}$, define

$$\begin{aligned} T &= \inf\{n \ge 0 : q_n \ge \varepsilon_{\mathrm{mac}}\}, \quad T_q^- = \inf\{n \ge 0 : q_n < g/(2\sqrt{N})\}, \\ T_u^+ &= \inf\{n \ge 0 : u_n > U/\sqrt{N}\}, \quad u_n = \max_{f \in A_{k-1}\setminus\{e_k\}} |m_f(X_{s+n})|, \\ T_{\mathrm{sur}}^- &= \inf\{n \ge 0 : \min_{f \in A_k} m_f(X_{s+n}) < \gamma_{-,k}/(2\sqrt{N})\}, \\ \nu &= T \wedge T_q^- \wedge T_u^+ \wedge T_{\mathrm{sur}}^- \wedge T_{\mathrm{old}} \wedge H. \end{aligned} \tag{220}$$

Empty maxima are zero; empty-exit times are infinite. Here T_{old} is the first failure of the old diagonal or old cross tube, and $H = Q_{k,N}^{\mathrm{mic}}$ for a fresh Regime-II application. When the old event starts at an earlier stage entrance s_k, replace the last cap by the remaining time $s_k + Q_{k,N}^{\mathrm{mic}} - s$; all noise estimates may still use the larger deterministic bound $Q_{k,N}^{\mathrm{mic}}$. This cap is chosen at s, not after success. The old geometry stop from Lemma B.12 is included when necessary; the bounds below show that all its active coordinates satisfy that geometry through the interval.

Exact recursion and separate errors The two normalized projection monomials in (131) are

$$\frac{p}{2}\mu_u\nu_j M_{iv}M_{uj}^{p-1}M_{uv}, \qquad \frac{p}{2}\mu_u\nu_v M_{iv}M_{uj}M_{uv}^{p-1}.$$

For new row and column entries, retaining the scalar damping and the direct nonlinear power gives

$$\begin{aligned} z_{f,n+1} &= (1 - h_N a_{f,n})z_{f,n} + h_N R_{f,n} + \xi_{f,n+1} + r_{f,n+1}^{\mathrm{pol}}, \\ a_{kb,n} &= \tfrac{p}{2}L_k q_n^p, \quad a_{bk,n} = pL_k q_n^p, \quad a_{ij,n} = 0 \ (i,j > k), \end{aligned} \tag{221}$$

$$|R_{f,n}| \le C_* u_n^{p-1} + C_D q_n^{p-1} u_n^2 + C_{\mathrm{old}} v_N^2. \tag{222}$$

In a future survivor, or in the leader after removing its logistic drift, the remainder bounds are

$$\left|D_f/\sqrt{N} - p\Lambda_f m_f^{p-1}\right| \le C_D(q^{p-1}u^2 + qu^p + u^{p+1}) + C_{\mathrm{old}}v_N^2, \tag{223}$$

$$\left|D_{kk}/\sqrt{N} - pL_k q^{p-1}(1-q^2)\right| \le C_D(q^{p-1}u^2 + qu^p + u^{p+1}) + C_{\mathrm{old}}v_N^2. \tag{224}$$

For the growing row, Lemma B.3's polar identity instead gives

$$|r_{kk,n+1}^{\mathrm{pol}}| \le \frac{\alpha_2\delta_N^2}{2N}\left(q_n + \frac{\sqrt{r}U + C_G\delta_N}{\sqrt{N}}\right), \tag{225}$$

$$P_k = 1 + \frac{2(\sqrt{r}U + C_G)}{g}, \quad L_N^{\mathrm{pol}} = \alpha_2 P_k 2^{d-1} g^{-d}\delta_N N^{(d-1)/2}, \tag{226}$$

$$D_N = \frac{C_D U^2}{N}\left[1 + (2U/g)^d + (2U/g)^{d+1}\right] + C_{\mathrm{old}}v_N^2(2\sqrt{N}/g)^{p-1} + L_N^{\mathrm{pol}}. \tag{227}$$

Here D_N is the displayed relative error budget, not the component drift D_{ij}.

Proof. The index choices $(u,v) = (k,k)$ in (131) give $-pL_k q^{p+1}$ for the leader, $-(p/2)L_k q^p z$ for a new row, and $-pL_k q^p z$ for a new column. The additional new-row term is $-(p/2)\mu_k\nu_b q^2 z^{p-1}$; combined with the direct term its coefficient is $p\mu_k\nu_b(1 - q^2/2)$, bounded in modulus by $p\lambda_+^2$. After these terms have been extracted, the remaining q^{p-1} terms contain two nonleader factors, the remaining q terms contain p such factors, and terms with no q contain $p+1$. The possible index cases are: both $u, v = k$, exactly one equal to k, neither equal to k, or an old index. The last case contains two old-cross factors by Lemma B.12's census. These cases exhaust the sum. Summing at most $2r^2$ projection coefficients of size $p\lambda_+^2/2$, and retaining the direct power, gives (222)–(224); $qu^p, u^{p+1} \le u^{p-1}$ for $q, u \le 1$ gives the coefficient $C_* = 3C_D$ in (222).

The row norm for the leader is at most $q + \sqrt{r}U/\sqrt{N}$, while the Euler change in that row is at most $C_G\delta_N/\sqrt{N}$. Multiplication by $\|B - I\|_{\mathrm{op}} \le \alpha_2\delta_N^2/(2N)$ gives (225). For $q \ge g/(2\sqrt{N})$ and $\delta_N \le 1$, division by $h_N q^{p-1} > 0$ gives

$$\frac{|r_{kk}^{\mathrm{pol}}|}{h_N q^{p-1}} \le \frac{\alpha_2\delta_N P_k}{2\sqrt{N}}q^{-d} \le L_N^{\mathrm{pol}}.$$

Likewise the three active terms in (224), divided by q^{p-1}, are at most U^2/N, $(U^2/N)(2U/g)^d$, and $(U^2/N)(2U/g)^{d+1}$. The old term contributes $C_{\mathrm{old}} v_N^2 (2\sqrt{N}/g)^{p-1}$. This verifies every term in (227). □

Leader growth, the common stop, and time On the simultaneous stopped noise event let

$$S_q(n) = \sum_{t<n} \mathbf{1}_{\{t<\nu\}} \xi_{kk,t+1}, \qquad \widetilde{q}_n = q_n - S_q(n), \quad \max_n |S_q(n)| \le a_N^{\mathrm{II}} \le g/(32\sqrt{N}), \tag{228}$$

$$\tfrac{15}{16} q_n \le \widetilde{q}_n \le \tfrac{17}{16} q_n, \quad \tfrac{16}{17} \widetilde{q}_n \le q_n \le \tfrac{16}{15} \widetilde{q}_n \quad (n < \nu). \tag{229}$$

If $D_N \le \frac{1}{2} p \lambda_-^2 (1 - \varepsilon_{\mathrm{mac}}^2)$, the exact update, (224)–(227), and the definition of κ give

$$\widetilde{q}_{n+1} - \widetilde{q}_n \ge \tfrac{1}{2} p \lambda_-^2 (1 - \varepsilon_{\mathrm{mac}}^2) h_N q_n^{p-1} \ge 2\kappa h_N \widetilde{q}_n^{p-1} > 0 \quad (n < \nu). \tag{230}$$

If $L_N^{\mathrm{pol}} \le 1$, its upper increment is $h_N(C_D + 1)$. Set

$$a_{\mathrm{base,N}} = 15g/(16\sqrt{N}), \quad a_\ell^{II} = \min\{2^\ell a_{\mathrm{base,N}}, \varepsilon_{\mathrm{mac}} + a_N^{\mathrm{II}}\}, \tag{231}$$

$$\widetilde{T}_\ell = \inf\{n : \widetilde{q}_n \ge a_\ell^{II}\}, \quad L_N = 1 + \lceil \log_2(1 + 2\varepsilon_{\mathrm{mac}}/a_{\mathrm{base,N}}) \rceil. \tag{232}$$

The alias $a_{\mathrm{base,N}}$ names the slab base in (231); it does not overwrite the exponent a_0 of (50). Use only the finite set $\mathcal{I}_N = \{\ell \ge 0 : 2^\ell a_{\mathrm{base,N}} < \varepsilon_{\mathrm{mac}} + a_N^{\mathrm{II}}\}$, without repeating the terminal level.

Proof. Monotonicity of $\widetilde{q}$ through the exit update gives

$$q_n \ge q_0 - a_N^{\mathrm{II}} \ge 31g/(32\sqrt{N}) > g/(2\sqrt{N}) \quad (n \le \nu).$$

Thus a candidate $T_q^- = \nu$ contradicts its defining inequality. On the pre-exit portion in a slab starting at $a = 2^\ell a_{\mathrm{base,N}}$, each auxiliary increment is at least $2\kappa h_N a^{d+1}$. If there are m pre-update visits to that slab, the first $m - 1$ increments between visits cover length less than or equal to its width a, so

$$m \le 1 + \frac{1}{2\kappa h_N a^d}. \tag{233}$$

This gives the integer accounting for the slab allowance:

$$H_\ell = 4/(\kappa h_N (a_\ell^{II})^d). \tag{234}$$

Before $\nu \le T$, $\widetilde{q}_n < \varepsilon_{\mathrm{mac}} + a_N^{\mathrm{II}}$, so every pre-update time is in one of these slabs and every slab is visited at most once. Consequently the common stop, even if it is a bad exit, satisfies

$$\nu \le L_N + \frac{a_{\mathrm{base,N}}^{-d}}{2\kappa h_N (1 - 2^{-d})} = L_N + \frac{T_N}{16 h_N},$$

$$h_N L_N \le T_N/2 \quad \Longrightarrow \quad h_N \nu \le 9T_N/16 < T_N. \tag{235}$$

No prior assertion $\nu = T$ was used. In particular the time budget used below to exclude bad exits is already valid.

At any $n \le \nu$, the last pre-update $q_{n-1} < \varepsilon_{\mathrm{mac}}$ and the upper increment give

$$\widetilde{q}_n \le \varepsilon_{\mathrm{mac}} + a_N^{\mathrm{II}} + h_N(C_D + 1) \le 2\varepsilon_{\mathrm{mac}}. \tag{236}$$

Sum (230) and use $q_n \le (16/15)\widetilde{q}_n$ to obtain, through any earlier exit,

$$h_N \sum_{n<\nu} q_n^{p-1} \le (16/15)^{p-1} \frac{\widetilde{q}_\nu - \widetilde{q}_0}{2\kappa} \le \frac{(16/15)^{p-1} \varepsilon_{\mathrm{mac}}}{\kappa} \le S_{1,k}, \tag{237}$$

$$h_N \sum_{n<\nu} q_n^p \le \varepsilon_{\mathrm{mac}} S_{1,k} = S_{2,k}. \tag{238}$$

For $h_N(C_D + 1)\sqrt{N}/g \le 1/2$, consecutive positive auxiliary values have ratio at most $3/2$. Hence

$$(\widetilde{q}_{n+1} - \widetilde{q}_n)\widetilde{q}_n^{-d} \le 2^d \int_{\widetilde{q}_n}^{\widetilde{q}_{n+1}} x^{-d}\, dx, \quad h_N q_n \le \frac{16/15}{2\kappa} (\widetilde{q}_{n+1} - \widetilde{q}_n)\widetilde{q}_n^{-d}.$$

Summing and enlarging only to the already defined reserve gives

$$h_N \sum_{n<\nu} q_n \le J_{k,N}, \tag{239}$$

$$J_{k,N} = \frac{(16/15)2^d}{\kappa} \begin{cases} \log(1 + 2\varepsilon_{\mathrm{mac}}/a_{\mathrm{base,N}}), & d = 1, \\ a_{\mathrm{base,N}}^{1-d}/(d-1), & d > 1. \end{cases} \tag{240}$$

The auxiliary level $\varepsilon_{\mathrm{mac}} + a_N^{\mathrm{II}}$ forces a raw macro crossing; it is not claimed to have the same hitting time as T. □

Upper exits, survivor comparison, and probability Retain the two polar budgets and the two nonleader budgets:

$$P_N^{\mathrm{sur}} = \frac{\alpha_2 \delta_N}{2N} T_N(\sqrt{r}U + C_G\delta_N), \tag{241}$$

$$P_N^{\mathrm{cross}} = \frac{\alpha_2 \delta_N}{2\sqrt{N}} \left[J_{k,N} + \frac{T_N}{\sqrt{N}}(\sqrt{r}U + C_G\delta_N) \right], \tag{242}$$

$$B_N^{\mathrm{up}} = 2a_N^{\mathrm{II}} + P_N^{\mathrm{cross}} + C_D S_{1,k} U^2/N + C_{\mathrm{old}} v_N^2 T_N, \tag{243}$$

$$B_N^{\mathrm{sur}} = a_N^{\mathrm{II}} + P_N^{\mathrm{sur}} + C_D \left(\frac{S_{1,k}U^2}{N} + \frac{\varepsilon_{\mathrm{mac}} T_N U^p}{N^{p/2}} + \frac{T_N U^{p+1}}{N^{(p+1)/2}} \right) + C_{\mathrm{old}} v_N^2 T_N. \tag{244}$$

The definition of N_k^{II} requires all the following for every larger N:

$$\begin{aligned} S_N \ge 1, \quad \delta_N \le 1, \quad U/\sqrt{N} \le 1, \quad v_N \le U/\sqrt{N}, \quad a_{\mathrm{base,N}} \le \varepsilon_{\mathrm{mac}}, \quad h_N p \lambda_+^2 \le 1/2, \\ a_N^{\mathrm{II}} \le g/(32\sqrt{N}), \quad D_N \le \tfrac{1}{2} p\lambda_-^2(1 - \varepsilon_{\mathrm{mac}}^2), \quad L_N^{\mathrm{pol}} \le 1, \\ a_N^{\mathrm{II}} + h_N(C_D + 1) \le \varepsilon_{\mathrm{mac}}, \quad h_N(C_D+1)\sqrt{N}/g \le 1/2, \\ h_N L_N \le T_N/2, \quad T_N \le h_N Q_{k,N}^{\mathrm{mic}}, \quad B_N^{\mathrm{up}} \le U/(16\sqrt{N}), \\ B_N^{\mathrm{sur}} \le c\gamma_{-,k}/[32(p-1)\sqrt{N}], \quad h_N p \lambda_+^2 (U/\sqrt{N})^d \le c/[64(p-1)]. \end{aligned} \tag{245}$$

Proof. Since $0 \le a_{f,n} \le p\lambda_+^2$ and $h_N p\lambda_+^2 \le 1/2$, the scalar products obtained by iterating (221) lie in $[0,1]$ and are nondecreasing as functions of the noise-increment index. Lemma B.4 applies pathwise to these products with $\zeta_j = \mathbf{1}_{\{j-1<\nu\}}\xi_{f,j}$ and $\max|S_j| \le a_N^{\mathrm{II}}$, so their weighted noise sum is at most $2a_N^{\mathrm{II}}$. Predictability of the terminal products is not needed. The polar row bounds integrated with (239) and $h_N\nu \le T_N$ are exactly (241)–(242). Iteration therefore yields, for $n \le \nu$,

$$u_n \le \rho g/\sqrt{N} + 2a_N^{\mathrm{II}} + P_N^{\mathrm{cross}} + C_* h_N \sum_{t<n} u_t^{p-1} + C_D h_N \sum_{t<n} q_t^{p-1} u_t^2 + C_{\mathrm{old}} v_N^2 h_N n. \tag{246}$$

The direct power contribution is at most

$$C_* T_N (U/\sqrt{N})^{d+1} = \frac{U}{\sqrt{N}} C_* C_T (U/g)^d \le \frac{U}{16\sqrt{N}}.$$

Use (243), (245) and $\rho g = U/4$ to get

$$u_n \le \frac{U}{4\sqrt{N}} + \frac{U}{16\sqrt{N}} + \frac{U}{16\sqrt{N}} = \frac{3U}{8\sqrt{N}} < \frac{U}{\sqrt{N}} \le \frac{L_k^{\mathrm{in}}}{\sqrt{N}}. \tag{247}$$

This contradicts a candidate $T_u^+ = \nu$, including a tie with another exit.

For $f \in A_k$, Lemma B.10 is now applied with $y_n = m_f(X_{s+n})$, $a = h_N p \Lambda_f$, $d = p-2$, $b = B_N^{\mathrm{sur}}$, $c = c_{\mathrm{ord}}$, initial $y_0 > \gamma_{-,k}/\sqrt{N}$, and stopped interval $[0, \nu]$. The two required inequalities follow separately from the last line of (245). To verify the denominator on this whole interval, use (217) and $n \le T_N/h_N$:

$$\begin{aligned} pdh_N(1+c)^{p-1} G_f(X_s) n &\le pd(1+c)^{p-1}\lambda_+^2 C_T \rho^d \\ &\le \frac{d(5/4)^{d+1}}{48(1+r^2)4^d} = \frac{5d(5/16)^d}{192(1+r^2)} < \frac{1}{4}. \end{aligned} \tag{248}$$

Here $c \le 1/4$, $C_D = p\lambda_+^2(1+r^2)$, and $d(5/16)^d \le 1$ for integer $d \ge 1$. Thus the upper denominator is at least $3/4$; the lower denominator is larger. At a candidate lower exit the endpoint completion of Lemma B.10 yields

$$y_f(\nu) \ge \underline{y}_{f,c}(\nu) \ge (1-c)y_f(0) > 3\gamma_{-,k}/(4\sqrt{N}) > \gamma_{-,k}/(2\sqrt{N}),$$

contradicting $T^-_{\mathrm{sur}} = \nu$. No survivor is lost.

For every future pair whose entrance ratio is larger than $B(c)$, both full-interval envelopes are defined by (248); their division gives

$$\frac{G_{e_\ell}(X_{s+n})}{G_f(X_{s+n})} \ge A(c)^d \frac{G_{e_\ell}(X_s)}{G_f(X_s)} \frac{1 - pdh_N(1+c)^{p-1}G_f(X_s)n}{1 - pdh_N(1-c)^{p-1}G_{e_\ell}(X_s)n} \ge A(c)^d \frac{G_{e_\ell}(X_s)}{G_f(X_s)}. \tag{249}$$

The final ratio of denominators is at least one for the same reason as in (281). The application is made once on the whole interval; the dyadic layers were not comparison restarts.

On the old-block event Lemma B.12 excludes old exits through the same updates, because all nonleader magnitudes stay below $U/\sqrt{N} \le U_k^{\mathrm{mic}}/\sqrt{N}$, the old crosses are at most $v_N \le U/\sqrt{N}$, and every diagonal is nonnegative. The current growing diagonal is allowed to be macroscopic in that lemma. The $H = Q_{k,N}^{\mathrm{mic}}$ cutoff is excluded by $\nu < T_N/h_N \le H$, unless the target has already occurred. In the earlier-stage absolute-horizon version, the explicit remaining-time calculation (256) below gives the same conclusion. Thus $\nu = T$, and at this single time

$$|m_f(X_{s+T})| \le C_{0,k}g/\sqrt{N}\ (f \in \mathcal{X}_k), \quad \frac{\gamma_{-,k}}{C_{0,k}\sqrt{N}} \le m_f(X_{s+T}) \le \frac{C_{0,k}g}{\sqrt{N}}\ (f \in A_k). \tag{250}$$

Use $C_{0,k} \ge 2$, $U \le g \le C_{0,k}g$, and the lower survivor bound just proved. The stochastic event is the intersection of the Gram event and the single stopped simultaneous martingale event at threshold a_N^{II}, constructed conditionally on $\mathcal{F}_s$. Its conditional failure is bounded by $Q_{k,N}^{\mathrm{mic}} e^{-c_R N} + N^{-3}$. The old-block event is intersected afterwards, with its separately proved probability; it is not used to assert conditional freshness. At $k = r$ omit the empty cross and survivor computations and their tests. □

Stage substitution and the ordering handoff *Proof.* At a stage entrance s_k with $X_{s_k} \in E_{k-1}^{\sharp}$, (49) gives, on this measurable entrance event,

$$\gamma_{2,k}/\sqrt{N} \le m_f(X_{s_k}) \le \gamma_{1,k}/\sqrt{N}\ (f \in A_{k-1}), \qquad G_{e_\ell}(X_{s_k})/G_f(X_{s_k}) \ge \Theta_k. \tag{251}$$

The fixed-reference error and accumulated old displacement are

$$\varepsilon_{k,N}^{\mathrm{rel}} = \frac{C_{\mathrm{act}}(U_k^{\mathrm{mic}})^{p+1}}{p\lambda_-^2\gamma_{2,k}^{p-1}N} + \frac{C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2}{p\lambda_-^2\gamma_{2,k}^{p-1}\log N}, \tag{252}$$

$$\varepsilon_{k,N}^{\mathrm{acc}} = \frac{C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2 C_\delta(A_k^{\mathrm{mic}}+1)}{\sqrt{N}\log N}, \tag{253}$$

$$\sqrt{N}\varepsilon_{k,N}^{\mathrm{acc}} = \frac{C_{\mathrm{old}}(C_{\chi,k}^{\mathrm{in}})^2 C_\delta(A_k^{\mathrm{mic}}+1)}{\log N}. \tag{254}$$

The designation "rel" in (252) uses the fixed reference drift in (197), used in (314). If $m_f \ge \gamma_{2,k}/\sqrt{N}$, it also bounds the relative error for the actual direct drift by (318). We do not assert this stronger state condition after the lower guard is relaxed. The stopped additive-error comparison uses (253).

By the strengthened selection clock (278) with stage-k substitution, $T_I \le \lceil A_k^{\mathrm{sel}} S_N \rceil$. The Regime-II common-stop estimate, already valid before excluding other exits, gives

$$T_{II} \le T_N/h_N = \frac{C_T}{C_\delta g^d} S_N = (A_k^{\mathrm{mac}} - 2)S_N. \tag{255}$$

Therefore, with $S_N \ge 1$,

$$T_I + T_{II} \le A_k^{\mathrm{sel}} S_N + 1 + (A_k^{\mathrm{mac}} - 2)S_N = A_k^{\mathrm{mic}} S_N - 2S_N + 1 < A_k^{\mathrm{mic}} S_N \le Q_{k,N}^{\mathrm{mic}}. \tag{256}$$

In particular the remaining absolute cutoff at selection is strictly larger than the Regime-II bound. The old event chosen on $[s_k, s_k + Q_{k,N}^{\mathrm{mic}}]$ covers both regimes; it is never extended to $s + Q_{k,N}^{\mathrm{mic}}$.

For $k < r$, at a macroscopic entrance σ the survivor band is $[4\gamma_{2,k+1}, \gamma_{1,k+1}/2]/\sqrt{N}$. If cleanup gives $|m_f(X_{\widehat{\tau}_k}) - m_f(X_\sigma)| \le \omega_{k,N}/\sqrt{N}$ and $\omega_{k,N} \le \gamma_{2,k+1}$, set

$$a = \omega_{k,N}/(4\gamma_{2,k+1}) \le 1/4, \qquad \kappa_{k,N} = \left(\frac{1-a}{1+a}\right)^d. \tag{257}$$

Then $(1-a)m_f(X_\sigma) \le m_f(X_{\widehat{\tau}_k}) \le (1+a)m_f(X_\sigma)$ for all survivors. Combining the two whole-interval ordering losses with these positive inequalities gives

$$\frac{G_{e_\ell}(X_{\widehat{\tau}_k})}{G_f(X_{\widehat{\tau}_k})} \ge \kappa_{k,N} A(c_{\mathrm{ord}})^{2d} \Theta_k. \tag{258}$$

The tail threshold is

$$N_k^{\mathrm{ord}} = \min\{n : \kappa_{k,N} A(c_{\mathrm{ord}})^{2d} \ge B_0^{-1} \text{ for every } N \ge n\}. \tag{259}$$

Its finiteness follows from $\omega_{k,N} \to 0$ and $A(c_{\mathrm{ord}})^{2d} > B_0^{-1}$ in (76). For $N \ge \max\{N_k^{\mathrm{cl}}, N_k^{\mathrm{ord}}\}$, (258) is at least $B_0^{-1}\Theta_k = \Theta_{k+1}$. At $k = r$ there are no such ratios, and $N_r^{\mathrm{ord}} = 1$ is our convention for empty future maxima. □

FIRST STAGE: THREE REGIMES AND EXACT CLOSURE

Write $q_1(X) = m_{e_1}(X)$ and retain the following sets:

$$\mathcal{O}_1^{\mathrm{fut}}(S) = \bigcap_{\ell=2}^{r} \bigcap_{f \in A_{\ell-1}\setminus\{e_\ell\}} \{X : G_{e_\ell}(X) \ge S G_f(X)\}, \tag{260}$$

$$E_{\mathrm{sel},1} = \{X : q_1(X) \ge \gamma_{+,1}/\sqrt{N},\ \gamma_{-,1}/\sqrt{N} < m_f(X) < \rho_1\gamma_{+,1}/\sqrt{N}\ (f \ne e_1)\}, \tag{261}$$

$$\mathcal{X}_1 = (\{i_1^*\} \times J_1) \cup (I_1 \times \{j_1^*\}), \tag{262}$$

$$E_{\mathrm{mac},1} = \left\{ X : \begin{array}{l} q_1(X) \ge \varepsilon_{\mathrm{mac}}, \quad |m_f(X)| \le C_{0,1}\gamma_{+,1}/\sqrt{N}\ (f \in \mathcal{X}_1), \\ 4\gamma_{2,2}/\sqrt{N} \le m_f(X) \le \gamma_{1,2}/(2\sqrt{N})\ (f \in A_1), \\ X \in \mathcal{O}_1^{\mathrm{fut}}(A(c_{\mathrm{ord}})^{2d}\Theta_1) \end{array} \right\}, \tag{263}$$

$$E_{\mathrm{mac},1}^{\mathrm{in}} = E_{\mathrm{mac},1}, \tag{264}$$

$$4\gamma_{2,2} = \gamma_{-,1}/C_{0,1}, \qquad \gamma_{1,2}/2 = C_{0,1}\gamma_{+,1}. \tag{265}$$

In (261), $f \in A_0 \setminus \{e_1\}$ is understood.

Regime I

Lemma B.13 (Regime I: microscopic leader selection)**.** *Assume $N \ge N_1^{\mathrm{mic}}$, set $C_{I,1} = A_1^{\mathrm{mic}}$ and $\eta_{I,1,N} = \eta_{1,N}^{\mathrm{cmp}}$. Uniformly over $X_{\mathrm{in}} \in \mathcal{C}_1^{\mathrm{init}} \cap \mathcal{C}_2^{\mathrm{init}}$, there is an event $\mathcal{G}_{I,1,N}$ with*

$$\mathbb{P}_{X_{\mathrm{in}}}(\mathcal{G}_{I,1,N}) \ge 1 - \eta_{I,1,N}, \tag{266}$$

$$\sigma_{\mathrm{sel},1} := T_{E_{\mathrm{sel},1}} \le Q_{I,1,N} := \lceil C_{I,1} d_{0,N}^{-1} (\log N) N^{p-2} \rceil, \tag{267}$$

$$X_{\sigma_{\mathrm{sel},1}} \in \mathcal{O}_1^{\mathrm{fut}}(A(c_{\mathrm{ord}})^d \Theta_1). \tag{268}$$

The leader reaches its selection line strictly before any competitor reaches its line; no active overlap first exits through its lower guard.

Proof. Step 1. State, stop, and common event. Set $y_f(n) = m_f(X_n)$, $a_f = \sqrt{N} m_f(X_{\mathrm{in}})$, $\Lambda_f = \lambda_i\lambda_j$, $e = e_1$, $g = \gamma_{+,1}$ and $c = \bar{c}$. With $\inf \varnothing = \infty$, define

$$\begin{aligned} T_e &= \inf\{n \in \mathbb{Z}_{\ge 0} : y_e(n) \ge g/\sqrt{N}\}, \\ T_f^+ &= \inf\{n \in \mathbb{Z}_{\ge 0} : y_f(n) \ge \rho_1 g/\sqrt{N}\}, \quad T_+ = \min_{f \in A_0\setminus\{e\}} T_f^+, \\ T_- &= \inf\{n \in \mathbb{Z}_{\ge 0} : \min_{f \in A_0} y_f(n) \le \gamma_{-,1}/\sqrt{N}\}, \\ \tau &= T_e \wedge T_+ \wedge T_- \wedge Q_{I,1,N}. \end{aligned} \tag{269}$$

The entrance conditions give $a_f \in [\gamma_{2,1}, \gamma_{1,1}]$, and (83) gives $\rho_1 g > (1+c)\gamma_{1,1}$. Hence none of the three state exits occurs at zero. Before τ, all active coordinates are positive and at most $g/\sqrt{N} \le U_1^{\mathrm{mic}}/\sqrt{N}$; the latter follows from $U_1^{\mathrm{mic}} = 2 + L_1^{\mathrm{out}} + C_{\chi,1}$, $L_1^{\mathrm{out}} = 2K_\lambda C_{0,1} g + 8$, and $C_{0,1} \ge 2$. Also $Q_{I,1,N} = Q_{1,N}^{\mathrm{mic}}$.

Use $\mathcal{G}_{I,1,N} = \mathcal{G}_{1,N}^{\mathrm{cmp}}(0,\tau)$ from (180). For each coordinate, the substitution in Lemma B.10 is

$$(y_n, a, d, b, y_0, H) = \big(y_f(n), h_N p \Lambda_f, p-2, b_{1,N}^{\mathrm{mic}}, a_f/\sqrt{N}, Q_{I,1,N}\big).$$

By (183), $b \le c_{\mathrm{ord}} y_0/[32(d+1)]$ and $a y_n^d \le c_{\mathrm{ord}}/[64(d+1)]$ at every pre-update state. Since $0 < c_{\mathrm{ord}} \le c < 1/2$, both the c and the c_{ord} applications are valid on this one event, with their respective denominator domains. Its conditional failure bound at time zero is (181), so no independent event or second charge is needed.

For $c' \in \{c, c_{\mathrm{ord}}\}$ write the concrete envelopes

$$\begin{aligned} \underline{y}_{f,c'}(n) &= \frac{(1-c')a_f}{\sqrt{N}}[1 - pdh_N(1-c')^{p-1} G_f(X_{\mathrm{in}}) n]^{-1/d}, \\ \overline{y}_{f,c'}(n) &= \frac{(1+c')a_f}{\sqrt{N}}[1 - pdh_N(1+c')^{p-1} G_f(X_{\mathrm{in}}) n]^{-1/d}. \end{aligned} \tag{270}$$

Lemma B.10 gives $\underline{y}_{f,c'}(n) \le y_f(n) \le \overline{y}_{f,c'}(n)$ through $n \le \tau$ where the relevant denominators are positive. The endpoint assertion is exactly (170)–(172), not an assumption about the exit state.

Step 2. Clock separation on the correct domain. Specializing (209)–(210) to $k = 1$ and dividing by $K_N > 0$ gives

$$L_e = \frac{1 - [(1-c)a_e/g]^d}{(1-c)^{d+1}\Lambda_e a_e^d}, \tag{271}$$

$$U_f = \frac{1 - [(1+c)a_f/(\rho_1 g)]^d}{(1+c)^{d+1}\Lambda_f a_f^d}. \tag{272}$$

For each competitor, (45) and the fixed greedy choice give $\Lambda_e a_e^d \ge \Theta_1 \Lambda_f a_f^d$. By (211), first using this inequality, then the amplitude endpoints, and finally (83),

$$\sup_{(a_e,a_f)\in\mathcal{D}_{1,e,f}} (L_e - U_f) \le -\frac{\Delta_1^{\mathrm{race}}}{(1+c)^{p-1}\lambda_+^2 \gamma_{1,1}^d} < 0. \tag{273}$$

The restricted supremum is essential. The admissible set in (208) records the hypothesis used to take this supremum.

Step 3. Integer clocks, poles, and the prescribed time constant. Let $t_c = t_e^-(c)$ and $t_o = t_e^-(c_{\mathrm{ord}})$, with t_e^- as in (209), and put

$$n_* = \lceil t_o \rceil, \qquad \bar{n} = \lceil t_c \rceil. \tag{274}$$

For $t = 1 - c' \in (0,1)$, the normalized clock is

$$\frac{1}{\Lambda_e}\big(a_e^{-d} t^{-d-1} - g^{-d} t^{-1}\big).$$

Its derivative is $\Lambda_e^{-1} t^{-2}[g^{-d} - (d+1)a_e^{-d} t^{-d}] < 0$, because $g > a_e$. Thus $c_{\mathrm{ord}} \le c$ implies $t_o \le t_c$ and $n_* \le \bar{n}$. The race threshold and (214)–(215) give

$$\bar{n} \le t_c + 1 < t_f^+(c), \qquad \bar{n} \le t_c + 1 < t_e^{\mathrm{pole}}(c). \tag{275}$$

The c_{ord} lower pole also survives rounding, without changing N_1^{race}:

$$\begin{aligned} t_e^{\mathrm{pole}}(c_{\mathrm{ord}}) - t_o &= \frac{K_N}{(1-c_{\mathrm{ord}})\Lambda_e g^d} = \frac{1-c}{1-c_{\mathrm{ord}}}\,[t_e^{\mathrm{pole}}(c) - t_c] \\ &\ge \frac{3}{4}\cdot 2 = \frac{3}{2} > 1, \qquad n_* < t_e^{\mathrm{pole}}(c_{\mathrm{ord}}). \end{aligned} \tag{276}$$

For every nonleader f, $G_e \geq \Theta_1 G_f$ and $\Theta_1 > B_0$ imply

$$(1 + c_{\mathrm{ord}})^{p-1} G_f \leq (1 + c)^{p-1} G_f < (1 - c)^{p-1} G_e. \tag{277}$$

Its upper c_{ord} pole, and therefore its lower c_{ord} pole, lie after the leader's lower c pole. The same holds with c_{ord} replaced by c for a nonleader. All nonleader envelopes needed up to n_* are consequently defined.

Let $S_N = d_{0,N}^{-1}(\log N) N^d$. The threshold N_1^{mic} gives $N \geq 3, d_{0,N} \leq 1$, so $S_N \geq 1$. From (50) and (173),

$$\begin{aligned} n_* &\leq 1 + \frac{N^{(p-1)/2}}{pd\delta_N (1 - c_{\mathrm{ord}})^{p-1} \Lambda_e a_e^d} \\ &\leq 1 + \frac{S_N}{pdC_\delta (1 - c_{\mathrm{ord}})^{p-1} \lambda_-^2 \gamma_{2,1}^d} \\ &\leq \left(2 + \frac{4}{pdC_\delta (1 - c_{\mathrm{ord}})^{p-1} \lambda_-^2 \gamma_{2,1}^d} \right) S_N = A_1^{\mathrm{sel}} S_N \leq \lceil A_1^{\mathrm{mic}} S_N \rceil = Q_{I,1,N}. \end{aligned} \tag{278}$$

Using the earlier c_{ord} clock gives the stated selection bound. Merely substituting $c = \bar{c}$ into the definition of A_1^{sel}, which contains c_{ord}, would not establish this chain.

Step 4. Excluding every first exit, including ties and the endpoint. For $n \leq \tau \wedge n_*$, all lower c envelopes are defined. Their monotonicity and (82) give, for every $f \in A_0$,

$$y_f(n) \geq \underline{y}_{f,c}(n) \geq (1 - c) m_f(X_{\mathrm{in}}) \geq \frac{(1 - c)\gamma_{2,1}}{\sqrt{N}} = \frac{2\gamma_{-,1}}{\sqrt{N}} > \frac{\gamma_{-,1}}{\sqrt{N}}. \tag{279}$$

If $T_- = \tau \leq n_*$, the definition of T_- supplies a coordinate at most $\gamma_{-,1}/\sqrt{N}$, contradicting (279) at that very update. For a competitor and $n \leq \tau \wedge n_*$, (275) gives $n < t_f^+(c)$ and hence

$$y_f(n) \leq \overline{y}_{f,c}(n) < \rho_1 g/\sqrt{N}.$$

Thus $T_+ = \tau \leq n_*$ is impossible, even simultaneously with T_e, T_-, or the cutoff. If $\tau > n_*$, the main lower c_{ord} envelope gives $y_e(n_*) \geq g/\sqrt{N}$, contradicting $T_e > n_*$. Therefore $\tau \leq n_*$. The preceding strict inequalities exclude the bad state exits. If $\tau = Q_{I,1,N} \leq n_*$, (278) forces $\tau = n_*$ and the same main lower envelope gives $T_e \leq \tau$; the cutoff cannot be the only exit. We have proved

$$T_e = \tau \leq n_*, \qquad T_e < T_+ \wedge T_-, \qquad X_{T_e} \in E_{\mathrm{sel},1}, \qquad T_{E_{\mathrm{sel},1}} = T_e. \tag{280}$$

The last equality holds because membership of $E_{\mathrm{sel},1}$ requires the leader threshold, which was not reached before T_e.

Step 5. All future ratios and the success event. Fix $2 \leq \ell \leq r$ and $f \in A_{\ell-1} \setminus \{e_\ell\}$. Both indices are nonleaders at stage one. At entrance $G_{e_\ell}/G_f \geq \Theta_1 > B(c_{\mathrm{ord}})$ by (70), (76). By (277), their needed denominators are positive through T_e. Raising the two envelope bounds to $d > 0$ and dividing gives

$$\begin{aligned} \frac{G_{e_\ell}(X_n)}{G_f(X_n)} &\geq A(c_{\mathrm{ord}})^d \frac{G_{e_\ell}(X_{\mathrm{in}})}{G_f(X_{\mathrm{in}})} \frac{1 - pdh_N(1 + c_{\mathrm{ord}})^{p-1} G_f(X_{\mathrm{in}}) n}{1 - pdh_N(1 - c_{\mathrm{ord}})^{p-1} G_{e_\ell}(X_{\mathrm{in}}) n} \\ &\geq A(c_{\mathrm{ord}})^d \Theta_1 \qquad (0 \leq n \leq T_e). \end{aligned} \tag{281}$$

Indeed $G_{e_\ell} > B(c_{\mathrm{ord}}) G_f$ implies that the coefficient subtracted in the denominator is larger than that subtracted in the numerator. The positive numerator is at least the positive denominator, so their ratio is at least one. Evaluating at $T_e = \sigma_{\mathrm{sel},1}$ gives (268) simultaneously for every prescribed pair, with no new random event. We have shown

$$\mathcal{G}_{I,1,N} \subseteq \{\sigma_{\mathrm{sel},1} \leq Q_{I,1,N},\ X_{\sigma_{\mathrm{sel},1}} \in E_{\mathrm{sel},1} \cap \mathcal{O}_1^{\mathrm{fut}}(A(c_{\mathrm{ord}})^d \Theta_1),\ T_e < T_+ \wedge T_-\}.$$

Together with (181), this proves (266)–(268). □

Role in the reconstruction. The first leader wins the microscopic race while every future ratio retains its directed margin. Lemma B.16 uses this selected state as the entrance to Lemma B.14.

Regime II

Lemma B.14 (Regime II: microscopic-to-macroscopic transport)**.** *Assume* $N \geq \max\{N_1^{\mathrm{mic}}, N_1^{II}\}$. *At a capped stopping time* σ*, suppose*

$$X_\sigma \in E_{\mathrm{sel},1} \cap \mathcal{O}_1^{\mathrm{fut}}(A(c_{\mathrm{ord}})^d \Theta_1). \tag{282}$$

Set $C_{II,1} = A_1^{\mathrm{mic}}$, $\eta_{II,1,N} = \eta_{1,N}^{\mathrm{cmp}}$. *With conditional probability at least* $1 - \eta_{II,1,N}$,

$$\begin{gathered} \sigma_{\mathrm{mac},1}^{[\sigma]} := \inf\{t \geq \sigma : q_1(X_t) \geq \varepsilon_{\mathrm{mac}}\}, \\ \sigma_{\mathrm{mac},1}^{[\sigma]} \leq \sigma + Q_{II,1,N}, \qquad X_{\sigma_{\mathrm{mac},1}^{[\sigma]}} \in E_{\mathrm{mac},1}, \\ Q_{II,1,N} = \lceil C_{II,1} S_N \rceil. \end{gathered} \tag{283}$$

The superscript specifies the lower time bound of this crossing. It distinguishes this restart crossing from the first crossing since stage entrance.

Proof. Apply the Regime-II argument above with

$$(k, s, g, \rho, U, v_N, \kappa, a_N^{\mathrm{II}}, H) = (1, \sigma, \gamma_{+,1}, \rho_1, 4\rho_1\gamma_{+,1}, 0, \kappa_{II,1}, a_{1,N}^{\mathrm{mic}}, Q_{1,N}^{\mathrm{mic}}).$$

The selected state gives $q_0 \geq g/\sqrt{N}$, all nonleaders in $(\gamma_{-,1}, \rho_1 g)/\sqrt{N}$, and all future ratios at least $A(c_{\mathrm{ord}})^d\Theta_1 > B(c_{\mathrm{ord}})$ by (76). Each numerical condition in (245) is supplied by $N \geq N_1^{II}$; all terms involving v_N vanish. The only stochastic inputs are the simultaneous stopped martingale event at a_N^{II} and the Gram event on H steps, with conditional failure at most $\eta_{1,N}^{\mathrm{cmp}}$.

If $q_0 \geq \varepsilon_{\mathrm{mac}}$, the hitting time is σ. The nonleader entrance bounds imply the output bounds because $C_{0,1} \geq 2$, $\rho_1 \leq 1/4$ and (265); the entrance ratio is larger than the required $A(c_{\mathrm{ord}})^{2d}\Theta_1$. Otherwise the Regime-II argument above gives the actual local crossing T, not the auxiliary crossing of $\widetilde{q}$, with

$$\widetilde{q}_n = q_1(X_{\sigma+n}) - \sum_{t<n} \xi_{e_1}(\sigma + t + 1), \quad |q_1(X_{\sigma+n}) - \widetilde{q}_n| \leq a_{1,N}^{\mathrm{mic}}, \tag{284}$$

$$\widetilde{q}_{n+1} - \widetilde{q}_n \geq 2\kappa_{II,1} h_N \widetilde{q}_n^{p-1} \quad (n < \nu). \tag{285}$$

The relative leader polar budget is (227); new crosses use (242), survivors use (241) and (244). The common-stop argument of the Regime-II argument above gives $h_N T \leq T_N$ and excludes each bad exit through its updated state. Its concrete output is

$$1 - pdh_N(1 + c_{\mathrm{ord}})^{p-1} G_f(X_\sigma) T \geq 3/4 \quad (f \in A_1), \tag{286}$$

$$|m_f(X_{\sigma+T})| \leq C_{0,1}\gamma_{+,1}/\sqrt{N} \quad (f \in \mathcal{X}_1), \tag{287}$$

$$\frac{\gamma_{-,1}}{C_{0,1}\sqrt{N}} \leq m_f(X_{\sigma+T}) \leq \frac{C_{0,1}\gamma_{+,1}}{\sqrt{N}} \quad (f \in A_1). \tag{288}$$

By (265), these are precisely the cross and interior survivor parts of $E_{\mathrm{mac},1}$. One whole-interval application of Lemma B.10 to each survivor, followed by the division in (249), yields

$$\frac{G_{e_\ell}(X_{\sigma+T})}{G_f(X_{\sigma+T})} \geq A(c_{\mathrm{ord}})^d \frac{G_{e_\ell}(X_\sigma)}{G_f(X_\sigma)} \geq A(c_{\mathrm{ord}})^{2d}\Theta_1. \tag{289}$$

Together with $q_1(X_{\sigma+T}) \geq \varepsilon_{\mathrm{mac}}$ these show $X_{\sigma+T} \in E_{\mathrm{mac},1}$ at one time. Also $T \leq T_N/h_N = (A_1^{\mathrm{mac}} - 2)S_N \leq Q_{II,1,N}$. No layer count is charged to the ordering comparison or to the probability. □

Role in the reconstruction. This supplies the first actual macroscopic crossing and a simultaneous survivor band. Lemma B.15 cleans the selected row and column from precisely this entrance state, and Lemma B.16 concatenates the three regimes.

Regime III

Lemma B.15 (Regime III: cleanup and preservation of the future maximum)**.** *Assume* $N \geq \max\{N_1^{\mathrm{cl}}, N_1^{\mathrm{ord}}\}$ *and* $X_\sigma \in E_{\mathrm{mac},1}$ *at a capped stopping time. Set* $A_{III,1} = A_1^{\mathrm{cl}}$, $\eta_{III,1,N} = \eta_{1,N}^{\mathrm{cl}}$ *and*

$$H_{III,1,N} = \lceil A_{III,1} h_N^{-1} \log N \rceil, \qquad \widehat{\tau}_1 = \sigma + H_{III,1,N}. \tag{290}$$

With conditional probability at least $1 - \eta_{III,1,N}$,

$$X_{\widehat{\tau}_1} \in \mathcal{R}_1^{[1]}(\varepsilon_{\mathrm{rec}}), \tag{291}$$

$$\max_{f \in A_1} \max_{n \leq H_{III,1,N}} |m_f(X_{\sigma+n}) - m_f(X_\sigma)| \leq \omega_{1,N}/\sqrt{N}, \tag{292}$$

$$\kappa_{1,N} = \left(\frac{1 - \omega_{1,N}/(4\gamma_{2,2})}{1 + \omega_{1,N}/(4\gamma_{2,2})} \right)^d \longrightarrow 1, \tag{293}$$

$$X_{\widehat{\tau}_1} \in \mathcal{O}_1^{\mathrm{fut}}(\kappa_{1,N} A (c_{\mathrm{ord}})^{2d} \Theta_1). \tag{294}$$

Proof. Apply Lemma B.20 with $k = 1$, $E_{\mathrm{mac},1}^{\mathrm{in}} = E_{\mathrm{mac},1}$, $\mathcal{I}_0 = \varnothing$, $\mathcal{I}_1 = \{(a,b) : a < b, a \leq 1\}$, $H = H_{III,1,N}$ and $U = U_1^{\mathrm{cl}}$. The absence of old blocks means $W_0 = 0$. The nonempty maximum has $W_1(0) \leq K_\lambda L_1^{\mathrm{in}}/\sqrt{N}$. The conditions (351)–(361) at $k = 1$ verify the contraction, coupling, error-floor, strip and survivor budgets; $N \geq N_1^{\mathrm{cl}}$ supplies exactly those conditions. Lemma B.20 gives the new diagonal bound (366), cross bound (367), and freezing bound (368), with $\eta_{1,N}^{\mathrm{cl}}$ from (362). All coordinates of the new row and column are in $\mathcal{X}_1$, so the first two bounds yield (291), and the third yields (292).

For $u_f = m_f(X_\sigma) \geq 4\gamma_{2,2}/\sqrt{N}$, put $a = \omega_{1,N}/(4\gamma_{2,2}) \leq 1/4$ by (361). Then

$$|m_f(X_{\widehat{\tau}_1}) - u_f| \leq \omega_{1,N}/\sqrt{N} \leq a u_f, \quad (1-a)u_f \leq m_f(X_{\widehat{\tau}_1}) \leq (1+a)u_f. \tag{295}$$

For $\ell \geq 2$, raising the lower numerator and upper denominator bounds to d gives

$$\frac{G_{e_\ell}(X_{\widehat{\tau}_1})}{G_f(X_{\widehat{\tau}_1})} \geq \left(\frac{1-a}{1+a} \right)^d \frac{G_{e_\ell}(X_\sigma)}{G_f(X_\sigma)} \geq \kappa_{1,N} A (c_{\mathrm{ord}})^{2d} \Theta_1. \tag{296}$$

This proves (294); no additional random input is used. □

Role in the reconstruction. This is the $k = 1$ specialization of the uniform cleanup in Lemma B.20. Its survivor-freezing bound and future ratios are used in Lemma B.16 to establish the first certified sharp set.

Exact entrance into the first sharp set

Lemma B.16 (Regime-to-sharp: closed first-stage transition)**.** *Retain*

$$N_{\mathrm{fs}} = \max\{N_1^{\mathrm{mic}}, N_1^{II}, N_1^{\mathrm{cl}}, N_1^{\mathrm{ord}}, N_1^{\mathrm{ceil}}\}, \tag{297}$$

$$C_{\mathrm{fs}} = 2A_1^{\mathrm{mic}} + A_1^{\mathrm{cl}} + 3, \quad \eta_{\mathrm{fs},N} = 2\eta_{1,N}^{\mathrm{cmp}} + \eta_{1,N}^{\mathrm{cl}}. \tag{298}$$

For $N \geq N_{\mathrm{fs}}$,

$$\inf_{X_{\mathrm{in}} \in \mathcal{C}_1^{\mathrm{init}} \cap \mathcal{C}_2^{\mathrm{init}}} \mathbb{P}_{X_{\mathrm{in}}}\{\tau_1^\sharp \leq C_{\mathrm{fs}} S_N\} \geq 1 - \eta_{\mathrm{fs},N}. \tag{299}$$

Proof. Actual first macroscopic crossing. Set $\sigma_{\mathrm{mac},1} = \inf\{t \geq 0 : q_1(X_t) \geq \varepsilon_{\mathrm{mac}}\}$, and cap selection by $s_1 = T_{E_{\mathrm{sel},1}} \wedge Q_{I,1,N}$. Before selection all estimates in Lemma B.13 hold, including the lower guard and future ratios. If $\sigma_{\mathrm{mac},1} \leq T_{E_{\mathrm{sel},1}}$, then on its good event, at that earlier crossing,

$$q_1 \geq \varepsilon_{\mathrm{mac}}, \quad |m_f| < \rho_1 \gamma_{+,1}/\sqrt{N} \leq C_{0,1}\gamma_{+,1}/\sqrt{N}, \quad m_f \geq 2\gamma_{-,1}/\sqrt{N} > \gamma_{-,1}/(C_{0,1}\sqrt{N}),$$

for the appropriate nonleader or survivor indices, and $G_{e_\ell}/G_f \geq A(c_{\mathrm{ord}})^d \Theta_1 \geq A(c_{\mathrm{ord}})^{2d}\Theta_1$. Thus this actual first crossing already belongs to $E_{\mathrm{mac},1}$ by (265). This argument also covers a tie with selection. If the macro crossing has not happened by selection, Lemma B.14 at s_1 gives its actual subsequent first crossing and its entrance set. This branch distinction uses the $\mathcal{F}_{s_1}$-measurable event $\{\sigma_{\mathrm{mac},1} \leq s_1\}$; it does not condition future noise on a future success event.

Cap the actual first crossing by $s_2 = \sigma_{\mathrm{mac},1} \wedge (Q_{I,1,N} + Q_{II,1,N})$, and apply Lemma B.15 at s_2 on the entrance event. Set $F_I = \mathcal{G}^c_{I,1,N}$. Let $B_1 \in \mathcal{F}_{s_1}$ be the event of valid capped selection with no previous macro hit, and let $B_2 \in \mathcal{F}_{s_2}$ be the event that the actual macro hit occurred within the cap with state in $E_{\mathrm{mac},1}$. Construct Regime-II and cleanup noise events $\mathcal{G}_{II}(s_1)$ and $\mathcal{G}_{III}(s_2)$ conditionally on their respective pasts. Define $F_{II} = B_1 \cap \mathcal{G}_{II}(s_1)^c$ and $F_{III} = B_2 \cap \mathcal{G}_{III}(s_2)^c$. The preceding deterministic implications show that total transition failure is contained in $F_I \cup F_{II} \cup F_{III}$. The conditional estimates give

$$\begin{aligned}
\mathbb{P}(F_I) &\le \eta^{\mathrm{cmp}}_{1,N}, \\
\mathbb{P}(F_{II}) &= \mathbb{E}[\mathbf{1}_{B_1}\mathbb{P}(\mathcal{G}_{II}(s_1)^c \mid \mathcal{F}_{s_1})] \le \eta^{\mathrm{cmp}}_{1,N}, \\
\mathbb{P}(F_{III}) &= \mathbb{E}[\mathbf{1}_{B_2}\mathbb{P}(\mathcal{G}_{III}(s_2)^c \mid \mathcal{F}_{s_2})] \le \eta^{\mathrm{cl}}_{1,N}, \\
\mathbb{P}((F_I \cup F_{II} \cup F_{III})^c) &\ge 1 - 2\eta^{\mathrm{cmp}}_{1,N} - \eta^{\mathrm{cl}}_{1,N}.
\end{aligned} \tag{300}$$

Both B_1 and B_2 concern finite-time entrance properties. The future-noise event $\mathcal{G}_{I,1,N}$ was not inserted into either conditional expectation; only its complement was separately charged. No independence between the regimes is asserted.

On the common success event put $\widehat{\tau}_1 = \sigma_{\mathrm{mac},1} + H_{III,1,N}$. At this one time, Lemma B.15 gives $\mathcal{R}^{[1]}_1$. The stronger threshold (361) gives

$$\frac{3\gamma_{2,2}}{\sqrt{N}} \le m_f(X_{\widehat{\tau}_1}) \le \frac{5\gamma_{1,2}}{8\sqrt{N}}, \quad f \in A_1, \qquad X_{\widehat{\tau}_1} \in \mathcal{C}^{\mathrm{act}}_{1,2}. \tag{301}$$

Moreover $N \ge N^{\mathrm{ord}}_1$ and (296) imply

$$G_{e_\ell}(X_{\widehat{\tau}_1})/G_f(X_{\widehat{\tau}_1}) \ge \kappa_{1,N} A (c_{\mathrm{ord}})^{2d} \Theta_1 \ge B_0^{-1}\Theta_1 = \Theta_2. \tag{302}$$

Hence

$$X_{\widehat{\tau}_1} \in \mathcal{R}^{[1]}_1 \cap \mathcal{C}^{\mathrm{act}}_{1,2} \cap \mathcal{C}^{\mathrm{gr}}_{2,2} = E^\sharp_1, \qquad \tau^\sharp_1 \le \widehat{\tau}_1. \tag{303}$$

The discrete hitting-time definition ensures $X_{\tau^\sharp_1} \in E^\sharp_1$ when it is finite; it does not identify this time with $\widehat{\tau}_1$. Finally (393) gives $S_N \ge 1$ and $H_{III,1,N} \le (A^{\mathrm{cl}}_1 + 1)S_N$, so

$$Q_{I,1,N} + Q_{II,1,N} + H_{III,1,N} \le 2(A^{\mathrm{mic}}_1 S_N + 1) + (A^{\mathrm{cl}}_1 + 1)S_N \le C_{\mathrm{fs}} S_N.$$

This and (300) prove (299). □

Role in the reconstruction. The sharp-set entrance closes the first stochastic stage. It is the base event in Theorem B.23 and supplies the first certified time used by Proposition B.22.

LATER STAGES, INDUCTION, AND THE SINGLETON STAGE

The stage-k restart state is $Y^{(k)} = X_{\tau^\sharp_{k-1}}$, not X_{in}. Retain

$$\mathcal{X}_k = (\{i^*_k\} \times (J_{k-1} \setminus \{j^*_k\})) \cup ((I_{k-1} \setminus \{i^*_k\}) \times \{j^*_k\}), \tag{304}$$

$$\mathcal{Z}_{k-1} = \bigcup_{a<k} \big[(\{i^*_a\} \times ([r] \setminus \{j^*_a\})) \cup (([r] \setminus \{i^*_a\}) \times \{j^*_a\})\big], \tag{305}$$

$$C^{\mathrm{out}}_{\chi,k} = C_{\chi,k}. \tag{306}$$

The exact restart identity

Proposition B.17 (Stage-dependent initialization sets)**.** *For $2 \le k \le r$,*

$$E^\sharp_{k-1} = \mathcal{R}^{[k-1]}_{\le k-1} \cap \mathcal{C}^{\mathrm{act}}_{1,k} \cap \mathcal{C}^{\mathrm{gr}}_{2,k}. \tag{307}$$

Proof. Substitute $s = k-1$ in (49). Its active index set is $A_{(k-1)}$, its lower and upper parameters are $\gamma_{2,(k-1)+1} = \gamma_{2,k}$ and $\gamma_{1,(k-1)+1} = \gamma_{1,k}$, and its ratio is $1 + \gamma_k/\gamma_{1,k} = \Theta_k$ by (73). These are exactly (46)–(47). At a finite discrete hitting time, $X_{\tau^\sharp_{k-1}}$ belongs to the hit set by definition; no probability is needed for this membership. □

Role in the reconstruction. This identity specifies the measurable restart invariant for Lemmas B.18 and B.19 and for the conditional stage induction in Theorem B.23.

Old-block stability and normalized drift cost

Lemma B.18 (Stability of the recovered block and its drift cost)**.** *Fix $2 \le k \le r$, $N \ge \max\{N_k^{\mathrm{old}}, N_k^{\mathrm{mic}}\}$, and a capped stopping time s with $X_s \in E_{k-1}^{\sharp}$. The two thresholds have separate roles: N_k^{old} supplies stability, whereas N_k^{mic} supplies the final smallness bound in* (314)*. Both are already assumed in Lemma B.19 and in $N_\star$. Use the absolute horizon*

$$H_{k,N}^{\mathrm{mic}} = Q_{k,N}^{\mathrm{mic}} \le (A_k^{\mathrm{mic}} + 1) S_N. \tag{308}$$

In the relabeled coordinates of (123)*, let the local active-geometry exit of Lemma B.12 be*

$$T_{\mathrm{geom}} = \inf\left\{ n \in \mathbb{Z}_{\ge 0} : \begin{array}{l} \text{some } q_a(X_{s+n}) \notin [0,1], \text{ or} \\ \|M(X_{s+n}) - \operatorname{diag} M(X_{s+n})\|_\infty > U_k^{\mathrm{mic}}/\sqrt{N} \end{array} \right\}, \quad \inf \varnothing = \infty. \tag{309}$$

There is the old-block event of Lemma B.12, with conditional failure at most $\eta_{k,N}^{\mathrm{old}}$ from (192)*, on which, simultaneously for $0 \le n \le H_{k,N}^{\mathrm{mic}} \wedge T_{\mathrm{geom}}$,*

$$m_{e_a}(X_{s+n}) \ge 1 - \varepsilon_{\mathrm{rec}} \quad (a < k), \tag{310}$$

$$\max_{f \in \mathcal{Z}_{k-1}} |m_f(X_{s+n})| \le C_{\chi,k}^{\mathrm{in}} \chi_N. \tag{311}$$

These stability bounds include the candidate exit update. The drift estimates below are used at pre-update times $0 \le n < H_{k,N}^{\mathrm{mic}} \wedge T_{\mathrm{geom}}$. For $f = (i,j) \in A_{k-1}$, write $\Lambda_f = \lambda_i \lambda_j$, $m_f = m_f(X_{s+n})$, and evaluate all drift terms at that same state. Then

$$D_f = \sqrt{N} p \lambda_i \lambda_j m_f^{p-1} - Q_f^{\mathrm{act}} - Q_f^{\mathrm{old}}, \tag{312}$$

$$|Q_f^{\mathrm{act}}| \le C_{\mathrm{act}} (U_k^{\mathrm{mic}})^{p+1} N^{-p/2}, \quad |Q_f^{\mathrm{old}}| \le C_{\mathrm{old}} (C_{\chi,k}^{\mathrm{in}})^2 \sqrt{N} \chi_N^2. \tag{313}$$

The all-active bound additionally requires $\max_{g \in A_{k-1}} |m_g| \le U_k^{\mathrm{mic}}/\sqrt{N}$; the old-term bound does not require this additional band. On that fully microscopic band the fixed-reference normalization is

$$\begin{aligned} \frac{|Q_f^{\mathrm{act}}| + |Q_f^{\mathrm{old}}|}{\sqrt{N} p \lambda_-^2 (\gamma_{2,k}/\sqrt{N})^{p-1}} &\le \frac{C_{\mathrm{act}} (U_k^{\mathrm{mic}})^{p+1}}{p \lambda_-^2 \gamma_{2,k}^{p-1} N} + \frac{C_{\mathrm{old}} (C_{\chi,k}^{\mathrm{in}})^2}{p \lambda_-^2 \gamma_{2,k}^{p-1} \log N} \\ &= \varepsilon_{k,N}^{\mathrm{rel}} \le \frac{c_{\mathrm{ord}}}{8}. \end{aligned} \tag{314}$$

The denominator in (314) *is the fixed reference in* (197)*. The conversions to a current-amplitude denominator are given in* (318)*. No pathwise lower bound $m_f \ge \gamma_{2,k}/\sqrt{N}$ is assumed. The cumulative estimate uses the stopped sum explicitly:*

$$\frac{\delta_N}{N} \sum_{0 \le n < H_{k,N}^{\mathrm{mic}} \wedge T_{\mathrm{geom}}} |Q_f^{\mathrm{old}}(X_{s+n})| \le \frac{C_{\mathrm{old}} (C_{\chi,k}^{\mathrm{in}})^2 C_\delta (A_k^{\mathrm{mic}} + 1)}{\sqrt{N} \log N}. \tag{315}$$

Proof. The parameters of Lemmas B.6–B.9 in Lemma B.12 are, explicitly,

$$\begin{aligned} &\mathcal{I} = \mathcal{I}_{k-1} = \{(a,b) : a < b,\ a < k\}, \quad \|z\| = \max_{\mathcal{I}_{k-1}} \|z_{ab}\|_{ab}, \\ &q_* = 1 - \varepsilon_{\mathrm{rec}}, \quad \kappa = \kappa_{\mathrm{old}}, \quad \varepsilon = C_{\mathrm{coup}} U_k^{\mathrm{mic}}/\sqrt{N}, \\ &B = K_\lambda b_{k,N}^{\mathrm{old}}, \quad P = K_\lambda b_N^{\mathrm{pol}}. \end{aligned} \tag{316}$$

Here $b_{k,N}^{\mathrm{old}}$ denotes the scalar noise threshold in (188), not the bold error floor in (190). The norm is the normalized norm (124). The pre-update geometry gives all $q_b \ge 0$, $q_a \ge 1 - \varepsilon_{\mathrm{rec}}$ for $a < k$, and $\|E\|_\infty \le U_k^{\mathrm{mic}}/\sqrt{N}$. Thus Lemma B.6 gives $\|I - h_N A\| \le 1 - 2\kappa_{\mathrm{old}} h_N$, and Lemma B.7 on $\mathcal{I}_{k-1}$ gives $\|r\| \le \varepsilon W_{k-1}$. The threshold (191) gives $\varepsilon \le \kappa_{\mathrm{old}}$ and $h_N L_{\mathrm{cross}} \le 1/2$, verifying Lemma B.8's deterministic conditions. Its floor is exactly $F_{k,N}^{\mathrm{old}}$ of (190).

For Lemma B.9 use $H = Q^{\mathrm{mic}}_{k,N}$, $\ell = L^{\mathrm{old}}_{k,N}$, $J = B^{\mathrm{old}}_{k,N}$, and $n_{\mathrm{sc}} = 2|\mathcal{I}_{k-1}| \le r^2$. Lemma B.12's stopped block-noise event supplies the bound B in every such block. Hence, with $C_{\mathrm{reset}} < 3$ and $W_{k-1}(0) \le K_\lambda C_{\chi,k-1}\chi_N$,

$$W_{k-1}(n) \le W_{k-1}(0) + 3F^{\mathrm{old}}_{k,N} \le K_\lambda C_{\chi,k-1}\chi_N + \boldsymbol{b}^{\mathrm{old}}_{k,N} < C^{\mathrm{in}}_{\chi,k}\chi_N.$$

The last inequality is precisely the first tube condition in (191); it contradicts any proposed first old-cross exit, including at the endpoint. Lemma B.12's strip proof uses $b_* = 1 - \varepsilon_{\mathrm{rec}}$, $c_* = 1 - \varepsilon_{\mathrm{rec}}/2$, drift at least $2\ell_{\mathrm{old}}$ on the entire strip, and the other (191) bounds. At an updated state from the strip the increment is at least

$$h_N(2\ell_{\mathrm{old}} - \ell_{\mathrm{old}}/8 - \ell_{\mathrm{old}}/8) - b^{\mathrm{pol}}_N \ge 13h_N\ell_{\mathrm{old}}/8 > 0.$$

From above c_* the total downward step is at most $c_* - b_*$, so it cannot land below b_*. This verifies (310) at the same times as (311). No block is bounded by its own initial value: the initial quantity is the maximum over $\mathcal{I}_{k-1}$.

For normalization, put $a_{\mathrm{ref}} = \gamma_{2,k}/\sqrt{N}$ and $J_{\mathrm{ref}} = \sqrt{N}p\lambda_-^2 a_{\mathrm{ref}}^{p-1} = p\lambda_-^2\gamma_{2,k}^{p-1}N^{-(p-2)/2} > 0$, solely in this calculation. Dividing the two bounds in (313) separately gives

$$\begin{aligned}
\frac{|Q^{\mathrm{act}}_f|}{J_{\mathrm{ref}}} &\le \frac{C_{\mathrm{act}}(U^{\mathrm{mic}}_k)^{p+1}N^{-p/2}}{p\lambda_-^2\gamma_{2,k}^{p-1}N^{-(p-2)/2}} = \frac{C_{\mathrm{act}}(U^{\mathrm{mic}}_k)^{p+1}}{p\lambda_-^2\gamma_{2,k}^{p-1}N}, \\
\frac{|Q^{\mathrm{old}}_f|}{J_{\mathrm{ref}}} &\le \frac{C_{\mathrm{old}}(C^{\mathrm{in}}_{\chi,k})^2N^{(p-1)/2}\chi_N^2}{p\lambda_-^2\gamma_{2,k}^{p-1}} = \frac{C_{\mathrm{old}}(C^{\mathrm{in}}_{\chi,k})^2}{p\lambda_-^2\gamma_{2,k}^{p-1}\log N}.
\end{aligned} \tag{317}$$

Here $N^{-p/2}/N^{-(p-2)/2} = N^{-1}$ and $N^{(p-1)/2}\chi_N^2 = (\log N)^{-1}$ by the definition of χ_N before (46). Adding the two lines gives (252). The defining conditions for N^{mic}_k in the microscopic estimates above, with $\mathbf{1}_{\{k\ge 2\}} = 1$, state exactly $\varepsilon^{\mathrm{rel}}_{k,N} \le c_{\mathrm{ord}}/8$. This proves (314). Lemma B.12 alone supplies the reference bound; the stated microscopic threshold supplies its last numerical inequality.

For any positive m_f on the same fully microscopic band, multiplication by the positive denominator ratios, without discarding a factor, also gives

$$\begin{aligned}
\frac{|Q^{\mathrm{act}}_f| + |Q^{\mathrm{old}}_f|}{\sqrt{N}m_f^{p-1}} &= \frac{|Q^{\mathrm{act}}_f| + |Q^{\mathrm{old}}_f|}{J_{\mathrm{ref}}}p\lambda_-^2\left(\frac{a_{\mathrm{ref}}}{m_f}\right)^{p-1} \le p\lambda_-^2\left(\frac{a_{\mathrm{ref}}}{m_f}\right)^{p-1}\varepsilon^{\mathrm{rel}}_{k,N}, \\
\frac{|Q^{\mathrm{act}}_f| + |Q^{\mathrm{old}}_f|}{\sqrt{N}p\Lambda_f m_f^{p-1}} &\le \frac{\lambda_-^2}{\Lambda_f}\left(\frac{a_{\mathrm{ref}}}{m_f}\right)^{p-1}\varepsilon^{\mathrm{rel}}_{k,N}.
\end{aligned} \tag{318}$$

If $m_f \ge a_{\mathrm{ref}}$, the second line is at most $\varepsilon^{\mathrm{rel}}_{k,N}$ because $\Lambda_f \ge \lambda_-^2$. On a relaxed positive guard, that additional state condition need not hold, so neither conversion factor is silently replaced by one. The stopped comparison uses the cumulative old forcing (315) in the fourth term of (179).

Indeed $N \ge N^{\mathrm{mic}}_k$ gives $S_N \ge 1$ and

$$\delta_N Q^{\mathrm{mic}}_{k,N} \le C_\delta(A^{\mathrm{mic}}_k + 1)N^{(p-1)/2}, \qquad \chi_N^2 = (\log N)^{-1}N^{-(p-1)/2}.$$

Each summand in (315) is evaluated before the geometry exit, and there are at most $Q^{\mathrm{mic}}_{k,N}$ such summands. Substituting the old bound of (313) gives exactly its displayed right side. This argument does not require the current leader to remain microscopic. All conclusions hold on the old-block event of Lemma B.12, with the same conditional failure (192); the normalization adds no event or probability charge. □

Role in the reconstruction. This converts old-block stability into the explicit small forcing seen by the new active rectangle. Lemma B.19 uses this budget for the next leader; Proposition B.22 uses it at stage two.

Roles of the stage estimates. Old diagonals stay recovered, old crosses stay in their stage tube, and active survivors may change multiplicatively in Regimes I–II. Only cleanup gives additive freezing by $\omega_{k,N}/\sqrt{N}$. These are different conclusions about different coordinate sets.

Later microscopic transport For $2 \le k \le r$, retain

$$E_{\mathrm{sel,k}} = \mathcal{R}^{[k]}_{\le k-1} \cap \left\{ X : \begin{array}{l} m_{e_k} \ge \gamma_{+,k}/\sqrt{N}, \\ \gamma_{-,k}/\sqrt{N} < m_f < \rho_k \gamma_{+,k}/\sqrt{N} \quad (f \in A_{k-1} \setminus \{e_k\}), \\ G_{e_\ell}/G_f \ge A(c_{\mathrm{ord}})^d \Theta_k \quad (\ell \ge k+1,\ f \in A_{\ell-1} \setminus \{e_\ell\}) \end{array} \right\}. \tag{319}$$

Define

$$\widetilde{E}^{\mathrm{in}}_{\mathrm{mac,k}} = \left\{ X : \begin{array}{l} m_{e_k} \ge \varepsilon_{\mathrm{mac}},\ m_{e_a} \ge 1 - \varepsilon_{\mathrm{rec}}\ (a < k), \\ \max_{f \in \mathcal{Z}_{k-1}} |m_f| \le C^{\mathrm{in}}_{\chi,k} \chi_N, \\ |m_f| \le L^{\mathrm{in}}_k / \sqrt{N}\ (f \in \mathcal{X}_k) \end{array} \right\}, \tag{320}$$

$$E^{\mathrm{in}}_{\mathrm{mac,k}} = \widetilde{E}^{\mathrm{in}}_{\mathrm{mac,k}} \cap \left\{ X : \begin{array}{l} 4\gamma_{2,k+1}/\sqrt{N} \le m_f \le \gamma_{1,k+1}/(2\sqrt{N})\ (f \in A_k), \\ G_{e_\ell}/G_f \ge A(c_{\mathrm{ord}})^{2d} \Theta_k\ (\ell \ge k+1) \end{array} \right\} \quad (k < r), \tag{321}$$

$$E^{\mathrm{in}}_{\mathrm{mac,r}} = \widetilde{E}^{\mathrm{in}}_{\mathrm{mac,r}}. \tag{322}$$

In the ratio condition $f \in A_{\ell-1} \setminus \{e_\ell\}$.

Lemma B.19 (Later Regimes I–II: selection and macro transport)**.** *For $2 \le k \le r$, $N \ge \max\{N^{\mathrm{mic}}_k, N^{\mathrm{old}}_k, N^{II}_k\}$ and a capped restart s_k with $X_{s_k} \in E^{\sharp}_{k-1}$, define the actual crossing*

$$\sigma_{\mathrm{mac,k}} = \inf\{t \ge s_k : m_{e_k}(X_t) \ge \varepsilon_{\mathrm{mac}}\}. \tag{323}$$

Then

$$\mathbb{P}\{\sigma_{\mathrm{mac,k}} - s_k \le A^{\mathrm{mic}}_k S_N + 1,\ X_{\sigma_{\mathrm{mac,k}}} \in E^{\mathrm{in}}_{\mathrm{mac,k}} \mid \mathcal{F}_{s_k}\} \ge 1 - \eta^{\mathrm{mic}}_{k,N}, \tag{324}$$

$$\eta^{\mathrm{mic}}_{k,N} = \eta^{\mathrm{old}}_{k,N} + 2\eta^{\mathrm{cmp}}_{k,N} \longrightarrow 0. \tag{325}$$

Proof. Construct the old event from s_k on the fixed absolute interval $[s_k, s_k + Q^{\mathrm{mic}}_{k,N}]$. Independently of any conditioning on its success, construct the stopped comparison event of Lemma B.11 for stage k on that interval. The exact substitution for Lemma B.13's calculation is

$$(A_0, e_1, \gamma_{2,1}, \gamma_{1,1}, \Theta_1, \gamma_{-,1}, \gamma_{+,1}, \rho_1, X_{\mathrm{in}}) \mapsto (A_{k-1}, e_k, \gamma_{2,k}, \gamma_{1,k}, \Theta_k, \gamma_{-,k}, \gamma_{+,k}, \rho_k, X_{s_k}).$$

The future index range becomes $\ell \ge k+1$; row-column weights remain $\Lambda_f = \lambda_i \lambda_j$. Proposition B.17 supplies the new initial amplitudes and ratios. The additive old term is the fourth term in (179), supplied by (315); hence (183) is valid with this tuple. Lemma B.13's clock algebra gives a selection within $\lceil A^{\mathrm{sel}}_k S_N \rceil$, strict nonleader guards and future ratios at least $A(c_{\mathrm{ord}})^d \Theta_k$. The old event gives $m_{e_a} \ge 1 - \varepsilon_{\mathrm{rec}}$ and old crosses at most $C^{\mathrm{in}}_{\chi,k} \chi_N \le C_{\chi,k} \chi_N$, so the selected state is precisely $E_{\mathrm{sel,k}}$. For the singleton use (84) and (215); no competitor comparison is performed.

If the actual macro crossing has occurred before or at selection, the same intermediate-state inequalities prove (320)–(321) at that crossing, exactly as in Lemma B.16: nonleaders are below $\rho_k g$, survivors above $2\gamma_{-,k}$ in units of $N^{-1/2}$, and future ratios above $A(c_{\mathrm{ord}})^d \Theta_k$. These dominate the required macro bounds by (89). If the leader is already macro at s_k, the same conclusion follows from the entry band: for $k < r$, $\gamma_{1,k} < \rho_k \gamma_{+,k}$ and $\gamma_{2,k} > \gamma_{-,k}/C_{0,k}$; at $k = r$ all nonleader conditions are empty.

Otherwise, at the capped selection restart s, use the Regime-II argument above with

$$(g, \rho, U, v_N, a^{\mathrm{II}}_N, \kappa) = (\gamma_{+,k}, \rho_k, 4\rho_k \gamma_{+,k}, C^{\mathrm{in}}_{\chi,k} \chi_N, a^{\mathrm{mic}}_{k,N}, \kappa_{II,k}).$$

The old absolute endpoint remains part of the stop. The threshold (245) verifies every leader, cross and survivor budget. The conclusion (250) is the macro band by (89); (249) changes the entrance order to $A(c_{\mathrm{ord}})^{2d} \Theta_k$. The elapsed time is covered by (255)–(256), so no old-event extension is used. Thus the actual first crossing in (323) has the asserted state.

For clarity about probability, let $B_s \in \mathcal{F}_s$ denote valid selection with no previous macro crossing. Construct a conditional Regime-II event $\mathcal{G}_{II}(s)$ whose failure on B_s is at most $\eta^{\mathrm{cmp}}_{k,N}$, without conditioning on $\mathcal{G}^{\mathrm{old}}$. The total failure is contained in

$$(\mathcal{G}^{\mathrm{old}})^c \cup (\mathcal{G}^{\mathrm{cmp}}_I)^c \cup (B_s \cap \mathcal{G}_{II}(s)^c).$$

Since $s \ge s_k$ is capped,

$$\mathbb{P}(B_s \cap \mathcal{G}_{II}(s)^c \mid \mathcal{F}_{s_k}) = \mathbb{E}[\mathbf{1}_{B_s} \mathbb{P}(\mathcal{G}_{II}(s)^c \mid \mathcal{F}_s) \mid \mathcal{F}_{s_k}] \le \eta^{\mathrm{cmp}}_{k,N}.$$

The other two conditional bounds are $\eta^{\mathrm{old}}_{k,N}$ and $\eta^{\mathrm{cmp}}_{k,N}$. Their sum is (325), proving (324). □

Role in the reconstruction. The next greedy pair reaches a macroscopic overlap while earlier pairs remain recovered and future leaders remain ordered. Proposition B.22 and Theorem B.23 pass its entrance state to Lemma B.20.

Quantitative uniform-cleanup interface We next fix the input budgets for the cleanup lemma.

Uniform cleanup with coupled norms and strip barriers Keep the normalized block norm (124) and the following constants:

$$q_\circ = \min\{1-\varepsilon_{\mathrm{rec}}, \varepsilon_{\mathrm{mac}}/2\}, \tag{326}$$

$$\kappa_{\mathrm{cl}} = \tfrac{1}{2} p c_p \lambda_-^2 q_\circ^p, \tag{327}$$

$$\ell_{\mathrm{grow}} = p\lambda_-^2 \min\{(\varepsilon_{\mathrm{mac}}/2)^{p-1}[1-(\varepsilon_{\mathrm{mac}}/2)^2], (1-\varepsilon_{\mathrm{rec}}/2)^{p-1}[1-(1-\varepsilon_{\mathrm{rec}}/2)^2]\}, \tag{328}$$

$$\ell_{\mathrm{bar}} = \tfrac{1}{2} p\lambda_-^2 (1-\varepsilon_{\mathrm{rec}})^{p-1}[1-(1-\varepsilon_{\mathrm{rec}}/2)^2], \tag{329}$$

$$A_k^{\mathrm{cl}} = 2 + \max\{4/\ell_{\mathrm{grow}}, (p+9)/(4\kappa_{\mathrm{cl}})\}, \tag{330}$$

$$H = H_{k,N}^{\mathrm{cl}} = \lceil A_k^{\mathrm{cl}} h_N^{-1} \log N \rceil, \qquad \widehat{\tau}_k = \sigma + H, \tag{331}$$

$$L_{\mathrm{cross}} = 3pK_\lambda \lambda_+^2, \quad T_{\mathrm{cl}} = 2 + L_{\mathrm{cross}}/\kappa_{\mathrm{cl}}, \quad C_{\mathrm{sur}} = K_\lambda^2 C_{\mathrm{old}}, \tag{332}$$

$$g_k^{\mathrm{band}} = \min\{\gamma_{2,k+1}, \gamma_{1,k+1}/8\} \quad (k<r), \qquad g_r^{\mathrm{band}} = 1. \tag{333}$$

The last convention is a numerical allowance only, not a stage-$(r+1)$ parameter. Retain all four kinds of cleanup error:

$$b_{\times,k,N} = \chi_N/(64K_\lambda T_{\mathrm{cl}}), \tag{334}$$

$$b_{\mathrm{sur},k,N} = \frac{\alpha_1 \delta_N}{N}\sqrt{2H\log(4r^2N^3)}, \tag{335}$$

$$b_{q,k} = \tfrac{1}{64}\min\{\varepsilon_{\mathrm{mac}}, \varepsilon_{\mathrm{rec}}\}, \tag{336}$$

$$b_{k,N}^{\mathrm{pol,act}} = \frac{\alpha_2 \delta_N^2}{2N^{3/2}}(\sqrt{r}U_k^{\mathrm{cl}} + C_G\delta_N), \tag{337}$$

$$\varepsilon_{k,N}^{\mathrm{coup,cl}} = C_{\mathrm{coup}} U_k^{\mathrm{cl}}/\sqrt{N}, \tag{338}$$

$$\boldsymbol{b}_{k,N}^{\times} = T_{\mathrm{cl}} K_\lambda b_{\times,k,N} + \frac{K_\lambda b_N^{\mathrm{pol}}}{\kappa_{\mathrm{cl}} h_N}, \tag{339}$$

$$J_{k,N}^{\times} = \frac{2K_\lambda^2 (U_k^{\mathrm{cl}})^2}{\kappa_{\mathrm{cl}} N} + 2Hh_N(\boldsymbol{b}_{k,N}^{\times})^2, \tag{340}$$

$$\omega_{k,N} = \sqrt{N}\left[(C_D + C_{\mathrm{old}})Hh_N \frac{(U_k^{\mathrm{cl}})^{p-1}}{N^{(p-1)/2}} + C_{\mathrm{sur}} J_{k,N}^{\times} + b_{\mathrm{sur},k,N} + H b_{k,N}^{\mathrm{pol,act}}\right]. \tag{341}$$

The bold symbol in (339) is the vector error floor; the scalar threshold (334) is different. Direct substitution of (114), (334) gives

$$\frac{\boldsymbol{b}_{k,N}^{\times}}{\chi_N} = \frac{1}{64} + \frac{K_\lambda \alpha_2}{2\kappa_{\mathrm{cl}}}\frac{\delta_N}{\sqrt{N}\chi_N}\sqrt{1+\alpha_2\delta_N^2/N}, \tag{342}$$

$$\frac{\delta_N}{\sqrt{N}\chi_N} = C_\delta d_{0,N}(\log N)^{-1/2} N^{-(p-3)/4}, \quad \sqrt{N}\chi_N^2 = (\log N)^{-1} N^{-(p-2)/2}, \tag{343}$$

$$\boldsymbol{b}_{k,N}^{\times} \le B_k^{\times}\chi_N, \quad B_k^{\times} = \frac{1}{64} + \frac{K_\lambda \alpha_2 C_\delta}{2\kappa_{\mathrm{cl}}}\sqrt{1+\alpha_2 C_\delta^2} \le B_\chi, \tag{344}$$

$$Hh_N \le A_k^{\mathrm{cl}} \log N + h_N. \tag{345}$$

For $N \geq 3, d_{0,N} \leq 1$, the four corresponding upper bounds for (341) are

$$\Omega_{k,N}^{(1)} = (C_D + C_{\mathrm{old}})(U_k^{\mathrm{cl}})^{p-1}(A_k^{\mathrm{cl}} \log N + h_N)N^{-(p-2)/2}, \tag{346}$$

$$\Omega_{k,N}^{(2)} = \frac{2C_{\mathrm{sur}}K_\lambda^2(U_k^{\mathrm{cl}})^2}{\kappa_{\mathrm{cl}}\sqrt{N}} + 2C_{\mathrm{sur}}(B_k^{\times})^2(A_k^{\mathrm{cl}} \log N + h_N)\sqrt{N}\chi_N^2, \tag{347}$$

$$(\Omega_{k,N}^{(3)})^2 = 2\alpha_1^2\left(A_k^{\mathrm{cl}}\frac{\delta_N}{\sqrt{N}} \log N + \frac{\delta_N^2}{N}\right)\log(4r^2N^3), \tag{348}$$

$$\Omega_{k,N}^{(4)} = \frac{\alpha_2}{2}\left(A_k^{\mathrm{cl}}\frac{\delta_N}{\sqrt{N}} \log N + \frac{\delta_N^2}{N}\right)(\sqrt{r}U_k^{\mathrm{cl}} + C_G\delta_N), \tag{349}$$

$$0 \leq \omega_{k,N} \leq \sum_{j=1}^{4} \Omega_{k,N}^{(j)} \longrightarrow 0. \tag{350}$$

Indeed (50) gives $(\delta_N/\sqrt{N}) \log N = C_\delta d_{0,N}N^{-d/2}$ and $\delta_N^2/N = C_\delta^2 d_{0,N}^2(\log N)^{-2}N^{-d}$. These identities and (343) make each term above tend to zero, including the logarithm in (348).

The threshold N_k^{cl} is the least tail threshold at least three for all of

$$d_{0,N} \leq 1, \quad L_{\mathrm{cross}}h_N \leq 1/2, \quad U_k^{\mathrm{cl}}/\sqrt{N} \leq 1, \quad C_{\chi,k}^{\mathrm{in}}\sqrt{N}\chi_N \leq L_k^{\mathrm{in}}, \tag{351}$$

$$d_N := \frac{C_{D,2}(U_k^{\mathrm{cl}})^2}{2N} \leq \min\{\ell_{\mathrm{grow}}, \ell_{\mathrm{bar}}\}/8, \tag{352}$$

$$\varepsilon_{k,N}^{\mathrm{coup,cl}} \leq \kappa_{\mathrm{cl}}, \tag{353}$$

$$Hb_N^{\mathrm{pol}} + (A_k^{\mathrm{cl}} \log N + h_N)d_N \leq b_{q,k}, \tag{354}$$

$$\tfrac{7}{8}\ell_{\mathrm{grow}}Hh_N \geq 1 + 2b_{q,k}, \tag{355}$$

$$b_N^{\mathrm{pol}} \leq h_N\ell_{\mathrm{bar}}/8, \tag{356}$$

$$h_N(C_D + \ell_{\mathrm{bar}}/8) + b_N^{\mathrm{pol}} \leq \varepsilon_{\mathrm{rec}}/4, \tag{357}$$

$$K_\lambda U_k^{\mathrm{cl}}N^{-1/2-\kappa_{\mathrm{cl}}A_k^{\mathrm{cl}}} + \boldsymbol{b}_{k,N}^{\times} \leq \tfrac{1}{4}C_{\chi,k}\chi_N, \tag{358}$$

$$K_\lambda C_{\chi,k}^{\mathrm{in}}\chi_N + \boldsymbol{b}_{k,N}^{\times} < C_{\chi,k}\chi_N, \tag{359}$$

$$K_\lambda L_k^{\mathrm{in}}/\sqrt{N} + \boldsymbol{b}_{k,N}^{\times} < L_k^{\mathrm{out}}/\sqrt{N}, \tag{360}$$

$$\omega_{k,N} \leq g_k^{\mathrm{band}}. \tag{361}$$

No threshold is enlarged. For example the first term of (358) divided by χ_N is

$$K_\lambda U_k^{\mathrm{cl}}\sqrt{\log N}\, N^{(p-3)/4-\kappa_{\mathrm{cl}}A_k^{\mathrm{cl}}} \longrightarrow 0,$$

since the exponent is at most -3. The remaining term is at most B_χ, whereas (92) gives $C_{\chi,k}/4 > B_\chi$; (359) and (360) have fixed strict reserves from (92) and (88). All other limits follow from (50) and (350).

Set the failure allowance

$$\begin{aligned}\eta_{k,N}^{\mathrm{cl}} = He^{-c_R N} + 4r^2\exp\left(-\frac{b_{\times,k,N}^2N^2}{2\alpha_1^2\delta_N^2H}\right) + 2r^2N^{-3} \\ + 2\exp\left(-\frac{b_{q,k}^2N^2}{2\alpha_1^2\delta_N^2H}\right) + 2krH\exp\left(-\frac{\ell_{\mathrm{bar}}^2N}{128\alpha_1^2}\right).\end{aligned} \tag{362}$$

Lemma B.20 (Uniform cleanup with coupled norms and strip barriers)**.** *For $1 \le k \le r$, $N \ge N_k^{\mathrm{cl}}$, and a capped stopping time σ with $X_\sigma \in E_{\mathrm{mac,k}}^{\mathrm{in}}$, there is an event $\mathcal{G}_{k,N}^{\mathrm{cl}}$ satisfying*

$$\mathbb{P}((\mathcal{G}_{k,N}^{\mathrm{cl}})^c \mid \mathcal{F}_\sigma) \le \eta_{k,N}^{\mathrm{cl}}, \tag{363}$$

$$m_{e_a}(X_{\sigma+n}) \ge 1 - \varepsilon_{\mathrm{rec}} \quad (a < k,\ 0 \le n \le H), \tag{364}$$

$$\max_{f \in \mathcal{Z}_{k-1}} |m_f(X_{\sigma+n})| \le C_{\chi,k}\chi_N \quad (0 \le n \le H), \tag{365}$$

$$m_{e_k}(X_{\widehat{\tau}_k}) \ge 1 - \varepsilon_{\mathrm{rec}}, \tag{366}$$

$$\max_{f \in \mathcal{X}_k} |m_f(X_{\widehat{\tau}_k})| \le C_{\chi,k}\chi_N, \tag{367}$$

$$\max_{f \in A_k} \max_{n \le H} |m_f(X_{\sigma+n}) - m_f(X_\sigma)| \le \omega_{k,N}/\sqrt{N}. \tag{368}$$

The last four displays are on $\mathcal{G}_{k,N}^{\mathrm{cl}}$; an empty maximum is zero.

Proof. Step 1. Stop, event, and pre-update geometry. Write $M_{ab}(n) = m_{i_a^*,j_b^*}(X_{\sigma+n})$ and $q_a = M_{aa}$. Let $\mathcal{V}$ be the state set specified by

$$q_k \ge \varepsilon_{\mathrm{mac}}/2, \quad q_a \ge 1 - \varepsilon_{\mathrm{rec}}\ (a < k), \quad \max_{\mathcal{Z}_{k-1}} |m_f| \le C_{\chi,k}\chi_N, \quad \max_{\mathcal{X}_k} |m_f| \le L_k^{\mathrm{out}}/\sqrt{N},$$

and, if $k < r$, $m_f \in [2\gamma_{2,k+1}, 3\gamma_{1,k+1}/4]/\sqrt{N}$ for every $f \in A_k$. Define

$$T_{\mathrm{bad}} = \inf\{n \in \mathbb{Z}_{\ge 1} : X_{\sigma+n} \notin \mathcal{V}\}, \qquad T = T_{\mathrm{bad}} \wedge H, \qquad I_n = \mathbf{1}_{\{n<T\}}. \tag{369}$$

The entrance and (359)–(360) give $X_\sigma \in \mathcal{V}$. Define five events: the Gram event on H updates; all stopped off-diagonal partial noise sums at most $b_{\times,k,N}$; all survivor partial noise sums at most $b_{\mathrm{sur},k,N}$; the new diagonal partial noise sum at most $b_{q,k}$; and each individual selected diagonal noise increment satisfying $|\xi_{aa}(\sigma+n+1)| \le h_N \ell_{\mathrm{bar}}/8$ for $a \le k, n < H$. Their intersection is $\mathcal{G}_{k,N}^{\mathrm{cl}}$. Empty coordinate collections impose no tests.

At $n < T$, old, new and survivor off-diagonal coordinates satisfy

$$\|E(n)\|_\infty \le \max\{C_{\chi,k}\chi_N, L_k^{\mathrm{out}}/\sqrt{N}, \mathbf{1}_{\{k<r\}} 3\gamma_{1,k+1}/(4\sqrt{N})\} \le U_k^{\mathrm{cl}}/\sqrt{N}. \tag{370}$$

Here $\sqrt{N}\chi_N \le 1$ for $N \ge 3, p \ge 3$ and (93) fixes U_k^{cl}. All diagonal entries are nonnegative by the same state partition, and all are at most one by (107).

Step 2. Two applications of the coupled matrix lemma. For $s \in \{k-1, k\}$ let $\mathcal{I}_s = \{(a,b) : a < b, a \le s\}$ and $W_s(n) = \max_{\mathcal{I}_s} \|Z_{ab}(n)\|_{ab}$. When a set is empty its maximum is zero. For each nonempty set, apply Lemma B.8 with

$$z_n = (Z_{ab}(n))_{(a,b)\in\mathcal{I}_s}, \quad A_n = \mathrm{diag}_{\mathcal{I}_s} A_{ab}(n), \quad h = h_N, \quad \kappa = \kappa_{\mathrm{cl}},$$
$$\varepsilon = \varepsilon_{k,N}^{\mathrm{coup,cl}}, \quad B = K_\lambda b_{\times,k,N}, \quad P = K_\lambda b_N^{\mathrm{pol}}, \quad r_n = (R_{ab}(n)/\sqrt{N})_{\mathcal{I}_s}, \quad H_{\mathrm{application}} = T. \tag{371}$$

The pre-update lower bound $q_a \ge q_\circ$ for $a \le s$ and Lemma B.6 give block contraction at least $pc_p\lambda_-^2 q_\circ^p = 2\kappa_{\mathrm{cl}}$ and operator norm at most L_{cross}. Lemma B.7 applied to this same $\mathcal{I}_s$ gives $\|r_n\| \le \varepsilon W_s(n)$. Conditions (351), (353) give the step-size and coupling hypotheses of Lemma B.8. Also $2\kappa_{\mathrm{cl}} \le L_{\mathrm{cross}}$, so $0 < 2\kappa_{\mathrm{cl}} h_N \le 1$. On the common off-diagonal noise event, (125) gives the cumulative vector bound B simultaneously for both collections. The Gram event and (114) give P for both. Thus their error floor is exactly (339), and Lemma B.8 gives

$$W_s(n) \le e^{-\kappa_{\mathrm{cl}} h_N n} W_s(0) + \boldsymbol{b}_{k,N}^{\times} \quad (s = k-1, k;\ 0 \le n \le T). \tag{372}$$

This is one noise event supporting two deterministic applications. No second cross-noise charge is made.

At entrance $W_{k-1}(0) \le K_\lambda C_{\chi,k}^{\mathrm{in}}\chi_N$, whereas

$$W_k(0) \le K_\lambda \max\{C_{\chi,k}^{\mathrm{in}}\chi_N, L_k^{\mathrm{in}}/\sqrt{N}\} \le K_\lambda L_k^{\mathrm{in}}/\sqrt{N}$$

by (351). Hence at a proposed old-cross exit, (372), (359) and coordinate domination (125) give $|m_f| < C_{\chi,k}\chi_N$, a contradiction. At a proposed new-cross exit, use (360) instead to obtain $|m_f| <$

$L_k^{\mathrm{out}}/\sqrt{N}$. At $n = H$, (358) and $W_k(0) \le K_\lambda U_k^{\mathrm{cl}}/\sqrt{N}$ give $W_k(H) \le C_{\chi,k}\chi_N/4$. All these conclusions concern maxima over sets; none replaces $W_s(0)$ by an individual block initial norm.

Use the energy conclusion (146) of Lemma B.8 with $s = k$ (no new lemma or event):

$$h_N \sum_{n<T} W_k(n)^2 \le \frac{2W_k(0)^2}{\kappa_{\mathrm{cl}}} + 2Hh_N(\boldsymbol{b}_{k,N}^{\times})^2 \le J_{k,N}^{\times}. \tag{373}$$

Any product of two selected cross coordinates, from the same or different blocks, is bounded by W_k^2.

Step 3. Survivor displacement and the exit update. For a survivor $i, j > k$, put $\mu = U_k^{\mathrm{cl}}/\sqrt{N} \le 1$. In the two monomials of (131), the disjoint index cases and selected cross factors are

$u > k, v > k$	all factors active, bound μ^{p+1}
$u \le k, v > k$	M_{uj}, M_{uv}
$u > k, v \le k$	M_{iv}, M_{uv}
$u \le k, v \le k$	M_{iv}, M_{uj}

Every listed factor occurs in both branches with positive exponent; every remaining factor is at most one in modulus. Thus the selected part has bound $pr^2\lambda_+^2 W_k^2 \le C_{\mathrm{sur}} W_k^2$, and the fully active part, together with the direct monomial, is at most $(C_D + C_{\mathrm{old}})\mu^{p-1}$. The whole survivor row has norm at most $\sqrt{r}U_k^{\mathrm{cl}}/\sqrt{N}$; multiplication of its Euler row bound by $\|B - I\|_{\mathrm{op}}$ gives exactly (337). Sum the exact recursion and use (373) and the survivor noise event:

$$\begin{aligned} |m_f(X_{\sigma+n}) - m_f(X_\sigma)| &\le (C_D + C_{\mathrm{old}})Hh_N\mu^{p-1} + C_{\mathrm{sur}}J_{k,N}^{\times} + b_{\mathrm{sur},k,N} + Hb_{k,N}^{\mathrm{pol,act}} \\ &= \omega_{k,N}/\sqrt{N} \qquad (n \le T). \end{aligned} \tag{374}$$

With (361) and the macro entrance, the updated survivor lies in

$$[3\gamma_{2,k+1}, 5\gamma_{1,k+1}/8]/\sqrt{N} \subset (2\gamma_{2,k+1}, 3\gamma_{1,k+1}/4)/\sqrt{N}.$$

This strict inclusion contradicts a first survivor exit, including at $T = H$.

Step 4. Entire-strip recovery and one-step crossing control. Apply the diagonal conclusion (130) of Lemma B.6 with $e \le U_k^{\mathrm{cl}}/\sqrt{N}$, giving

$$D_{aa}/\sqrt{N} = pL_a q_a^{p-1}(1 - q_a^2) + e_a, \quad |e_a| \le d_N \quad (a \le k, n < T). \tag{375}$$

For every $q \in [1 - \varepsilon_{\mathrm{rec}}, 1 - \varepsilon_{\mathrm{rec}}/2]$, the logistic term is at least $2\ell_{\mathrm{bar}}$. Using (352), the single-step noise event and (356),

$$q_a(n+1) - q_a(n) \ge h_N(2\ell_{\mathrm{bar}} - \ell_{\mathrm{bar}}/8 - \ell_{\mathrm{bar}}/8) - h_N\ell_{\mathrm{bar}}/8 = 13h_N\ell_{\mathrm{bar}}/8 > 0. \tag{376}$$

If $q_a(n) > 1 - \varepsilon_{\mathrm{rec}}/2$, (115), the same noise event and (357) give

$$q_a(n+1) \ge q_a(n) - [h_N(C_D + \ell_{\mathrm{bar}}/8) + b_N^{\mathrm{pol}}] > 1 - 3\varepsilon_{\mathrm{rec}}/4 > 1 - \varepsilon_{\mathrm{rec}}.$$

These two cases prove preservation above $1 - \varepsilon_{\mathrm{rec}}$ by induction, including a start exactly on that boundary. They apply to every old diagonal and to the new one after it reaches $1 - \varepsilon_{\mathrm{rec}}/2$.

Before the new leader reaches $1-\varepsilon_{\mathrm{rec}}/2$, its pre-update value is in $[\varepsilon_{\mathrm{mac}}/2, 1-\varepsilon_{\mathrm{rec}}/2]$. The derivative of $q^{p-1}(1-q^2)$ is $q^{p-2}[(p-1) - (p+1)q^2]$, so the sole interior critical point is a maximum. Its minimum on this interval is at an endpoint; (328), (352) give effective drift at least $7\ell_{\mathrm{grow}}/8$. Through a candidate lower exit,

$$q_k(n) \ge q_k(0) + \tfrac{7}{8}\ell_{\mathrm{grow}}h_N n - b_{q,k} - Hb_N^{\mathrm{pol}} \ge \varepsilon_{\mathrm{mac}} - 2b_{q,k} > \varepsilon_{\mathrm{mac}}/2$$

by (336), (354). Thus the lower exit is impossible. Once all state exits are excluded, if no upper hit occurs by H, (355) gives

$$q_k(H) - q_k(0) \ge \tfrac{7}{8}\ell_{\mathrm{grow}}Hh_N - b_{q,k} - Hb_N^{\mathrm{pol}} \ge 1. \tag{377}$$

This contradicts $q_k(0) \ge \varepsilon_{\mathrm{mac}} > 0$ and $q_k(H) \le 1$. If the leader starts above the upper target, this hitting time is zero and the strip argument alone applies.

Step 5. Closed stop and conditional probability. Every bound above uses states $n < T$ and their next updates; hence it holds at a candidate $T_{\mathrm{bad}} \le H$. Steps 2–4 contradict each defining failure of $\mathcal{V}$ individually. They also contradict any combination of simultaneous failures. Therefore $T_{\mathrm{bad}} > H$ (or is infinite), $T = H$, the new leader hits its upper target by H, and the endpoint cross bound and (374) give (364)–(368).

Conditionally on $\mathcal{F}_\sigma$, Lemma B.3 and (101) bound the five event failures by

$$He^{-c_R N}, \quad 2r^2 e^{-b_{\times,k,N}^2 N^2/(2\alpha_1^2\delta_N^2 H)}, \quad 2r^2 e^{-\log(4r^2N^3)}, \quad 2e^{-b_{q,k}^2 N^2/(2\alpha_1^2\delta_N^2 H)},$$

and $2kH\exp[-\ell_{\mathrm{bar}}^2 N/(128\alpha_1^2)]$, respectively. The last exponent follows from

$$\frac{(h_N \ell_{\mathrm{bar}}/8)^2}{2\alpha_1^2\delta_N^2/N^2} = \frac{\ell_{\mathrm{bar}}^2 N}{128\alpha_1^2}.$$

Each of these terms is at most its allocated term in (362); their union gives (363). This is conditional on the past at the capped restart only. Uniformity follows because every bound depends only on the displayed entrance constants. □

Role in the reconstruction. This cleanup contracts new cross overlaps, keeps old ones stable, and freezes surviving microscopic overlaps. Lemma B.15 invokes its first-stage case; Lemma B.21 uses its general output to certify the next sharp set.

Lemma B.21 (Regime III to the exact sharp set)**.** *For $X_\sigma \in E_{\mathrm{mac,k}}^{\mathrm{in}}$ and $N \ge \max\{N_k^{\mathrm{cl}}, N_k^{\mathrm{ord}}\}$, on the event of Lemma B.20,*

$$X_{\widehat{\tau}_k} \in E_k^\sharp, \qquad \tau_k^\sharp \le \widehat{\tau}_k \quad \textit{when } \sigma \ge \tau_{k-1}^\sharp. \tag{378}$$

Proof. At $\widehat{\tau}_k$, (364)–(367) give all selected diagonal and cross conditions. The cross coordinates for a new row or column split between $\mathcal{Z}_{k-1}$ (crosses to old positions) and $\mathcal{X}_k$ (crosses to the remaining active positions). Hence these outputs give the whole $\mathcal{R}_{\le k}^{[k]}$, not merely its active subrectangle. If $k = r$, this is $E_r^\sharp$ by (49) and the proof of set membership is finished.

For $k < r$, use (368), (361) on the entrance band to get

$$\sqrt{N} m_f(X_{\widehat{\tau}_k}) \ge 4\gamma_{2,k+1} - \omega_{k,N} \ge 3\gamma_{2,k+1} \ge \gamma_{2,k+1},$$
$$\sqrt{N} m_f(X_{\widehat{\tau}_k}) \le \gamma_{1,k+1}/2 + \omega_{k,N} \le 5\gamma_{1,k+1}/8 \le \gamma_{1,k+1}.$$

This is $\mathcal{C}_{1,k+1}^{\mathrm{act}}$ at the same endpoint. The application (257)–(259), now with the present $\omega_{k,N}$ and macro entrance ratios $A(c_{\mathrm{ord}})^{2d}\Theta_k$, gives every future ratio at least Θ_{k+1}. Therefore this endpoint also lies in $\mathcal{C}_{2,k+1}^{\mathrm{gr}}$. Their intersection with $\mathcal{R}_{\le k}^{[k]}$ is exactly $E_k^\sharp$ by (49). If the stage entrance precedes σ, the constructed endpoint is an admissible time in the infimum (49), so $\tau_k^\sharp \le \widehat{\tau}_k$. □

Role in the reconstruction. The separate leader, old-block, and survivor estimates now hold at one common endpoint in $E_k^\sharp$. Proposition B.22 and Theorem B.23 use that endpoint for the next restart.

Stage two written out

Proposition B.22 (Recovery of the second spike)**.** *Assume*

$$N \ge \max\{N_2^{\mathrm{mic}}, N_2^{II}, N_2^{\mathrm{old}}, N_2^{\mathrm{cl}}, N_2^{\mathrm{ord}}, N_2^{\mathrm{ceil}}\}.$$

On the first-stage success event,

$$Y^{(2)} = X_{\tau_1^\sharp} \in \mathcal{R}_1^{[1]} \cap \mathcal{C}_{1,2}^{\mathrm{act}} \cap \mathcal{C}_{2,2}^{\mathrm{gr}}, \tag{379}$$

$$C_2 = 2A_2^{\mathrm{mic}} + A_2^{\mathrm{cl}} + 3, \qquad \eta_{2,N} = \eta_{2,N}^{\mathrm{old}} + 2\eta_{2,N}^{\mathrm{cmp}} + \eta_{2,N}^{\mathrm{cl}}, \tag{380}$$

$$\mathbb{P}\{\tau_2^\sharp - \tau_1^\sharp \le C_2 S_N \mid \mathcal{F}_{\tau_1^\sharp}\} \ge 1 - \eta_{2,N}, \tag{381}$$

$$X_{\tau_2^\sharp} \in \mathcal{R}_1^{[2]} \cap \mathcal{R}_2^{[2]} \cap \mathcal{C}_{1,3}^{\mathrm{act}} \cap \mathcal{C}_{2,3}^{\mathrm{gr}} = E_2^\sharp. \tag{382}$$

For $r = 2$, the last two factors in (382) *impose no condition and no third-stage parameters are used. Membership in* (382) *is asserted when the certified time is finite.*

Proof. Set the deterministic cap $D_1 = \lceil C_{\mathrm{fs}} S_N \rceil$ and $s = \tau_1^\sharp \wedge D_1$. Let $B_1 = \{\tau_1^\sharp \le D_1,\ X_{\tau_1^\sharp} \in E_1^\sharp\} \in \mathcal{F}_s$. On B_1, $s = \tau_1^\sharp$ and (379) follows from Proposition B.17. The active set is $A_1 = I_1 \times J_1$, the old set is $\mathcal{Z}_1$, and the new leader is e_2. Its five-parameter tuple is exactly

$$(\gamma_{+,2}, \gamma_{-,2}, \gamma_2, \gamma_{1,2}, \gamma_{2,2}). \tag{383}$$

Apply Lemma B.12 through Lemma B.18 with $k = 2$, $H = Q_{2,N}^{\mathrm{mic}}$ and the block set $\mathcal{I}_1$. It gives $m_{e_1} \ge 1 - \varepsilon_{\mathrm{rec}}$, old crosses at most $C_{\chi,2}^{\mathrm{in}} \chi_N$, and cumulative old forcing (315) with $k = 2$. The first spike remains in the full Stiefel update. Its forcing is included in $b_{2,N}^{\mathrm{mic}}$ and, in Regime II, in $v_N = C_{\chi,2}^{\mathrm{in}} \chi_N$; the scalar comparison uses the accumulated bound (315), while (314) describes the fixed-reference cost on the fully microscopic band.

Lemma B.19 with this exact tuple gives the actual macro crossing σ in $[s, s + Q_{2,N}^{\mathrm{mic}}]$ and the state $E_{\mathrm{mac},2}^{\mathrm{in}}$, except for conditional failure $\eta_{2,N}^{\mathrm{mic}}$. Set $t = \sigma \wedge (s + Q_{2,N}^{\mathrm{mic}})$ and $B_2 = \{\sigma \le s + Q_{2,N}^{\mathrm{mic}}, X_\sigma \in E_{\mathrm{mac},2}^{\mathrm{in}}\} \in \mathcal{F}_t$. At t, condition only on $\mathcal{F}_t$ and apply Lemma B.20 on this entrance event. With $k = 2$, it controls the old collection $\mathcal{I}_1$ and all selected blocks $\mathcal{I}_2$, uses U_2^{cl}, $H_{2,N}^{\mathrm{cl}}$, and floor $\boldsymbol{b}_{2,N}^{\times}$, and gives the five conclusions (364)–(368). Thus e_1 stays recovered, e_2 recovers, all their crosses enter the stage-2 tube, and A_2 is frozen by $\omega_{2,N}/\sqrt{N}$ when $r > 2$.

The localized cleanup failure satisfies

$$\mathbb{P}(B_2 \cap (\mathcal{G}_{2,N}^{\mathrm{cl}}(t))^c \mid \mathcal{F}_s) = \mathbb{E}[\mathbf{1}_{B_2} \mathbb{P}((\mathcal{G}_{2,N}^{\mathrm{cl}}(t))^c \mid \mathcal{F}_t) \mid \mathcal{F}_s] \le \eta_{2,N}^{\mathrm{cl}}.$$

Together with Lemma B.19 this gives failure at most $\eta_{2,N}$ on B_1, uniformly over its state. Lemma B.21 puts the common constructed endpoint $\widehat{\tau}_2 = \sigma + H_{2,N}^{\mathrm{cl}}$ in $E_2^\sharp$. Therefore $\tau_2^\sharp \le \widehat{\tau}_2$ and finite hitting implies (382). Finally (393) and the ceiling inequality give

$$Q_{2,N}^{\mathrm{mic}} + H_{2,N}^{\mathrm{cl}} \le (A_2^{\mathrm{mic}} + 1) S_N + (A_2^{\mathrm{cl}} + 1) S_N \le C_2 S_N, \quad C_2 - (A_2^{\mathrm{mic}} + A_2^{\mathrm{cl}} + 2) = A_2^{\mathrm{mic}} + 1 > 0. \tag{384}$$

On B_1, $\mathcal{F}_s$ agrees with the stopped past $\mathcal{F}_{\tau_1^\sharp}$, proving (381). No conditioning on the first stage's future success event is used to generate new noise. □

Role in the reconstruction. The $k = 2$ calculation shows how to restart after a random certified time without conditioning on future noise. Theorem B.23 repeats this conditional-expectation step for every later stage.

Induction through the penultimate stage

Theorem B.23 (Stage-to-stage induction)**.** *For $N \ge N_\star$ as in* (392)*, retain*

$$C_k = 2A_k^{\mathrm{mic}} + A_k^{\mathrm{cl}} + 3, \quad \eta_{k,N} = \eta_{k,N}^{\mathrm{old}} + 2\eta_{k,N}^{\mathrm{cmp}} + \eta_{k,N}^{\mathrm{cl}} \quad (2 \le k \le r), \tag{385}$$

$$\inf_{X \in E_{k-1}^\sharp} \mathbb{P}_X\{T_{E_k^\sharp} \le C_k S_N\} \ge 1 - \eta_{k,N}. \tag{386}$$

There is an event G_N with $\mathbb{P}(G_N) \ge 1 - \eta_{\mathrm{fs},N} - \sum_{k=2}^r \eta_{k,N}$ on which

$$X_{\tau_k^\sharp} \in E_k^\sharp, \qquad 0 \le \tau_k^\sharp - \tau_{k-1}^\sharp \le C_k S_N \quad (1 \le k \le r). \tag{387}$$

Here and only as an alias in (387)*, $C_1 := C_{\mathrm{fs}}$, and $\eta_{1,N} := \eta_{\mathrm{fs},N}$.*

Proof. For a deterministic $X \in E_{k-1}^\sharp$, apply Lemma B.19 at zero, cap its actual macro time by $Q_{k,N}^{\mathrm{mic}}$, and apply Lemma B.20 on the measurable macro-entrance event. The exact conditional-expectation calculation of Proposition B.22, with 2 replaced by k, gives failure at most $\eta_{k,N}$. The parameters are $A_{k-1}, e_k, \mathcal{Z}_{k-1}, \mathcal{X}_k, A_k$, the stage-$k$ five-gamma tuple, and times $Q_{k,N}^{\mathrm{mic}}, H_{k,N}^{\mathrm{cl}}$. Lemma B.21 gives the common endpoint in $E_k^\sharp$ and

$$Q_{k,N}^{\mathrm{mic}} + H_{k,N}^{\mathrm{cl}} \le (A_k^{\mathrm{mic}} + A_k^{\mathrm{cl}} + 2) S_N \le C_k S_N. \tag{388}$$

Every constant is uniform over the starting state, giving (386).

To make the random induction explicit, set $D_0 = 0$ and $D_j = \sum_{a=1}^{j} \lceil C_a S_N \rceil$, and define

$$B_j = \bigcap_{a=1}^{j} \{\tau_a^\sharp < \infty,\ X_{\tau_a^\sharp} \in E_a^\sharp,\ \tau_a^\sharp - \tau_{a-1}^\sharp \le C_a S_N\}.$$

These assertions concern past times at $\tau_j^\sharp$. On B_{j-1} we have $\tau_{j-1}^\sharp \le D_{j-1}$. Use the capped restart $s_j = \tau_{j-1}^\sharp \wedge D_{j-1}$; B_{j-1} is then an $\mathcal{F}_{s_j}$-measurable event and $s_j = \tau_{j-1}^\sharp$ on it. By (307), the induction invariant at this state is exactly

$$\mathcal{R}_{\le j-1}^{[j-1]} \cap \mathcal{C}_{1,j}^{\mathrm{act}} \cap \mathcal{C}_{2,j}^{\mathrm{gr}}.$$

The old tube becomes $\mathcal{R}_{\le j-1}^{[j]}$, the active rectangle partitions disjointly into $\{e_j\} \cup \mathcal{X}_j \cup A_j$, and Lemma B.21 gives the next invariant at the common endpoint. Apply the uniform transition just proved at s_j. Then

$$\begin{aligned} \mathbb{P}(B_{j-1} \setminus B_j) &= \mathbb{E}\left[\mathbf{1}_{B_{j-1}} \mathbb{P}(B_j^c \mid \mathcal{F}_{s_j})\right] \le \eta_{j,N} \mathbb{P}(B_{j-1}) \le \eta_{j,N}, \\ \mathbb{P}(B_j^c) &\le \mathbb{P}(B_{j-1}^c) + \eta_{j,N}. \end{aligned} \tag{389}$$

The base is Lemma B.16, $\mathbb{P}(B_1^c) \le \eta_{\mathrm{fs},N}$. Induction gives $\mathbb{P}(B_r^c) \le \eta_{\mathrm{fs},N} + \sum_{j=2}^{r} \eta_{j,N}$. Take $G_N = B_r$, proving (387). No conditioning on B_r is used to claim freshness at an earlier restart. □

Role in the reconstruction. This induction provides the uniform stage events and their time bounds. Corollary B.24 handles the final singleton and the concluding proof of Theorem B.1 identifies its hitting times with τ_k^B.

The final singleton stage

Corollary B.24 (Final-stage extension and recovery of all spikes)**.** *Under Theorem B.23, $A_{r-1} = \{e_r\}$ and $A_r = \varnothing$. On G_N,*

$$X_{\tau_r^\sharp} \in \bigcap_{a=1}^{r} \mathcal{R}_a^{[r]}(\varepsilon_{\mathrm{rec}}) = E_r^\sharp, \qquad \|X_{\tau_r^\sharp} - V P_\pi\|_F^2 \le 2r\varepsilon_{\mathrm{rec}}, \quad \pi(j_k^*) = i_k^*. \tag{390}$$

Proof. Exactly one undeleted row and one undeleted column remain. In the stage-r Regime-I calculation, the competitor minimum is over an empty set and equals infinity. The sole threshold is (84); the lower guard and integer pole margin (215), including the c_{ord} refinement (276), suffice. Regime II has $u_n = 0$, no survivor stopping time and no ordering application. Lemma B.20 at $k = r$ uses $\mathcal{I}_{r-1}$ and $\mathcal{I}_r$; these sets in fact coincide, since $a < b \le r$ forces $a \le r - 1$. No additional noise charge is needed for that coincidence. Also $\mathcal{X}_r = A_r = \varnothing$; all crosses of the new principal row or column to old positions already belong to $\mathcal{Z}_{r-1}$. Old stability controls $m_{i_r^*, j_a^*}$ and $m_{i_a^*, j_r^*}$ for each $a < r$, and new leader recovery gives $m_{e_r} \ge 1 - \varepsilon_{\mathrm{rec}}$ at the common endpoint. Thus all factors of $E_r^\sharp$ hold together. The certified time belongs to that set by its definition.

Define P_π by requiring column j_k^* of VP_π to be $v_{i_k^*}$. Distinct deleted labels define a permutation. Since both frames have orthonormal columns,

$$\|X - VP_\pi\|_F^2 = \mathrm{tr}(X^T X) + \mathrm{tr}(P_\pi^T V^T V P_\pi) - 2\,\mathrm{tr}(X^T V P_\pi) = 2r - 2\sum_{k=1}^{r} m_{e_k}(X).$$

At $X = X_{\tau_r^\sharp}$, each summand is at least $1 - \varepsilon_{\mathrm{rec}}$, proving the stated $2r\varepsilon_{\mathrm{rec}}$ bound. No γ_{r+1}, $\gamma_{1,r+1}$, $\gamma_{2,r+1}$ or Θ_{r+1} is introduced. □

Role in the reconstruction. The last active pair needs no competitor comparison. Its diagonal overlap sum gives the Frobenius recovery guarantee in Theorem B.1.

Raw threshold and certified recovery times

Proposition B.25 (Exact time alignment)**.** *On stage-k success, let $s_k = \tau^\sharp_{k-1}$, let $\sigma_{\mathrm{mac,k}}$ be the actual first crossing since s_k as constructed in Lemmas B.16 and B.19, and let $\widehat{\tau}_k = \sigma_{\mathrm{mac,k}} + H^{\mathrm{cl}}_{k,N}$. Then*

$$\sigma_{\mathrm{mac,k}} \le \beta^{\mathrm{rec}}_k \le \tau^\sharp_k \le \widehat{\tau}_k, \qquad 0 \le \tau^\sharp_k - \beta^{\mathrm{rec}}_k \le H^{\mathrm{cl}}_{k,N}. \tag{391}$$

Proof. By (77), $\varepsilon_{\mathrm{mac}} \le (1-\varepsilon_{\mathrm{rec}})/4 < 1-\varepsilon_{\mathrm{rec}}$. At every $s_k \le t < \sigma_{\mathrm{mac,k}}$ the definition of the first crossing gives $m_{e_k}(X_t) < \varepsilon_{\mathrm{mac}} < 1-\varepsilon_{\mathrm{rec}}$, so no raw recovery hit has occurred. This proves $\sigma_{\mathrm{mac,k}} \le \beta^{\mathrm{rec}}_k$; equality is allowed if one update overshoots both levels. Every state of $E^\sharp_k$ has $m_{e_k} \ge 1-\varepsilon_{\mathrm{rec}}$, so the infimum defining β^{rec}_k is at most the infimum defining $\tau^\sharp_k$. Lemma B.16 for $k=1$, or Lemma B.21 for $k \ge 2$, gives $X_{\widehat{\tau}_k} \in E^\sharp_k$, hence $\tau^\sharp_k \le \widehat{\tau}_k$. Subtracting $\beta^{\mathrm{rec}}_k \ge \sigma_{\mathrm{mac,k}}$ from this last inequality yields $\tau^\sharp_k - \beta^{\mathrm{rec}}_k \le \widehat{\tau}_k - \sigma_{\mathrm{mac,k}} = H^{\mathrm{cl}}_{k,N}$. These times are ordered, not identified. □

Role in the reconstruction. This distinguishes the observable-level crossing, the raw leader threshold, and certified recovery. It explains the cleanup delay between the leader becoming large and the stage being safe for the next restart.

COEFFICIENT LEDGER AND PROOF STATUS

The constant selection order is: primitive model parameters; concentration, polynomial and norm constants; $\bar{c}, B_0, \Theta_k, c_{\mathrm{ord}}$; $\varepsilon_{\mathrm{mac}}$; the stage-by-stage ρ_k and five-gamma recursion; cross radii and microscopic/cleanup scales; time coefficients; finite tail thresholds; and total time and failure allowances. Every constant is specified by the definitions and dictionary below. In particular C_δ is input, and no unnamed constant is enlarged in this reconstruction.

$$N_\star = \max_{1\le k\le r}\{N^{\mathrm{mic}}_k, N^{\mathrm{race}}_k, N^{II}_k, N^{\mathrm{cl}}_k, N^{\mathrm{old}}_k, N^{\mathrm{ord}}_k, N^{\mathrm{ceil}}_k\}, \quad N^{\mathrm{old}}_1 = 1, \tag{392}$$

$$N^{\mathrm{ceil}}_k = \min\{n : S_N \ge 1,\ H^{\mathrm{cl}}_{k,N} \le (A^{\mathrm{cl}}_k + 1)S_N \text{ for every } N \ge n\}, \tag{393}$$

$$C_{\mathrm{all}} = C_{\mathrm{fs}} + \sum_{k=2}^{r} C_k, \qquad \eta_{\mathrm{all},N} = \eta_{\mathrm{fs},N} + \sum_{k=2}^{r} \eta_{k,N} \longrightarrow 0. \tag{394}$$

The step-size sequence in (50) is fixed. Each minimum is a tail threshold for that sequence, not a computable convergence modulus from finitely many observations of an arbitrary sequence. In particular

$$\frac{H^{\mathrm{cl}}_{k,N}}{S_N} \le \frac{A^{\mathrm{cl}}_k}{C_\delta}(\log N)N^{-d/2} + S_N^{-1} \longrightarrow 0,$$

which verifies the finiteness needed in (393). The integer comparison (276) uses the factor-two reserve in the race thresholds. (314) uses the existing N^{mic}_k smallness condition. Lemma B.18 also uses N^{old}_k; both already occur in (392), so the same threshold $N_\star$ suffices.

From Theorem B.23 and the definitions of G_N,

$$G_N \subseteq \left\{\tau^\sharp_r = \sum_{k=1}^{r}(\tau^\sharp_k - \tau^\sharp_{k-1}) \le C_{\mathrm{all}}S_N\right\},$$

and therefore the uniform total-time conclusion is

$$\inf_{X_{\mathrm{in}} \in \mathcal{C}^{\mathrm{init}}_1 \cap \mathcal{C}^{\mathrm{init}}_2} \mathbb{P}_{X_{\mathrm{in}}}\{\tau^\sharp_r \le C_{\mathrm{all}}S_N\} \ge 1 - \eta_{\mathrm{all},N}. \tag{395}$$

For the cross-noise terms in (192) and (362), the respective exponents grow at least as fixed positive multiples of $\sqrt{N}/d_{0,N}$ and $\sqrt{N}/(d_{0,N}\log N)$ for all sufficiently large N. They dominate the logarithms of the polynomial horizons from $d^{-1}_{0,N} \le N^{a_0}$. The remaining exponential terms and N^{-3} terms vanish; since r is fixed their sum in (394) vanishes.

Fixed-reference drift cost and domain The denominator of (314) is the positive, deterministic scale in (197). The termwise calculation (317), the microscopic threshold, and the current-amplitude conversions (318) establish the estimate on $N \geq \max\{N_k^{\mathrm{old}}, N_k^{\mathrm{mic}}\}$. The uniform stage transition and the global threshold $N_\star$ use the same domain. The time comparison in Lemma B.21 applies at every stage restart $\sigma \geq \tau_{k-1}^\sharp$.

COMPLETE COEFFICIENT DICTIONARY

The following dictionary records the explicit coefficient formulas (396)–(419). All numerical constants in the argument have these definitions.

Primitive polynomial, noise, and norm coefficients

$$C_\Phi := p\Lambda_2^2, \qquad \Lambda_2 := \left(\sum_{a=1}^{r} \lambda_a^2\right)^{1/2}, \tag{396}$$

$$C_G := r^{1/4}\sqrt{\alpha_2}, \qquad \alpha_2 := 2p^2\Lambda_2^4 + 64K^2p^2\Lambda_2^2(r\log 5 + 3), \tag{397}$$

$$C_D := p\lambda_+^2(1+r^2), \tag{398}$$

$$C_{D,1} := p\lambda_+^2[(p-1) + r^2(p+1)], \tag{399}$$

$$C_{D,2} := p\lambda_+^2[(p-1)(p-2) + r^2p(p+1)], \tag{400}$$

$$C_{\mathrm{act}} := pr^2\lambda_+^2, \qquad C_{\mathrm{old}} := 2pr^2\lambda_+^2, \tag{401}$$

$$C_\lambda := \frac{\max_{1\leq i,j\leq r}\lambda_i\lambda_j}{\min_{1\leq i,j\leq r}\lambda_i\lambda_j}, \tag{402}$$

$$K_\lambda := 1 + \left(\frac{\lambda_+}{\lambda_-}\right)^{2/p}, \qquad C_{\mathrm{coup}} := K_\lambda^2 C_{D,2}. \tag{403}$$

The second equality in (397) follows by substituting $C_\Phi = p\Lambda_2^2$, $L_A = p\Lambda_2$, $B_G = 32K^2L_A^2(r\log 5 + 3)$, and $\alpha_2 = 2C_\Phi^2 + 2B_G$. Thus it contains no intermediate coefficient whose value is unstated.

Coefficients for stability and growth The coefficients are fixed by the following formulas.

$$\begin{aligned} L_{\mathrm{cross}} &:= 3pK_\lambda\lambda_+^2, & C_{\mathrm{reset}} &:= 1 + (1-e^{-1})^{-1} < 3, \\ D_{\mathrm{old}} &:= 2 + L_{\mathrm{cross}}/\kappa_{\mathrm{old}}, & T_{\mathrm{cl}} &:= 2 + L_{\mathrm{cross}}/\kappa_{\mathrm{cl}}, \\ C_{\mathrm{sur}} &:= K_\lambda^2 C_{\mathrm{old}}, & C_{T,k} &:= \frac{8(16/15)^{p-2}}{\kappa_{II,k}(1-2^{-(p-2)})}, \\ C_* &:= 3C_D. \end{aligned} \tag{404}$$

The conservative coupling coefficient is $K_\lambda^2 C_{D,2}$ as recorded above; the stronger monomial estimate proved in the text is $K_\lambda C_D$. The conservative choice permits one coefficient to be used throughout the stage estimates.

$$\begin{aligned} b_{k,N}^{\mathrm{old}} &= \frac{\chi_N}{192K_\lambda D_{\mathrm{old}}}, & \boldsymbol{b}_{k,N}^{\mathrm{old}} &= \frac{\chi_N}{64} + \frac{3K_\lambda b_N^{\mathrm{pol}}}{\kappa_{\mathrm{old}}h_N}, \\ b_{\times,k,N} &= \frac{\chi_N}{64K_\lambda T_{\mathrm{cl}}}, & \boldsymbol{b}_{k,N}^{\times} &= \frac{\chi_N}{64} + \frac{K_\lambda b_N^{\mathrm{pol}}}{\kappa_{\mathrm{cl}}h_N}. \end{aligned} \tag{405}$$

The Regime-II selection condition is

$$\rho_k \leq \tfrac{1}{4}(48C_DC_{T,k})^{-1/(p-2)}.$$

It is imposed before computing $\gamma_{+,k}$. Therefore the five gamma families are computed from this ρ_k using their recursive formulas.

Stage-dependent deterministic coefficients For every $1 \le k \le r$, first compute

$$\kappa_{II,k} := \tfrac{1}{4} p \lambda_-^2 (1 - \varepsilon_{\mathrm{mac}}^2)(16/17)^{p-1}, \tag{406}$$

$$S_{1,k} := \frac{2(16/15)^{p-1}\varepsilon_{\mathrm{mac}}}{\kappa_{II,k}}, \qquad S_{2,k} := \varepsilon_{\mathrm{mac}} S_{1,k}, \tag{407}$$

$$c_8 := \max\{8, 2/(1-\bar{c})\}. \tag{408}$$

Then the Regime-II propagation coefficient is exactly

$$C_{0,k} := c_8 K_\lambda \exp(C_D S_{2,k})(1 + C_D S_{1,k}) \left(1 + \rho_k + \frac{\gamma_{+,k}}{\gamma_{-,k}}\right). \tag{409}$$

The cross-tube reserve is computed from

$$\kappa_0 := \tfrac{1}{2} p c_p \lambda_-^2 \min\{(1-\varepsilon_{\mathrm{rec}})^p, \min\{1-\varepsilon_{\mathrm{rec}}, \varepsilon_{\mathrm{mac}}/2\}^p\}, \tag{410}$$

$$B_\chi := 1 + \frac{3K_\lambda \alpha_2 C_\delta}{2\kappa_0}\sqrt{1 + \alpha_2 C_\delta^2}, \tag{411}$$

$$C_{\chi,0} := 2, \qquad C_{\chi,k}^{\mathrm{in}} := 8K_\lambda^2 C_{\chi,k-1} + 8B_\chi,$$
$$C_{\chi,k}^{\mathrm{out}} = C_{\chi,k} := 2K_\lambda C_{\chi,k}^{\mathrm{in}} + 8B_\chi. \tag{412}$$

The time coefficients are

$$A_k^{\mathrm{sel}} := 2 + \frac{4}{p(p-2)C_\delta(1-c_{\mathrm{ord}})^{p-1}\lambda_-^2 \gamma_{2,k}^{p-2}}, \tag{413}$$

$$A_k^{\mathrm{mac}} := 2 + \frac{8(16/15)^{p-2}\gamma_{+,k}^{-(p-2)}}{C_\delta \kappa_{II,k}(1 - 2^{-(p-2)})}, \tag{414}$$

$$A_k^{\mathrm{mic}} := A_k^{\mathrm{sel}} + A_k^{\mathrm{mac}}, \tag{415}$$

$$A_k^{\mathrm{cl}} := 2 + \max\{4/\ell_{\mathrm{grow}}, (p+9)/(4\kappa_{\mathrm{cl}})\}, \tag{416}$$

$$C_{I,1} := A_1^{\mathrm{mic}}, \qquad C_{II,1} := A_1^{\mathrm{mic}}, \tag{417}$$

$$C_{\mathrm{fs}} := 2A_1^{\mathrm{mic}} + A_1^{\mathrm{cl}} + 3, \qquad C_k := 2A_k^{\mathrm{mic}} + A_k^{\mathrm{cl}} + 3 \quad (2 \le k \le r), \tag{418}$$

$$C_{\mathrm{all}} := C_{\mathrm{fs}} + \sum_{k=2}^{r} C_k. \tag{419}$$

Here C_δ is the given step-size input. The symbols $\mathcal{C}_{1,k}^{\mathrm{act}}$ and $\mathcal{C}_{2,k}^{\mathrm{gr}}$ denote sets, not numerical coefficients; their definitions are (46) and (47).

Crosscheck of the relabeled drift identities. In the independent row and column relabeling of (123), set

$$\mathscr{G}(M) := \Lambda_\mu M^{\circ(p-1)} \Lambda_\nu.$$

Direct substitution into (100) gives

$$\frac{D_{ab}(M)}{\sqrt{N}} = p\mu_a \nu_b M_{ab}^{p-1} - \frac{p}{2}\sum_{c,d=1}^{r}\left(\mu_c \nu_b M_{ad} M_{cd} M_{cb}^{p-1} + \mu_c \nu_d M_{ad} M_{cb} M_{cd}^{p-1}\right), \tag{420}$$

$$\frac{D(M)}{\sqrt{N}} = p\left[\mathscr{G}(M) - M\,\mathrm{Sym}(M^{\mathsf{T}}\mathscr{G}(M))\right]. \tag{421}$$

In the two-monomial census, the first coefficient is $\mu_c \nu_b$, whereas the second is $\mu_c \nu_d$. Both are bounded by λ_+^2. Their factor supports yield the old-block and all-active error bounds used in the microscopic comparison.

For clarity, at a diagonal $\overline{M}$ and in a direction H, put $\dot{\mathscr{G}} = \mathscr{G}'(\overline{M})[H]$. Differentiation of the identity gives

$$\frac{D'(\overline{M})[H]}{\sqrt{N}} = p\dot{\mathscr{G}} - pH\overline{M}\mathscr{G}(\overline{M}) - p\overline{M}\,\mathrm{Sym}\Big(H^{\mathsf{T}}\mathscr{G}(\overline{M}) + \overline{M}\dot{\mathscr{G}}\Big). \tag{422}$$

For $p \geq 3$, the off-diagonal entries of $\dot{\mathscr{G}}$ vanish. Consequently the two-coordinate off-diagonal Jacobian is still exactly (126). On the diagonal the identity gives $D_{aa}(\overline{M})/\sqrt{N} = pL_a q_a^{p-1}(1-q_a^2)$, as used in (130). The letters m_a, m_b in the diagonal differentiation calculation refer to q_a, q_b. The indexed $\nu_b = \lambda_{j_b^*}$ are loss weights; they are distinct from the unindexed noise standard deviation ν.

Conclusion of Theorem B.1. The sets $E_s^\sharp(\varepsilon_{\mathrm{rec}})$ in (49) are exactly the sets $E_s^\sharp$ in the theorem. Therefore the recursively defined certified hitting times $\tau_s^\sharp$ in (49) equal τ_s^B pathwise. The uniform conditional estimate (386), its first-stage counterpart (299), and the capped stopping-time induction (389) give, for fixed r,

$$\inf_{X_0 \in \mathcal{C}_N} \mathbb{P}_{X_0}\{\tau_s^B - \tau_{s-1}^B \leq C_s S_N,\ X_{\tau_s^B} \in E_s^\sharp\ (s=1,\dots,r)\} \geq 1 - \eta_{\mathrm{fs,N}} - \sum_{s=2}^{r} \eta_{s,N} = 1 - o(1).$$

At the entrance to stage k, its active leader is at most $\gamma_{1,k}/\sqrt{N} < 1 - \varepsilon_{\mathrm{rec}}$ for all sufficiently large N; at a hit of $E_k^\sharp$ it is at least $1 - \varepsilon_{\mathrm{rec}}$. Thus $\tau_{k-1}^B < \tau_k^B$ on this event, including $k = 1$ because $X_0 \in \mathcal{C}_N$. Summing the stage bounds yields $\tau_r^B \leq (\sum_s C_s) S_N$. Since the distinct greedy pairs define the permutation $\pi(j_s^*) = i_s^*$,

$$\|X_{\tau_r^B} - VP_\pi\|_F^2 = 2r - 2\sum_{s=1}^{r} m_{i_s^* j_s^*}(X_{\tau_r^B}) \leq 2r\varepsilon_{\mathrm{rec}},$$

as asserted. □

Role of the reconstruction theorem. The theorem identifies the order of the latent greedy matching and proves a polynomial-time certified recovery event. Its sharp-set formulation is the stagewise dynamical guarantee required when the reporting result refers to recovered spikes; Proposition B.25 specifies how the earlier raw threshold relates to that certified event.

Scope and attribution. The proof in this section is a quantitative reconstruction and refinement of the sequential-recovery argument of Ben Arous et al. (2026); we do not claim substantive novelty for this reconstruction. The explicit constants, stage-dependent estimates, and conditional bounds at stopping-time restarts are recorded to supply the precise dynamical inputs needed in the subsequent proof of the observable reporting guarantee. Accordingly, the auxiliary results and their detailed proofs are included as technical preparations for that analysis, rather than as independent contributions to the theory of sequential recovery.

C HESSIAN SPECTRAL OBSERVABILITY

We prove the uniform observable approximation in Theorem 3.1. The original fresh-noise Hessian is compared conditionally with an auxiliary Gaussian matrix; its bulk and edge are then transferred to the lower-bulk eigenvalue. We give the probability-space construction and the index-counting step explicitly. The matrix universality results of Brailovskaya & van Handel (2024) are invoked by citation, without restating their theorems.

C.1 INDEPENDENT COORDINATES OF THE SYMMETRIZED NOISE

We use the raw-noise convention of (3). In particular, independence is a property of the ordered coordinates of $\widehat{W}^t$, rather than the entries of $W^t = \mathrm{Sym}(\widehat{W}^t)$. Write $\mathscr{T}_N = \mathrm{Sym}^p(\mathbb{R}^N)$, with the Frobenius inner product inherited from the full tensor space. Let $\mathcal{I}_{N,p}$ index the orbits of $[N]^p$ under permutations of the p positions. For an orbit $\mathcal{O}_\alpha$ of size d_α, set

$$E_\alpha = \frac{1}{\sqrt{d_\alpha}} \sum_{\mathbf{a} \in \mathcal{O}_\alpha} e_{a_1} \otimes \cdots \otimes e_{a_p}, \qquad M_N := |\mathcal{I}_{N,p}| = \binom{N+p-1}{p} = O_p(N^p). \tag{423}$$

The orbit supports are disjoint, so $(E_\alpha)_{\alpha \in \mathcal{I}_{N,p}}$ is an orthonormal basis of $\mathscr{T}_N$. Define

$$\zeta_\alpha^t := \frac{\langle \widehat{W}^t, E_\alpha \rangle_F}{\nu} = \frac{1}{\nu\sqrt{d_\alpha}} \sum_{\mathbf{a} \in \mathcal{O}_\alpha} \widehat{W}_{\mathbf{a}}^t, \qquad W^t = \nu \sum_{\alpha \in \mathcal{I}_{N,p}} \zeta_\alpha^t E_\alpha. \tag{424}$$

Lemma C.1 (Independent orbit coordinates)**.** *The family $(\zeta_\alpha^t)_{\alpha,t}$ is independent, centered, and has unit variance. Its members need not be identically distributed, but for a constant depending only on the raw-noise parameters,*

$$\sup_{\alpha,t} \|\zeta_\alpha^t\|_{\psi_2} \le CK/\nu.$$

Consequently (424) *is an independent-summand decomposition of each fresh symmetrized observation.*

Proof. Distinct orbits contain disjoint ordered coordinates, and raw tensors are independent across t. Thus the normalized orbit sums are independent. Centering and the variance identity follow from the independence and variance ν^2 of each raw coordinate: $\mathbb{E}(\zeta_\alpha^t)^2 = (\nu^2 d_\alpha)^{-1} d_\alpha \nu^2 = 1$. The sub-Gaussian bound for a normalized sum of independent centered variables, for example (Vershynin, 2018, Proposition 2.6.1), gives

$$\|\zeta_\alpha^t\|_{\psi_2}^2 \le \frac{C}{\nu^2 d_\alpha} \sum_{\mathbf{a}\in\mathcal{O}_\alpha} \|\widehat{W}_{\mathbf{a}}^t\|_{\psi_2}^2 \le CK^2/\nu^2.$$

Finally, orthogonal projection onto $\mathscr{T}_N$ gives the expansion in (424). □

Role of Lemma C.1. It supplies the independent self-adjoint summands needed to apply Brailovskaya & van Handel (2024) conditionally on the past. Treating the entries of the symmetrized tensor as independent would not justify that application.

C.2 Auxiliary probability space and covariance-matched proxy

Here is the probability-space convention used throughout this spectral argument. For each fixed N, the original experiment can be realized on the canonical product

$$\Omega_N = \mathrm{St}(N,r) \times ((\mathbb{R}^N)^{\otimes p})^{\mathbb{N}}, \qquad \mathbb{P}_N = \mu_{0,N} \otimes \mu_{\widehat{W},N}^{\otimes\mathbb{N}},$$

where $\mu_{0,N}$ is the initialization law and $\mu_{\widehat{W},N}$ is the raw-noise law. Its natural filtration is the completion of $\mathcal{F}_{t,N}^0 = \sigma(X_0, \widehat{W}^1, \dots, \widehat{W}^t)$; this is the observation filtration of Section 2.1, since $\widehat{Y}^t$ and $\widehat{W}^t$ differ by the fixed signal. Measurable iteration of the update makes X_{t-1} $\mathcal{F}_{t-1,N}$-measurable, while all the coordinates $(\zeta_\alpha^t)_\alpha$ are independent of $\mathcal{F}_{t-1,N}$. One may equally start from any filtered realization with these laws: the product construction specifies precisely the extension used below.

To make the Gaussian comparison simultaneous in t and j, take a second canonical factor

$$\Omega_{G,N} = \mathbb{R}^{\mathcal{I}_{N,p}\times\mathbb{N}}, \qquad \mathbb{P}_{G,N} = \mathcal{N}(0,1)^{\otimes(\mathcal{I}_{N,p}\times\mathbb{N})},$$

with coordinate maps g_α^t. On $\widetilde{\Omega}_N = \Omega_N \times \Omega_{G,N}$ use $\widetilde{\mathbb{P}}_N = \mathbb{P}_N \otimes \mathbb{P}_{G,N}$ and the completed filtration

$$\widetilde{\mathcal{F}}_{t,N} = \overline{\sigma(X_0, \widehat{W}^1, \dots, \widehat{W}^t, g_\alpha^s : \alpha \in \mathcal{I}_{N,p},\ 1 \le s \le t)}^{\widetilde{\mathbb{P}}_N}. \tag{425}$$

The first-coordinate projection π_1 lifts the original trajectory without changing it. In particular, if τ is any original $\mathcal{F}_{t,N}$-stopping time, its lift $\tau \circ \pi_1$ is a $\widetilde{\mathcal{F}}_{t,N}$-stopping time, with the same pathwise values on the first coordinate and the same law. No comparison argument below conditions the original process on a future event. At each time t, the Gaussian vector $(g_\alpha^t)_\alpha$ is independent both of $\widetilde{\mathcal{F}}_{t-1,N}$ and of the entire original noise sequence. The proxy matrix, whose coefficients depend on X_{t-1}, is in general *not* independent of the past.

Fix a Borel measurable selection $X \mapsto Q_X$ of an orthonormal basis of $\mathrm{span}(X)^\perp$; a deterministic pivoted Gram–Schmidt rule gives such a selection. This affects only matrix coordinates, not the eigenvalues in (17). At $X = X_{t-1}$, write $x = x_{j,t-1}$ and put

$$\begin{aligned} B_{j,t,\alpha}(X) &:= \frac{p\lambda_j}{\sqrt{N}} \left[\langle E_\alpha, x^{\otimes p}\rangle_F I_{d_N} - (p-1) Q_X^\top E_\alpha(x^{\otimes(p-2)}) Q_X \right], \\ Z_{j,t,\alpha} &:= \nu \zeta_\alpha^t B_{j,t,\alpha}(X_{t-1}), \\ \mathscr{G}_{j,t} &:= \nu \sum_{\alpha\in\mathcal{I}_{N,p}} g_\alpha^t B_{j,t,\alpha}(X_{t-1}). \end{aligned} \tag{426}$$

Then $A_{j,t} = \sum_\alpha Z_{j,t,\alpha}$ in (16). Conditional on $\widetilde{\mathcal{F}}_{t-1,N}$, the $B_{j,t,\alpha}$ are deterministic real symmetric matrices, the $Z_{j,t,\alpha}$ are independent and centered, and $\mathscr{G}_{j,t}$ is a centered jointly Gaussian real symmetric matrix. Moreover,

$$\begin{aligned}\widetilde{\mathbb{E}}\big[(A_{j,t})_{ab}(A_{j,t})_{cd} \mid \widetilde{\mathcal{F}}_{t-1,N}\big] &= \nu^2 \sum_\alpha (B_{j,t,\alpha})_{ab}(B_{j,t,\alpha})_{cd} \\ &= \widetilde{\mathbb{E}}\big[(\mathscr{G}_{j,t})_{ab}(\mathscr{G}_{j,t})_{cd} \mid \widetilde{\mathcal{F}}_{t-1,N}\big].\end{aligned} \tag{427}$$

We use the same g_α^t for every column j at time t. Thus even the cross-covariances of column proxies match those of the corresponding noise compressions; their independence across columns is neither stated nor needed.

C.3 Uniform comparison with the Gaussian proxy

All matrix parameters and expectations in this subsection are conditional on $\widetilde{\mathcal{F}}_{t-1,N}$, with X_{t-1} held fixed. This regular conditional law exists on the canonical standard Borel raw space; identities with the completed filtration are understood almost surely. For the decomposition $A_{j,t} = \sum_\alpha Z_{j,t,\alpha}$, use the parameters

$$\begin{aligned}\sigma(A_{j,t})^2 &:= \Big\|\widetilde{\mathbb{E}}[A_{j,t}^2 \mid \widetilde{\mathcal{F}}_{t-1,N}]\Big\|_{\mathrm{op}}, \\ \sigma_*(A_{j,t})^2 &:= \sup_{\|u\|=\|v\|=1} \widetilde{\mathbb{E}}[|u^\top A_{j,t} v|^2 \mid \widetilde{\mathcal{F}}_{t-1,N}], \\ \overline{R}(A_{j,t})^2 &:= \widetilde{\mathbb{E}}\Big[\max_\alpha \|Z_{j,t,\alpha}\|_{\mathrm{op}}^2 \,\Big|\, \widetilde{\mathcal{F}}_{t-1,N}\Big].\end{aligned}$$

To calculate them, use Parseval in $\mathscr{T}_N$ and the orthogonality $Q_X^\top x = 0$. For $y_a = Q_X e_a$, let $T_{ab} = \mathrm{Sym}(y_a \otimes y_b \otimes x^{\otimes(p-2)})$.

Lemma C.2 (Covariance of transverse tensor contractions)**.** *For $1 \le a, b, c, d \le d_N$ and the T_{ab} above,*

$$\langle T_{ab}, T_{cd}\rangle_F = \frac{\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}}{p(p-1)}, \quad \langle x^{\otimes p}, T_{ab}\rangle_F = 0, \quad \sum_\alpha \langle E_\alpha, x^{\otimes p}\rangle_F^2 = 1. \tag{428}$$

Proof. For distinct positions $k, l \in [p]$, let $R_{ab}^{k,l}$ be the pure tensor with y_a in slot k, y_b in slot l, and x in every other slot. Symmetrization gives $T_{ab} = [p(p-1)]^{-1} \sum_{k \ne l} R_{ab}^{k,l}$. Because each y_a is orthogonal to x, the inner product of $R_{ab}^{k,l}$ with $R_{cd}^{m,n}$ vanishes unless $(m,n) = (k,l)$ or $(m,n) = (l,k)$. In the two cases the inner products are $\delta_{ac}\delta_{bd}$ and $\delta_{ad}\delta_{bc}$, respectively. Summing over the $p(p-1)$ ordered pairs proves the first identity. Each T_{ab} has a factor perpendicular to x, proving the second. The third follows from Parseval and $\|x^{\otimes p}\|_F = 1$. □

Role of Lemma C.2. This covariance computation fixes the normalization of the Gaussian proxy's GOE part and gives the exact variance and entrywise bounds below; the next subsection uses the same formula to identify its spectral edge.

In particular, for fixed unit u, summing over an orthonormal basis of the complement yields $(d_N + 1)/[p(p-1)]$ for the squared contracted term. Direct substitution into (426) therefore gives

$$\begin{aligned}\sigma(A_{j,t})^2 &= \frac{\nu^2 \lambda_j^2}{N}\big[p^2 + p(p-1)(d_N+1)\big], \\ \sigma_*(A_{j,t})^2 &\le \frac{\nu^2 \lambda_+^2}{N}\big[p^2 + 2p(p-1)\big], \\ \|B_{j,t,\alpha}\|_{\mathrm{op}} &\le \frac{p^2 \lambda_+}{\sqrt{N}}, \qquad \overline{R}(A_{j,t}) \le C_R \sqrt{\frac{\log N}{N}}.\end{aligned} \tag{429}$$

The last bound uses $M_N = O_p(N^p)$ and the uniform sub-Gaussian maximum estimate. All constants are independent of the predictable state, t, and j; for large N, $0 < c_\sigma \le \sigma(A_{j,t}) \le C_\sigma$.

Proposition C.3 (Conditional Gaussian comparison on a polynomial horizon)**.** *Suppose $Q_N \le N^B$ for a fixed $B < \infty$, with p, r fixed. On the auxiliary product space, there is a deterministic $\Delta_N \downarrow 0$ such that*

$$\widetilde{\mathbb{P}}_N \left(\max_{\substack{t \le Q_N \\ j \le r}} d_{\mathrm{H}}(\operatorname{sp} A_{j,t}, \operatorname{sp} \mathscr{G}_{j,t}) > \Delta_N \right) \longrightarrow 0. \tag{430}$$

For each fixed $\varphi \in C_c^\infty(\mathbb{R})$, writing $S_\varphi(M) := d_N^{-1} \operatorname{Tr} \varphi(M)$, one also has

$$\max_{\substack{t \le Q_N \\ j \le r}} |S_\varphi(A_{j,t}) - S_\varphi(\mathscr{G}_{j,t})| \longrightarrow 0 \quad \textit{in } \widetilde{\mathbb{P}}_N\textit{-probability}. \tag{431}$$

Proof. Hausdorff comparison. Let $L > B + 3$, set $x_N = L \log N$, and choose

$$R_N = C_\rho (\log N / N)^{1/4}.$$

By (429), taking C_ρ large enough ensures $R_N \ge \overline{R}(A_{j,t})^{1/2} \sigma(A_{j,t})^{1/2} + \sqrt{2}\, \overline{R}(A_{j,t})$ uniformly. Also, the raw-coordinate sub-Gaussian tail and $\|B_{j,t,\alpha}\|_{\mathrm{op}} \le p^2 \lambda_+ / \sqrt{N}$ give

$$\widetilde{\mathbb{P}}_N \left(\max_\alpha \|Z_{j,t,\alpha}\|_{\mathrm{op}} > R_N \,\middle|\, \widetilde{\mathcal{F}}_{t-1,N} \right) \le q_N := C N^p e^{-cNR_N^2} \le C N^p e^{-c\sqrt{N \log N}}. \tag{432}$$

We apply (Brailovskaya & van Handel, 2024, Theorem 2.8) to the regular conditional law of the independent centered matrices $Z_{j,t,\alpha}$. The Gaussian comparator required there is exactly $\mathscr{G}_{j,t}$ by (427). The theorem's bounded-maximal-summand event and (432) give

$$\widetilde{\mathbb{P}}_N \left(d_{\mathrm{H}}(\operatorname{sp} A_{j,t}, \operatorname{sp} \mathscr{G}_{j,t}) > C \varepsilon_{j,t,N} \,\middle|\, \widetilde{\mathcal{F}}_{t-1,N} \right) \le d_N e^{-x_N} + q_N, \tag{433}$$

where the parameter bounds in (429) yield

$$\begin{aligned} \varepsilon_{j,t,N} &= \sigma_*(A_{j,t}) \sqrt{x_N} + R_N^{1/3} \sigma(A_{j,t})^{2/3} x_N^{2/3} + R_N x_N \\ &\le e_N := C_L \left[N^{-1/2} (\log N)^{1/2} + N^{-1/12} (\log N)^{3/4} + N^{-1/4} (\log N)^{5/4} \right] = o(1). \end{aligned} \tag{434}$$

Take expectations in (433), then union bound over at most rQ_N pairs. This gives failure probability at most $rQ_N(N^{1-L} + q_N) = o(1)$ and proves (430) with a deterministic $\Delta_N = C e_N$ (or its decreasing envelope).

Expected smooth statistics. The unbounded-summand condition for (Brailovskaya & van Handel, 2024, Theorem 2.11) holds uniformly, since

$$\frac{\overline{R}(A_{j,t}) (\log d_N)^{5/3}}{\sigma(A_{j,t})} \le C \frac{(\log N)^{13/6}}{\sqrt{N}} = o(1).$$

For fixed $\varphi \in C_c^\infty(\mathbb{R})$, that theorem applied conditionally and then traced gives

$$\begin{aligned} \left| \widetilde{\mathbb{E}}_{t-1} S_\varphi(A_{j,t}) - \widetilde{\mathbb{E}}_{t-1} S_\varphi(\mathscr{G}_{j,t}) \right| &\le C \|\varphi\|_{W^{3,1}} \left[\sigma_*(A_{j,t}) + \overline{R}(A_{j,t})^{1/10} \sigma(A_{j,t})^{9/10} \right] \\ &\le \beta_{N,\varphi} := C_\varphi \left[N^{-1/2} + \left(\frac{\log N}{N} \right)^{1/20} \right] = o(1). \end{aligned} \tag{435}$$

Here $\widetilde{\mathbb{E}}_{t-1}$ conditions on $\widetilde{\mathcal{F}}_{t-1,N}$ and the normalization $d_N^{-1} \operatorname{Tr}$ turns the operator-norm estimate into the displayed scalar estimate.

Fluctuations of the smooth statistics. The Gaussian part of (Brailovskaya & van Handel, 2024, Proposition 5.10), applied conditionally to each diagonal matrix element and then union bounded over d_N of them, gives for $x > 0$

$$\widetilde{\mathbb{P}}_N \left(|S_\varphi(\mathscr{G}_{j,t}) - \widetilde{\mathbb{E}}_{t-1} S_\varphi(\mathscr{G}_{j,t})| > C_\varphi \sqrt{x/N} \,\middle|\, \widetilde{\mathcal{F}}_{t-1,N} \right) \le 4 d_N e^{-c_0 x}. \tag{436}$$

For $A_{j,t}$ the summands are unbounded. Set $Z_\alpha^{(R)} = Z_{j,t,\alpha} \mathbf{1}_{\{\|Z_{j,t,\alpha}\|_{\mathrm{op}} \le R_N\}}$, $A_{j,t}^{(R)} = \sum_\alpha Z_\alpha^{(R)}$, and center each truncated summand: $\widehat{Z}_\alpha^{(R)} = Z_\alpha^{(R)} - \widetilde{\mathbb{E}}_{t-1} Z_\alpha^{(R)}$. This centering is needed even though

the original summands are centered, since the orbit coordinates need not have symmetric laws. The $\widehat{Z}_\alpha^{(R)}$ remain conditionally independent, are bounded by $2R_N$, and satisfy

$$\Big\| \sum_\alpha \widetilde{\mathbb{E}}_{t-1}[(\widehat{Z}_\alpha^{(R)})^2] \Big\|_{\mathrm{op}} \le \sigma(A_{j,t})^2 \le C_\sigma^2, \qquad \sigma_*(A_{j,t}^{(R)}) \le \sigma_*(A_{j,t}).$$

Indeed, truncation can only decrease the variance of a centered scalar coefficient, and the conditional sum of the squared deterministic coefficient matrices is positive semidefinite. Conditional matrix Bernstein (Tropp, 2015, Theorem 1.6.2) and integration of its tail give

$$\widetilde{\mathbb{E}}_{t-1}\big\|A_{j,t}^{(R)} - \widetilde{\mathbb{E}}_{t-1}A_{j,t}^{(R)}\big\|_{\mathrm{op}}^2 \le C\left[C_\sigma^2 \log(2d_N) + R_N^2 \log^2(2d_N)\right] \le C \log N.$$

These inputs meet the bounded-summand hypotheses in the non-Gaussian part of (Brailovskaya & van Handel, 2024, Proposition 5.10); hence, with a constant depending on φ,

$$\begin{aligned} &\widetilde{\mathbb{P}}_N\left(|S_\varphi(A_{j,t}^{(R)}) - \widetilde{\mathbb{E}}_{t-1}S_\varphi(A_{j,t}^{(R)})| > \chi_{N,\varphi}(x) \,\middle|\, \widetilde{\mathcal{F}}_{t-1,N}\right) \le 4d_N e^{-c_0 x}, \\ &\chi_{N,\varphi}(x) := C_\varphi\left[(R_N + R_N^2)x + \left\{N^{-1/2} + R_N^{1/2}(\log N)^{1/4} + R_N(\log N)^{1/2}\right\}\sqrt{x}\right]. \end{aligned} \tag{437}$$

On the event $\max_\alpha \|Z_{j,t,\alpha}\|_{\mathrm{op}} \le R_N$, $A_{j,t} = A_{j,t}^{(R)}$. Because $|S_\varphi(M)| \le \|\varphi\|_\infty$, their conditional means differ by at most $2\|\varphi\|_\infty q_N$; therefore (437) applies to $A_{j,t}$ after adding $2\|\varphi\|_\infty q_N$ to the threshold and q_N to the failure probability.

Choose $x = L_0 \log N$ with $c_0 L_0 > B + 3$. Explicitly,

$$\chi_{N,\varphi}(L_0 \log N) = O_\varphi\left(N^{-1/8}(\log N)^{7/8} + N^{-1/4}(\log N)^{5/4} + N^{-1/2}(\log N)^{3/2}\right) = o(1).$$

Combining the two fluctuation bounds with (435), and using the tower property, gives a deterministic threshold

$$\eta_{N,\varphi} := \beta_{N,\varphi} + \chi_{N,\varphi}(L_0 \log N) + C_\varphi\sqrt{\log N/N} + 2\|\varphi\|_\infty q_N = o(1)$$

whose uniform failure probability is at most $rQ_N(8d_N e^{-c_0 L_0 \log N} + q_N) = o(1)$. This proves (431). □

Role of Proposition C.3. The Hausdorff statement controls how far the lowest noise eigenvalue can fall below the Gaussian edge. Smooth-statistic comparison controls a positive fraction of noise eigenvalues in a fixed interval immediately above that edge. Both inputs are needed: Hausdorff distance compares sets of eigenvalues and alone does not control the multiplicity at the index ℓ_N in (17). They will be combined with the Gaussian bulk and edge bounds in the following subsection.

C.4 The exact Gaussian proxy and its spectral edge

Fix a time $t \le Q_N$, a column $j \le r$, and condition on the enlarged past $\widetilde{\mathcal{F}}_{t-1,N}$. Write $X = X_{t-1}$, $x = x_{j,t-1}$, $d_N = N - r$, and let Q_X be the measurable orthonormal frame of $\mathrm{span}(X)^\perp$ chosen above. All expectations in this subsection are under the conditional law for almost every past history. In terms of the same independent Gaussian coordinates $\{g_\alpha^t\}_\alpha$ used to construct $\mathscr{G}_{j,t}$, set

$$\mathcal{W}_G^t := \nu \sum_{\alpha \in \mathcal{I}_{N,p}} g_\alpha^t E_\alpha. \tag{438}$$

The orbit tensors E_α are an orthonormal basis of $\mathrm{Sym}^p(\mathbb{R}^N)$. Hence, for symmetric p-tensors T, S,

$$\widetilde{\mathbb{E}}_N[\langle \mathcal{W}_G^t, T\rangle_F \langle \mathcal{W}_G^t, S\rangle_F \mid \widetilde{\mathcal{F}}_{t-1,N}] = \nu^2 \langle T, S\rangle_F. \tag{439}$$

In particular, the proxy for the noise compression $A_{j,t}$ in (16) has the exact representation

$$\mathscr{G}_{j,t} = \frac{p\lambda_j}{\sqrt{N}}\langle \mathcal{W}_G^t, x^{\otimes p}\rangle_F I_{d_N} - \frac{p(p-1)\lambda_j}{\sqrt{N}} Q_X^\top \mathcal{W}_G^t(x^{\otimes(p-2)})Q_X. \tag{440}$$

The contraction identities in Lemma C.2 identify the random matrix part of the proxy as GOE and make its scalar shift independent of that matrix. We use their normalization to determine the edge in (17).

Let H_d denote the normalized GOE matrix with centered jointly Gaussian entries and covariance

$$\mathbb{E}(H_d)_{ab}(H_d)_{cd} = d^{-1}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}). \tag{441}$$

In particular, its diagonal entries have variance $2/d$ and its off-diagonal entries variance $1/d$. Its limiting semicircle support is $[-2, 2]$.

Proposition C.4 (GOE form of the conditional Gaussian proxy)**.** *For each fixed (j, t), conditionally on $\widetilde{\mathcal{F}}_{t-1,N}$,*

$$\mathscr{G}_{j,t} \overset{\mathrm{d}}{=} a_{j,N} H_{d_N} + \theta_{j,t} I_{d_N}, \qquad a_{j,N} = \nu\lambda_j \sqrt{\frac{p(p-1)d_N}{N}}, \qquad \theta_{j,t} \sim \mathcal{N}\left(0, \frac{\nu^2 p^2 \lambda_j^2}{N}\right), \tag{442}$$

where H_{d_N} and $\theta_{j,t}$ are conditionally independent. This is a marginal law for each (j, t); it makes no independence assertion across columns or times. With $a_j := \nu\lambda_j\sqrt{p(p-1)}$, we have $a_{j,N} = a_j\sqrt{1 - r/N}$ and

$$\rho_{p,j} := 2a_j = 2\nu\lambda_j\sqrt{p(p-1)}. \tag{443}$$

Proof. Set $\Gamma_{ab} = \langle \mathcal{W}_G^t, T_{ab}\rangle_F$ and $\chi = \langle \mathcal{W}_G^t, x^{\otimes p}\rangle_F$. Equations (439) and (428) show that

$$\widetilde{\mathbb{E}}_N[\Gamma_{ab}\Gamma_{cd} \mid \widetilde{\mathcal{F}}_{t-1,N}] = \frac{\nu^2}{p(p-1)}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}), \qquad \chi \sim \mathcal{N}(0, \nu^2).$$

The same equations give zero conditional cross covariance between χ and every Γ_{ab}. Joint Gaussianity then implies $\chi \perp\!\!\!\perp \Gamma$ under the conditional law. Comparison with (441) yields $\Gamma \overset{\mathrm{d}}{=} \nu\sqrt{d_N/[p(p-1)]}\, H_{d_N}$. Substitution into (440) gives the coefficient $-a_{j,N}$ on H_{d_N} and the scalar shift $p\lambda_j\chi/\sqrt{N}$. The centered GOE law is invariant under a global sign change, giving (442). □

Role of the GOE identification. The exact conditional law allows the same bulk and lower-edge estimates at every predictable state; the union bound below then covers the polynomial observation horizon in Theorem 3.1.

Write the eigenvalues of a self-adjoint $d_N \times d_N$ matrix in increasing order as $\mu_1(M) \le \cdots \le \mu_{d_N}(M)$, and let $\mu_M = d_N^{-1}\sum_{k=1}^{d_N}\delta_{\mu_k(M)}$ be its empirical spectral measure. Let $\mu_{\mathrm{sc},j}$ be the semicircle law obtained by multiplying a standard semicircle variable by a_j:

$$\mu_{\mathrm{sc},j}(dz) = \frac{2\sqrt{(\rho_{p,j}^2 - z^2)_+}}{\pi\rho_{p,j}^2}\, dz, \qquad \operatorname{supp}\mu_{\mathrm{sc},j} = [-\rho_{p,j}, \rho_{p,j}]. \tag{444}$$

Proposition C.5 (Uniform Gaussian bulk and lower edge)**.** *Assume p, r are fixed and $Q_N \le N^B$ for fixed $B < \infty$. For every fixed $\varphi \in C_c^\infty(\mathbb{R})$,*

$$\max_{t\le Q_N}\max_{j\le r}\left|\int \varphi\, d\mu_{\mathscr{G}_{j,t}} - \int \varphi\, d\mu_{\mathrm{sc},j}\right| \xrightarrow{\widetilde{\mathbb{P}}_N} 0. \tag{445}$$

Moreover, for every fixed $\eta > 0$,

$$\widetilde{\mathbb{P}}_N\left(\max_{t\le Q_N}\max_{j\le r}|\mu_1(\mathscr{G}_{j,t}) + \rho_{p,j}| \ge \eta\right) \longrightarrow 0. \tag{446}$$

Proof. Fix φ and a past history. Under the conditional law in Proposition C.4, put $\mathsf{T}_\varphi(M) = d_N^{-1}\operatorname{Tr}\varphi(M)$. The GOE expected-distribution estimate (Kholopov et al., 2008, Theorem 1) is stated for $H_{d_N}/\sqrt{2}$, whose diagonal and off-diagonal variances are $1/d_N$ and $1/(2d_N)$. Its limiting density is $\pi^{-1}\sqrt{(2 - z^2)_+}$. With $\psi(z) = \varphi(\sqrt{2}a_j z)$, integration by parts against its $O(d_N^{-1})$ expected-CDF error gives

$$\left|\widetilde{\mathbb{E}}_N\mathsf{T}_\varphi(a_j H_{d_N}) - \int \varphi\, d\mu_{\mathrm{sc},j}\right| \le \frac{C\|\psi'\|_{L^1}}{d_N}. \tag{447}$$

Since $|a_{j,N} - a_j| = O(N^{-1})$ uniformly in j, the Lipschitz bound in the scale parameter and $\mathbb{E}[d_N^{-1} \operatorname{Tr} H_{d_N}^2] = 1 + d_N^{-1}$ produce an $O_\varphi(N^{-1})$ change in the expected statistic. The independent centered shift $\theta_{j,t}$ has variance $\nu^2 p^2 \lambda_j^2 / N$; a second-order Taylor expansion in that shift adds at most $\|\varphi''\|_\infty \nu^2 p^2 \lambda_j^2/(2N)$. Thus, uniformly over j, t and histories,

$$\left| \widetilde{\mathbb{E}}_N[\mathsf{T}_\varphi(\mathscr{G}_{j,t}) \mid \widetilde{\mathcal{F}}_{t-1,N}] - \int \varphi \, d\mu_{\mathrm{sc},j} \right| \le C_\varphi / N. \tag{448}$$

Here and below an unconditional expectation of a GOE law in a conditional-law calculation means expectation with respect to its fresh Gaussian coordinates.

For concentration, realize the independent GOE coordinates by standard Gaussians Z_{ab}, $a \le b$, with $(H_{d_N})_{aa} = \sqrt{2/d_N} Z_{aa}$ and $(H_{d_N})_{ab} = Z_{ab}/\sqrt{d_N}$ for $a < b$; realize the scalar shift by another independent standard Gaussian g_0. For two Gaussian coordinate vectors Z, Z', we have $\|H_{d_N}(Z) - H_{d_N}(Z')\|_F = \sqrt{2/d_N}\|Z - Z'\|_2$. By Hoffman–Wielandt, Cauchy–Schwarz, and the Lipschitz bound on φ, the map

$$(Z, g_0) \mapsto \mathsf{T}_\varphi\left(a_{j,N} H_{d_N}(Z) + \frac{\nu p \lambda_j}{\sqrt{N}} g_0 I_{d_N} \right)$$

has Euclidean Lipschitz constant at most $C_\varphi/\sqrt{N}$, uniformly in j and the realized past. Gaussian concentration yields

$$\widetilde{\mathbb{P}}_N\left(|\mathsf{T}_\varphi(\mathscr{G}_{j,t}) - \widetilde{\mathbb{E}}_N[\mathsf{T}_\varphi(\mathscr{G}_{j,t}) \mid \widetilde{\mathcal{F}}_{t-1,N}]| > u \,\middle|\, \widetilde{\mathcal{F}}_{t-1,N} \right) \le 2e^{-c_\varphi N u^2}. \tag{449}$$

Take $u = K_B\sqrt{\log N/N}$ with $c_\varphi K_B^2 > B + 2$, combine (448) and (449), and sum the conditional bounds over $j \le r$ and $t \le Q_N$. This proves (445); no independence between the matrices indexed by different j is needed.

For the edge, the same conditional representation gives

$$\mu_1(\mathscr{G}_{j,t}) + \rho_{p,j} \overset{\mathrm{d}}{=} a_{j,N}[\mu_1(H_{d_N}) + 2] + 2(a_j - a_{j,N}) + \theta_{j,t}. \tag{450}$$

Let $a_* = \min_{j \le r} a_j > 0$, and put $\bar\eta = \min\{\eta, a_*\}$. For all large N, $2|a_j - a_{j,N}| \le \bar\eta/4$, uniformly in j, and $u_{j,N} = \bar\eta/(4a_{j,N}) \in (0, 1]$. The $\beta = 1$ Hermite-ensemble edge bounds (Ledoux & Rider, 2010, Theorem 1), applied to $\sqrt{d_N} H_{d_N}$, give the two deviation tails $C\exp(-cd_N^2 u_{j,N}^3)$ and $C\exp(-cd_N u_{j,N}^{3/2})$; GOE sign symmetry transfers them to the lower edge. The Gaussian tail for $\theta_{j,t}$ therefore gives, for constants depending only on the fixed model parameters,

$$\begin{aligned} &\widetilde{\mathbb{P}}_N\left(|\mu_1(\mathscr{G}_{j,t}) + \rho_{p,j}| \ge \eta \mid \widetilde{\mathcal{F}}_{t-1,N} \right) \\ &\quad \le C_1 e^{-c_1 \bar\eta^3 d_N^2} + C_2 e^{-c_2 \bar\eta^{3/2} d_N} + 2e^{-c_3 \bar\eta^2 N}. \end{aligned} \tag{451}$$

Summing over at most rQ_N queries proves (446). □

Role of the uniform Gaussian estimates. The smooth-statistic limit counts a positive fraction of eigenvalues inside a fixed interval immediately above the lower edge. The smallest-eigenvalue bound controls the opposite side; both enter the lower-bulk lemma below after comparison with $A_{j,t}$.

C.5 From Gaussian comparison to a lower-bulk eigenvalue

The preceding conditional BvH analysis supplies two uniform events on the enlarged probability space: the Hausdorff distance between $\operatorname{sp}(A_{j,t})$ and $\operatorname{sp}(\mathscr{G}_{j,t})$ tends to zero uniformly for $j \le r, t \le Q_N$, and for each fixed $\varphi \in C_c^\infty(\mathbb{R})$ the difference between their normalized smooth spectral statistics tends to zero uniformly. Hausdorff distance compares sets of spectral values and does not control eigenvalue multiplicities. The smooth-statistic assertion supplies the missing count.

Lemma C.6 (Uniform lower-bulk transfer)**.** *Under these two BvH comparisons and Proposition C.5, let ℓ_N be deterministic and satisfy $1 \le \ell_N \le d_N$ and $\ell_N/d_N \to 0$. In particular, ℓ_N is allowed to remain fixed. Then for every fixed $\eta > 0$,*

$$\widetilde{\mathbb{P}}_N\left(\max_{t \le Q_N} \max_{j \le r} |\mu_{\ell_N}(A_{j,t}) + \rho_{p,j}| \ge \eta \right) \longrightarrow 0. \tag{452}$$

Proof. Fix a tolerance $\eta > 0$. For each of the finitely many columns, choose $0 \le \varphi_j \le 1$ smooth, nonzero, and supported in $(-\rho_{p,j}, -\rho_{p,j} + \eta/2)$. Then

$$c_\eta := \min_{j \le r} \int \varphi_j \, d\mu_{\mathrm{sc},j} > 0.$$

The uniform Gaussian bulk conclusion and the smooth-statistic comparison show that, with probability tending to one, for every $t \le Q_N, j \le r$,

$$\int \varphi_j \, d\mu_{A_{j,t}} \ge c_\eta/2.$$

Since φ_j is supported below $-\rho_{p,j} + \eta/2$ and bounded by one, this yields

$$\#\{k : \mu_k(A_{j,t}) \le -\rho_{p,j} + \eta/2\} \ge (c_\eta/2) d_N. \tag{453}$$

As $\ell_N/d_N \to 0$, the right side exceeds ℓ_N for all sufficiently large N. Thus $\mu_{\ell_N}(A_{j,t}) \le -\rho_{p,j} + \eta/2$ simultaneously for every query.

On the uniform Hausdorff event, let $\Delta_N = o(1)$ bound the distance of the two spectra, and combine it with the uniform Gaussian lower-edge event. For all sufficiently large N, with probability tending to one and simultaneously for every t, j,

$$\mu_{\ell_N}(A_{j,t}) \ge \mu_1(A_{j,t}) \ge \mu_1(\mathscr{G}_{j,t}) - \Delta_N > -\rho_{p,j} - \eta/2.$$

For the last strict inequality, invoke the Gaussian edge assertion at tolerance $\eta/4$ and take $\Delta_N \le \eta/4$. The two bounds give $|\mu_{\ell_N}(A_{j,t}) + \rho_{p,j}| < \eta$ for all queries on an event of probability tending to one, proving (452). □

Role of the lower-bulk lemma. The estimate applies equally to $\ell_N - r$ and $\ell_N + r$ when $r + 1 \le \ell_N \le d_N - r$ eventually, including a fixed index $\ell_N \ge r + 1$. Rank-r interlacing for (16) then gives the Hessian approximation (32) in Theorem 3.1, and hence the approximation of $S_{j,t}$ to the overlap sum.

C.6 Transfer to the transverse Hessian and observable scores

Recall from (16) that, for $X = X_{t-1}$ and the latent score $F_{j,t}$ in (13),

$$\mathscr{K}_{j,t} = A_{j,t} + p\lambda_j F_{j,t} I_{d_N} - D_{j,t}, \qquad \operatorname{rank}(D_{j,t}) \le r. \tag{454}$$

The sign or size of $D_{j,t}$ is immaterial to the following rank interlacing bound. Indeed, for the increasingly ordered eigenvalues $\mu_1 \le \cdots \le \mu_{d_N}$ and $r + 1 \le \ell_N \le d_N - r$,

$$\mu_{\ell_N - r}(A_{j,t}) + p\lambda_j F_{j,t} \le \mu_{\ell_N}(\mathscr{K}_{j,t}) \le \mu_{\ell_N + r}(A_{j,t}) + p\lambda_j F_{j,t}. \tag{455}$$

The index assumptions in (17) imply the required range for all sufficiently large N, including when ℓ_N stays fixed: $\ell_N - r \ge 1$ and $\ell_N + r \le d_N$.

Theorem C.7 (Uniform lower-bulk Hessian approximation)**.** *For any fixed $B < \infty$, $Q_N \le N^B$, and every $\epsilon > 0$, under the spectral index conditions of* (17)*,*

$$\mathbb{P}_N\left(\max_{t \le Q_N} \max_{j \le r} \left| \frac{\mu_{\ell_N}(\mathscr{K}_{j,t}) + \rho_{p,j}}{p\lambda_j} - F_{j,t} \right| > \epsilon \right) \longrightarrow 0. \tag{456}$$

Proof. Put $k_N^- := \ell_N - r$ and $k_N^+ := \ell_N + r$. Both are positive integers with $k_N^\pm/d_N \to 0$; they need not diverge. Apply Lemma C.6 to these two index sequences. By (455), uniformly over the rQ_N pairs (j, t),

$$\mu_{k_N^-}(A_{j,t}) + \rho_{p,j} \le \mu_{\ell_N}(\mathscr{K}_{j,t}) - p\lambda_j F_{j,t} + \rho_{p,j} \le \mu_{k_N^+}(A_{j,t}) + \rho_{p,j}.$$

The two outer quantities converge uniformly to zero in probability under the auxiliary product law. The middle event depends only on the original SGD trajectory and its fresh tensor samples, so its probability is identical under the auxiliary and original laws. Finally $p\lambda_j \ge p\lambda_r > 0$, which gives (456). □

Role in the observable approximation. This theorem supplies the Hessian-score part of Theorem 3.1. The rank bound removes the planted deformation from an index beyond its at most r exceptional directions.

Proof of Theorem 3.1. For a predictable X_{t-1}, the symmetric contraction of the raw tensor satisfies the exact identity

$$U_{j,t} - F_{j,t} = \frac{1}{\sqrt{N}} \langle \widehat{W}^t, x_{j,t-1}^{\otimes p} \rangle.$$

Conditionally on the past, the coefficient tensor on the right has Frobenius norm one. Independence and the uniform sub-Gaussian bound on the raw coordinates therefore give, uniformly over predictable X_{t-1} and j, a tail bound

$$\mathbb{P}_N(|U_{j,t} - F_{j,t}| > u \,|\, \mathcal{F}_{t-1,N}) \le 2\exp(-cNu^2/K^2).$$

Take $u = C_B\sqrt{\log N/N}$ with $cC_B^2/K^2 > B+2$, then sum the bound over the at most $rQ_N \le rN^B$ pairs. This proves (31).

By (17), $S_{j,t}$ is the first term inside (456); that theorem gives its uniform approximation to $F_{j,t}$. The triangle inequality also gives

$$\max_{j,t} |S_{j,t} - U_{j,t}| \le \max_{j,t} |S_{j,t} - F_{j,t}| + \max_{j,t} |U_{j,t} - F_{j,t}|.$$

The first term tends to zero in probability and the second is $O(\sqrt{\log N/N})$ with high probability. Applying each estimate with half of the requested tolerance proves (32). All bounds were conditional on an arbitrary realized predictable state, so their constants are uniform in the initialization. □

Role in the reporting theorem. Theorem 3.1 supplies the uniform score error used in the fixed-window threshold comparisons of the reporting result. Its conclusion concerns observable approximations and does not by itself assert a stage margin or a reconstruction time.

D REPORTING-TIME TRANSFER AND NUMERICAL GUARANTEES

This section derives the stage event $\mathcal{D}_N$ from the path estimates in Appendix B and transfers it to the fixed-window monitor in (21)–(23). Proposition D.1 supplies two-sided stage duration bounds and fixed threshold margins before and after certification. Theorem 3.1, proved in Appendix C, makes both exact observed scores uniformly close to $F_{j,t}$. The final subsection includes indexed eigensolver errors and the conditions required by each numerical centering scheme.

D.1 HIGH-PROBABILITY STAGE MARGINS FROM THE RECONSTRUCTION

We verify the stage event used by the reporting transfer from the path estimates in Appendix B.3. All stage sets and stopping times follow (49). We use the overlap convention of Section 2.2. Appendix B is written in positive-overlap notation. For even p, its positive active-band and recovered-overlap conditions are understood in terms of $z_{ij}(X) = \operatorname{sgn}(m_{ij}(X_0))m_{ij}(X)$; the certified sets and stopping times are interpreted accordingly. Cross-overlap conditions remain bounds on $|m_{ij}|$. Multiplying a scalar recursion by its frozen sign preserves the direct growth term as a positive multiple of z_{ij}^{p-1} for even p; the absolute remainder, conditional noise, and weighted block bounds are unchanged. Moreover, $m_{ij}^p = z_{ij}^p$ for even p, so the observable-score targets below retain their original definition in terms of the raw overlaps m_{ij}.

Definition of the stage event. For fixed constants $c_d, C_d, C_{\mathrm{st}}, \eta > 0$, put

$$L_N := C_{\mathrm{st}} h_N^{-1} \log N, \qquad b_{0,N} := 1.$$

On trajectories with finite certified times, also put

$$b_{k-1,N} := \tau_{k-1}^B + 2w_N + 1, \qquad 2 \le k \le r.$$

Define $\mathcal{D}_N = \mathcal{D}_N(c_d, C_d, C_{\mathrm{st}}, \eta)$ using the durations $D_{k,N} = \tau_k^B - \tau_{k-1}^B$ from (11). It is the event on which

$$0 = \tau_0^B < \tau_1^B < \cdots < \tau_r^B, \qquad \tau_r^B + 2w_N \le Q_N,$$
$$c_d S_N \le D_{k,N} \le C_d S_N \qquad (1 \le k \le r),$$

and, for every $k \le r$, the latent score in (13) satisfies the following margins, with the last line required for every $j \in J_{k-1} \setminus \{j_k^*\}$:

$$F_{j_k^*,t} \le \theta - \eta, \qquad b_{k-1,N} \le t < \tau_k^B - L_N,$$
$$F_{j_k^*,t} \ge \theta + \eta, \qquad \tau_k^B < t \le \tau_k^B + 2w_N,$$
$$F_{j,t} \le \theta - \eta, \qquad b_{k-1,N} \le t \le \tau_k^B + 2w_N.$$

All time intervals above are restricted to integer observation times. If a certified time is infinite, $\mathcal{D}_N$ does not occur. For $k \ge 2$, the preceding stage's competitor margin extends through $b_{k-1,N} - 1$, so the two stages leave no gap at their handoff. No leader-score margin is imposed on $[\tau_k^B - L_N, \tau_k^B]$. The proposition below proves the high-probability occurrence of this event for a suitable fixed choice of constants; it is not an additional assumption of Theorem 3.2.

Proposition D.1 (High-probability stage margins)**.** *Assume the hypotheses of Theorem B.1, and use the initialization and certified-stage conventions of Appendix B. Let the observation horizon, threshold, and window satisfy the conditions in* (12) *and* (20)*; in particular,*

$$Q_N \le N^B, \qquad Q_N/S_N \longrightarrow \infty, \qquad 0 < \theta < \lambda_-(1 - \varepsilon_{\mathrm{rec}})^p, \qquad h_N w_N \longrightarrow 0.$$

There exist fixed constants $c_d, C_d, \eta, C_{\mathrm{st}} > 0$ such that the event $\mathcal{D}_N$ defined above satisfies, for all sufficiently large N,

$$\inf_{x \in \mathcal{C}_N} \mathbb{P}_N^x(\mathcal{D}_N) \ge 1 - \eta_{\mathrm{fs},N} - \sum_{k=2}^{r} \eta_{k,N} - Q_N e^{-c_R N} \longrightarrow 1. \tag{457}$$

Here $\eta_{\mathrm{fs},N}$ and $\eta_{k,N}$ are the existing failure allowances in (298) *and* (385)*.*

Proof. All constants below are independent of N and of the allowed initial state. We retain $e_k = (i_k^*, j_k^*)$ and write $s_k = \tau_{k-1}^B$ on successful stage entrances. Put $q_k(u) := z_{i_k^* j_k^*}(X_u)$, where $z = m$ on the positive branch. The proof uses the path estimates in Appendix B, not only the endpoint statement of Theorem B.1.

A common event retaining the path estimates. Retain the past-measurable endpoint events B_k and deterministic caps from the induction leading to (389), with $B_0 = \Omega$. At each capped stage restart, let $\mathcal{G}_{k,N}$ be the composite good event used in the original stage proof: the Regime-I comparison, old-block stability when $k \geq 2$, Regime-II transport on its measurable valid entrance, and cleanup on its measurable valid entrance. A component whose valid entrance does not occur is taken to be vacuous. The conditional-expectation arguments in (300), (325), and the proof leading to (380)–(384) give, on B_{k-1}, conditional failure at most $\eta_{\mathrm{fs},N}$ for $k = 1$ and $\eta_{k,N}$ otherwise. Define

$$\mathcal{P}_N := \bigcap_{k=1}^{r} \big(B_{k-1}^{\mathrm{c}} \cup \mathcal{G}_{k,N}\big), \qquad \epsilon_N^{\mathrm{path}} := \eta_{\mathrm{fs},N} + \sum_{k=2}^{r} \eta_{k,N}.$$

Since B_{k-1} is measurable at the corresponding capped restart, the tower property and a union bound give

$$\sup_{x \in \mathcal{C}_N} \mathbb{P}_N^x(\mathcal{P}_N^{\mathrm{c}}) \leq \epsilon_N^{\mathrm{path}} \longrightarrow 0.$$

The deterministic implications $\mathcal{G}_{1,N} \subseteq B_1$ and $B_{k-1} \cap \mathcal{G}_{k,N} \subseteq B_k$ imply inductively that $\mathcal{P}_N \subseteq B_r$ and that all actual stages enjoy the original path estimates. No earlier good event involving future observations is inserted into a conditional freshness argument. In particular, on $\mathcal{P}_N$,

$$\begin{aligned} \tau_k^B - s_k \leq C_k S_N, \qquad \sigma_{\mathrm{mac},k} \leq \tau_k^B \leq \sigma_{\mathrm{mac},k} + H_{k,N}^{\mathrm{cl}}, \\ H_{k,N}^{\mathrm{cl}} = \lceil A_k^{\mathrm{cl}} h_N^{-1} \log N \rceil. \end{aligned} \tag{458}$$

by (387) and (391), with $H_{k,N}^{\mathrm{cl}}$ defined in (331). This refines the endpoint-only event denoted by $G_N = B_r$ after (389).

A uniform lower bound on every stage duration. Put $d = p - 2$, $c = \bar{c}$, and define

$$\zeta_k := \left(\frac{(1+c)\gamma_{1,k}}{\gamma_{+,k}} \right)^d, \qquad \underline{c}_k := \frac{1 - \zeta_k}{pdC_\delta (1+c)^{p-1} \lambda_+^2 \gamma_{1,k}^d}.$$

By (83)–(84), $0 < \zeta_k < 1$, including the final singleton stage. Let $T_{\mathrm{sel},k}$ be the elapsed first hitting time of $q_k \geq \gamma_{+,k}/\sqrt{N}$ after s_k. The entrance band (251), the upper envelope (97), and $h_N S_N N^{-d/2} = C_\delta$ give

$$q_k(s_k + n) \leq \frac{(1+c)\gamma_{1,k}}{\sqrt{N}} \left[1 - pdC_\delta (1+c)^{p-1} \lambda_+^2 \gamma_{1,k}^d \frac{n}{S_N} \right]^{-1/d}. \tag{459}$$

This holds through the selection update whenever the displayed bracket is positive. Indeed, Lemma B.11 controls the exit update. The Regime-I proof leading to (279) permits both c_{ord} and $\bar{c}$ on the same comparison event; the later-stage proof of (324) makes the same substitution. If $T_{\mathrm{sel},k} < \underline{c}_k S_N$, the bracket in (459) at that time is strictly greater than $\zeta_k > 0$, so the right-hand side is strictly less than $\gamma_{+,k}/\sqrt{N}$, a contradiction. For large N the certified threshold $1 - \varepsilon_{\mathrm{rec}}$ exceeds the selection threshold. Therefore

$$\tau_k^B - s_k \geq T_{\mathrm{sel},k} \geq \underline{c}_k S_N.$$

This argument allows an update to cross both thresholds simultaneously. Take $c_d = \min_k \underline{c}_k > 0$ and $C_d = \max_k C_k < \infty$.

Nonselected overlaps throughout each stage. There is a fixed $K_* < \infty$ such that, on $\mathcal{P}_N$,

$$\max_{s_k \leq u \leq \tau_k^B} \max_{\substack{j \in J_{k-1},\, i \leq r \\ (i,j) \neq e_k}} |m_{ij}(X_u)| \leq K_* N^{-1/2}, \qquad k \leq r. \tag{460}$$

For completeness, this is a whole-interval consequence of the existing proof. The Regime-I strict upper guards hold through selection, by the stopping construction (269) and the exit exclusion (279), with the stage substitution in the proof of (324). The Regime-II bounds (246)–(247) control all active nonleaders throughout transport; exclusion of the common bad exits identifies its endpoint with the actual macro crossing. Lemma B.12 controls overlaps in old recovered rows and still-active columns by a fixed multiple of χ_N. During cleanup, the state domain used in (369) is preserved by (372) and

(374)–(377): new crosses remain $O(N^{-1/2})$, survivors remain microscopic, and old crosses remain $O(\chi_N)$. Finally, (458) places the first certified time before the constructed cleanup endpoint. Since

$$\chi_N = (\log N)^{-1/2} N^{-(p-1)/4} \le N^{-1/2}$$

for all sufficiently large N, and r is fixed, these bounds yield (460). In particular, throughout $s_k \le u \le \tau_k^B$,

$$\sum_i \lambda_i m_{ij}(X_u)^p \le r\lambda_+ K_*^p N^{-p/2} = o(1), \qquad j \in J_{k-1} \setminus \{j_k^*\}. \tag{461}$$

The absolute-value bound proves this for odd as well as even p.

A global one-step bound. Let $\mathcal{H}_N$ be the event that the Gram bound in Lemma B.3 holds for all Q_N updates. Conditional application of (111) and a union bound give

$$\sup_{x \in \mathcal{C}_N} \mathbb{P}_N^x(\mathcal{H}_N^c) \le Q_N e^{-c_R N}.$$

On this event, write $g_t = \nabla_{\mathrm{St}} \mathcal{L}(X_{t-1}; Y^t)$. The bound $\|g_t\|_F \le C_G\sqrt{N}$, with $C_G = r^{1/4}\sqrt{\alpha_2}$ as in (57), tangency, and the exact polar formula give

$$\|X_t - X_{t-1}\|_{\mathrm{op}} \le C_G h_N + \frac{\alpha_2 h_N^2}{2}\sqrt{1+\alpha_2 h_N^2} \le C_{\mathrm{step}} h_N$$

for a fixed C_{step} and large N. Consequently, on this single event,

$$|m_{ij}(X_u) - m_{ij}(X_v)| \le C_{\mathrm{step}} h_N |u - v|, \qquad 0 \le u, v \le Q_N. \tag{462}$$

The times u, v may be random. Moreover, the upper stage bounds already imply $\tau_r^B + 2w_N \le Q_N$ on $\mathcal{P}_N$ for large N, because

$$\tau_r^B \le \Big(\sum_k C_k\Big) S_N, \qquad \frac{w_N}{S_N} = \frac{h_N w_N}{C_\delta N^{(p-2)/2}} \longrightarrow 0, \qquad Q_N/S_N \longrightarrow \infty.$$

Thus (462) covers all times used below.

From a fixed small overlap to certification. Choose

$$a := \min\left\{\frac{\varepsilon_{\mathrm{mac}}}{2}, \left(\frac{\theta}{4\lambda_+}\right)^{1/p}\right\} > 0, \qquad \sigma_{a,k} := \inf\{u \ge s_k : q_k(u) \ge a\}.$$

Work on $\mathcal{P}_N \cap \mathcal{H}_N$. Before selection the leader is $O(N^{-1/2})$; (462) shows that the selection update also ends at $o(1)$. Hence, for large N, $\sigma_{a,k}$ occurs after selection and no later than $\sigma_{\mathrm{mac},k}$. Using selection as the local origin, the same auxiliary process from (228)–(230) satisfies, up to the macro crossing,

$$\widetilde{q}_n = q_n - S_q(n), \qquad |S_q(n)| \le a_{k,N}^{\mathrm{mic}} \le \frac{\gamma_{+,k}}{32\sqrt{N}}, \qquad \widetilde{q}_{n+1} - \widetilde{q}_n \ge 2\kappa_{II,k} h_N \widetilde{q}_n^{p-1} > 0.$$

Here $a_{k,N}^{\mathrm{mic}}$ denotes the existing cumulative-noise allowance, not the fixed level a. At $\sigma_{a,k}$, $\widetilde{q} \ge a/2$ for large N. Thereafter each pre-crossing increment is at least $2\kappa_{II,k} h_N (a/2)^{p-1}$, while every pre-crossing value is less than $\varepsilon_{\mathrm{mac}} + a_{k,N}^{\mathrm{mic}}$. Summing only up to the last pre-crossing state gives

$$\sigma_{\mathrm{mac},k} - \sigma_{a,k} \le 1 + C_{a,k} h_N^{-1}, \qquad C_{a,k} := \frac{\varepsilon_{\mathrm{mac}} + 1}{2\kappa_{II,k}(a/2)^{p-1}}.$$

No comparison is restarted at $\sigma_{a,k}$ and no additional probability cost is incurred. By (458), the fixed choice

$$C_{\mathrm{st}} := \max_k (A_k^{\mathrm{cl}} + C_{a,k} + 3), \qquad L_N := C_{\mathrm{st}} h_N^{-1} \log N$$

ensures $\tau_k^B - \sigma_{a,k} \le L_N$. The leader stays positive before $\sigma_{a,k}$ by the microscopic and transport estimates. Therefore, for $s_k + 1 \le t < \tau_k^B - L_N$, its pre-update state satisfies

$$\sum_i \lambda_i m_{ij_k^*}(X_{t-1})^p \le \lambda_+ a^p + (r-1)\lambda_+ K_*^p N^{-p/2} \le \theta/4 + o(1), \tag{463}$$

where (460) controls the other overlaps in its column.

Persistence for two windows after certification. Put $b_{\mathrm{rec}} = 1 - \varepsilon_{\mathrm{rec}}$ and

$$\Delta_N := 2C_{\mathrm{step}} h_N w_N \longrightarrow 0, \qquad \beta_N := K_* N^{-1/2} + \Delta_N \longrightarrow 0.$$

At $u = \tau_k^B$, certification gives $q_k(u) \geq b_{\mathrm{rec}}$, and (460) bounds every other overlap in the still-active columns. Applying (462), for $\tau_k^B < t \leq \tau_k^B + 2w_N$ we obtain

$$\sum_i \lambda_i m_{ij}(X_{t-1})^p \leq r\lambda_+ \beta_N^p = o(1), \qquad j \in J_{k-1} \setminus \{j_k^*\}, \tag{464}$$

$$\sum_i \lambda_i m_{ij_k^*}(X_{t-1})^p \geq \lambda_-(b_{\mathrm{rec}} - \Delta_N)^p - (r-1)\lambda_+ \beta_N^p = \lambda_- b_{\mathrm{rec}}^p - o(1) > \theta. \tag{465}$$

Only $o(1)$ overlap displacement is needed here; microscopic bounds need not persist after certification. The argument also covers the final stage.

Finally take

$$\eta := \frac{1}{8} \min\{\theta, \lambda_-(1 - \varepsilon_{\mathrm{rec}})^p - \theta\} > 0.$$

Equations (461), (463), (464), and (465) give all required threshold margins for large N. The pre-certification bounds start at $s_k + 1$, which is no later than the required $b_{k-1,N}$. The duration and horizon conditions were proved above. Also

$$\frac{L_N}{S_N} = \frac{C_{\mathrm{st}} \log N}{C_\delta N^{(p-2)/2}} \longrightarrow 0,$$

so $L_N + 4w_N < c_d S_N$ eventually, as needed by Proposition D.3. We have shown $\mathcal{P}_N \cap \mathcal{H}_N \subseteq \mathcal{D}_N$ for all sufficiently large N. The two uniform failure bounds prove (457). □

D.2 From stage margins to fixed-window reports

Use the durations $D_{k,N}$ from (11) and the stage starting indices $b_{k-1,N}$ defined in Appendix D.1. Recall that

$$L_N = C_{\mathrm{st}} h_N^{-1} \log N. \tag{466}$$

On $\mathcal{D}_N$, $c_d S_N \leq D_{k,N} \leq C_d S_N$ and $\tau_r^B + 2w_N \leq Q_N$. For each stage k, its leader's latent score is at most $\theta - \eta$ on $[b_{k-1,N}, \tau_k^B - L_N)$ and at least $\theta + \eta$ on $(\tau_k^B, \tau_k^B + 2w_N]$; all other columns in J_{k-1} have latent scores at most $\theta - \eta$ throughout $[b_{k-1,N}, \tau_k^B + 2w_N]$. For $k \geq 2$, the preceding stage's nonleader bound continues through $\tau_{k-1}^B + 2w_N = b_{k-1,N} - 1$, so the two bounds cover the handoff without an unguarded sample. The first stage begins at $b_{0,N} = 1$.

Lemma D.2 (A complete window after a stage time)**.** *For any integer $s \geq 0$ and integer $w \geq 1$, the window $\mathcal{W}_{\lfloor s/w \rfloor + 2}$ lies entirely in $(s, s + 2w]$ and ends no later than $s + 2w$.*

Proof. Write $s = aw + b$ with $0 \leq b < w$. The indicated window is $\{(a+1)w + 1, \ldots, (a+2)w\}$. Its first time exceeds s, and its endpoint satisfies $(a+2)w - s = 2w - b \leq 2w$. □

Role of Lemma D.2. It converts the post-certification score margin in $\mathcal{D}_N$ into a qualifying decision by $\tau_k^B + 2w_N$, for every alignment of the fixed grid.

Proposition D.3 (Pathwise fixed-window transfer)**.** *Fix a sufficiently large N such that $L_N + 4w_N < c_d S_N$. Let $T_{j,t}$ be scores evaluated using the pre-update states and current observations. Write $\mathfrak{Q}_N(T)$ for the pairs (j,t) actually queried by the monitor; reported columns and inactive windows need no score. Suppose $\mathcal{D}_N$ holds and*

$$\max_{(j,t) \in \mathfrak{Q}_N(T)} |T_{j,t} - F_{j,t}| < \eta/2. \tag{467}$$

Apply the fixed-window reporting rule to T, with no inactive windows, or with a deterministic inactive prefix ending at a grid endpoint $a_N < \tau_1^B - L_N$. Then all r reports occur by Q_N, and, for $1 \leq k \leq r$,

$$\widehat{j}_k = j_k^*, \qquad -L_N \leq \widehat{\tau}_k - \tau_k^B \leq 2w_N. \tag{468}$$

In particular,

$$\max_{k \leq r} \left| \frac{\widehat{\tau}_k - \widehat{\tau}_{k-1}}{\tau_k^B - \tau_{k-1}^B} - 1 \right| \leq \frac{2(L_N + 2w_N)}{c_d S_N}, \qquad \widehat{\tau}_0 = \tau_0^B = 0. \tag{469}$$

Proof. At every queried pair, the score event preserves both strict sides of the threshold: $F_{j,t} \le \theta - \eta$ implies $T_{j,t} < \theta$, and $F_{j,t} \ge \theta + \eta$ implies $T_{j,t} > \theta$. Thus a full window of high leader scores has all w_N indicators equal to one and qualifies for every $f \in (0, 1]$; a column with low scores throughout a window cannot qualify.

For $k = 1$, all nonleaders have low scores from $t = 1$ through $\tau_1^B + 2w_N$. The leader has low scores before $\tau_1^B - L_N$, so no report can occur earlier than that boundary. The inactive prefix, if present, ends before it. If the leader has not already been reported, Lemma D.2 provides a complete active window after τ_1^B and by $\tau_1^B + 2w_N$. Every leader score in it is high, every competitor score is low, and the rule reports j_1^* at that endpoint. A possible earlier report in $[\tau_1^B - L_N, \tau_1^B]$ can only be of the same leader.

Proceed by induction. Suppose the first $k-1$ reported columns are $j_1^*, \dots, j_{k-1}^*$ and their report times satisfy (468). After the previous report and until $\tau_{k-1}^B + 2w_N$, the preceding stage's nonleader margin keeps every still-unreported column below θ. From the next integer time $b_{k-1,N} = \tau_{k-1}^B + 2w_N + 1$, the stage-$k$ margin takes over. It keeps all competitors low through $\tau_k^B + 2w_N$, and keeps j_k^* low before $\tau_k^B - L_N$. The assumed lower stage duration gives $\tau_k^B - L_N > \tau_{k-1}^B + 4w_N$; in particular, these intervals have the stated order and contain any intervening completed windows. Thus the next report cannot precede $\tau_k^B - L_N$ or identify a competitor. If it has not occurred before certification, the complete post-stage window of Lemma D.2 reports j_k^* by $\tau_k^B + 2w_N$. This establishes the induction, including the final stage. Since $\tau_r^B + 2w_N \le Q_N$, no report used above is censored by the horizon.

Put $e_k = \widehat{\tau}_k - \tau_k^B$, $e_0 = 0$. The localization gives $|e_k| \le L_N + 2w_N$. Therefore

$$\left| \frac{\widehat{\tau}_k - \widehat{\tau}_{k-1}}{D_{k,N}} - 1 \right| = \frac{|e_k - e_{k-1}|}{D_{k,N}} \le \frac{2(L_N + 2w_N)}{c_d S_N},$$

as claimed. □

Role of Proposition D.3. This is the common deterministic step in the proofs of Theorem 3.2 and the numerical transfer theorem below. It uses the fixed grid, the one-report-per-window rule, the reset, and the two-stage handoff exactly as implemented by the algorithms.

Proof of Theorem 3.2. By Proposition D.1, $\inf_{x \in \mathcal{C}_N} \mathbb{P}_N^x(\mathcal{D}_N) \to 1$. The large-$N$ separation required by Proposition D.3 follows from (6) and $h_N w_N \to 0$:

$$\frac{h_N^{-1} \log N}{S_N} = \frac{\log N}{C_\delta N^{(p-2)/2}} \longrightarrow 0, \qquad \frac{w_N}{S_N} = \frac{h_N w_N}{C_\delta N^{(p-2)/2}} \longrightarrow 0. \tag{470}$$

For $T = U$, (31) and $\sqrt{\log N / N} \to 0$ give

$$\inf_{x \in \mathcal{C}_N} \mathbb{P}_N^x \left(\max_{j \le r,\, t \le Q_N} |U_{j,t} - F_{j,t}| < \eta/2 \right) \longrightarrow 1.$$

For $T = S$ with its prescribed admissible index, (32) gives the same conclusion with $S_{j,t}$ in place of $U_{j,t}$. For either fixed choice, intersect its score event with $\mathcal{D}_N$. The bound over all (j, t) implies (467) over the pairs actually queried. Proposition D.3, with $a_N = 0$, therefore gives all r reports, their order $(j_k^*)_{k=1}^r$, and

$$-L_N \le \widehat{\tau}_k^{\mathrm{fw}} - \tau_k^B \le 2w_N \qquad (1 \le k \le r).$$

The two applications use the same stage margins; they do not require the two monitors to cross the threshold or report at identical times.

For either choice, the pathwise localization proves (33) with $C = C_{\mathrm{st}}$. Equations (469) and (470) prove (34), and the same event supplies $\mathcal{E}_N^{\mathrm{fin}}$. All probability bounds are uniform in $x \in \mathcal{C}_N$. □

Role of Theorem 3.2. It combines the stage margins established in Proposition D.1 with uniform observable approximation to obtain the report order and inter-report clocks under the stated initialization, step-size, and window conditions.

D.3 Numerical Hessian scores and calibrated centers

Numerical calibration uses a different extension from the auxiliary Gaussian product in Appendix C. Let $\overline{\Omega}_N = \Omega_N \times \Omega_{\mathrm{num,N}}$, where the second factor carries independent null raw tensors for each potential query and any solver random seeds. Its law $\overline{\mathbb{P}}_N^x$ has original marginal $\mathbb{P}_N^x$. A query state is fixed before its null batch is revealed. The numerical filtration reveals these auxiliary variables only when their corresponding queries occur. Thus an adaptive query has fresh conditional null samples, and the numerical report times are stopping times in that filtration. The Gaussian comparison coordinates used to prove the spectral theorem are never inputs to the monitor.

Lemma D.4 (Uniform center from statewise null medians)**.** *Suppose at most $M_N^{\mathrm{qry}} \le rQ_N$ possibly adaptive null calibration queries are made and every query uses B_N^{cal} independent noise-only replicates with the raw law of the model. Write $\widehat{\zeta}_q^{0,\mathrm{exact}}$ for the median of the exact eigenvalues in query q. For every fixed $u > 0$, the exact null indexed eigenvalue satisfies*

$$q_N(u) := \sup_{X \in \mathrm{St}(N,r),\, j \le r} \mathbb{P}\{|h_{j,N}^0(X; W^0) + \rho_{p,j}| > u\} \longrightarrow 0. \tag{471}$$

Consequently, if $\log(M_N^{\mathrm{qry}} \vee 2) = o(B_N^{\mathrm{cal}})$, the exact empirical null medians obey

$$\max_{q \text{ queried}} |\widehat{\zeta}_q^{0,\mathrm{exact}} + \rho_{p,j_q}| \longrightarrow 0 \quad \textit{in } \overline{\mathbb{P}}_N^x\textit{-probability}, \tag{472}$$

uniformly in deterministic x; the maximum of an empty set is zero. A uniform error at most η_N^{eig} in every null replicate moves each empirical median by at most η_N^{eig}.

Proof. For deterministic X, the null compression (25) is precisely the noise matrix $A_{j,t}$ of (16), with an independent raw tensor. The conditional lower-bulk comparison and Lemma C.6 have bounds independent of the predictable state. Applied to a deterministic state, they give (471), including a fixed admissible ℓ_N.

Fix $u > 0$ and take N sufficiently large that $q_N(u) \le 1/4$. Conditionally on the past and the query state, at least three quarters of each batch's exact null eigenvalues lie in $[-\rho_{p,j_q} - u, -\rho_{p,j_q} + u]$ in expectation. Hoeffding's inequality bounds the probability that no strict majority lies there by $e^{-B_N^{\mathrm{cal}}/8}$. When a strict majority lies in an interval, the empirical median (including the average-of-two convention) lies in that interval. Taking conditional expectations and a union bound over at most M_N^{qry} queries gives failure probability at most $M_N^{\mathrm{qry}} e^{-B_N^{\mathrm{cal}}/8} = o(1)$, uniformly in the adaptive states. If all approximations of a finite sample are within η_N^{eig} of their exact values, order statistics and their middle-value average move by at most that amount. □

Role of Lemma D.4. It supplies the center consistency used by the numerical reporting theorem for independent-null centering. Its conclusion uses the uniform spectral edge and independent same-law null batches.

Theorem D.5 (Conditional numerical reporting clocks)**.** *Assume the conditions of Theorem 3.2. Also require the index conditions in* (17)*: $r + 1 \le \ell_N \le d_N$ and $\ell_N/d_N \to 0$ (so $\ell_N + r \le d_N$ eventually). Run Algorithm 2 with one fixed choice of $\widehat{S}^{\rho}$, $\widehat{S}^{\mathrm{null}}$, or $\widehat{S}^{\mathrm{burn}}$. Suppose the indexed eigensolver has uniform error $\eta_N^{\mathrm{eig}} \to 0$ over its successful calls, each call has conditional certificate failure probability at most ζ_N^{eig}, and*

$$M_N^{\mathrm{solve}} \zeta_N^{\mathrm{eig}} \longrightarrow 0, \quad M_N^{\mathrm{solve}} := \begin{cases} rQ_N, & \textit{analytic center}, \\ rQ_N(1 + B_N^{\mathrm{cal}}), & \textit{independent null center}, \\ r(Q_N + B_N^{\mathrm{burn}}), & \textit{burn-in center}. \end{cases} \tag{473}$$

For null centering, suppose $\log(rQ_N \vee 2) = o(B_N^{\mathrm{cal}})$ and the null samples have the independent same-law construction above. For burn-in centering, suppose (29) *holds, the prefix solves obey the same error bound, and*

$$\max_{i,j \le r} \max_{t \le B_N^{\mathrm{burn}}} |m_{ij,t-1}| \longrightarrow 0 \quad \textit{in probability, uniformly for } x \in \mathcal{C}_N. \tag{474}$$

Define the numerical finite-report event

$$\mathcal{E}_N^{\mathrm{num}} := \{0 = \tau_0^B < \cdots < \tau_r^B \le Q_N\} \cap \{0 = \widehat{\tau}_0 < \cdots < \widehat{\tau}_r \le Q_N\}. \tag{475}$$

Also write

$$\mathcal{L}_N^{\mathrm{num}} := \bigcap_{k=1}^{r} \{\widehat{j}_k = j_k^*,\ -L_N \le \widehat{\tau}_k - \tau_k^B \le 2w_N\},$$
$$\mathcal{I}_N^{\mathrm{num}}(\epsilon) := \left\{ \max_{k \le r} \left| \frac{\widehat{\tau}_k - \widehat{\tau}_{k-1}}{D_{k,N}} - 1 \right| \le \epsilon \right\}. \tag{476}$$

Then, for every $\epsilon > 0$, under the joint numerical law,

$$\inf_{x \in \mathcal{C}_N} \overline{\mathbb{P}}_N^x(\mathcal{E}_N^{\mathrm{num}} \cap \mathcal{L}_N^{\mathrm{num}} \cap \mathcal{I}_N^{\mathrm{num}}(\epsilon)) \longrightarrow 1. \tag{477}$$

Proof. Theorem 3.1 yields uniform exact spectral-score error $o_{\overline{\mathbb{P}}}(1)$ relative to $F_{j,t}$; the original trajectory has the same law under the numerical extension. A union bound over the at most M_N^{solve} solver calls and (473) gives a joint event on which all required indexed eigenvalues meet the prescribed error. Uniform convergence in probability permits deterministic $\varepsilon_N^{\mathrm{H}}, \eta_N^{\mathrm{cal}} \downarrow 0$ such that the respective exact spectral and center error bounds hold with probability tending to one.

The analytic center has zero calibration error. In null mode, Lemma D.4 and the null-replicate solver bounds make $\max_{(j,t)\in\mathfrak{Q}_N(\widehat{S})} |\widehat{\zeta}_{j,N}^0(X_{t-1}) + \rho_{p,j}| = o_{\overline{\mathbb{P}}}(1)$. In burn-in mode, the exact uniform Hessian approximation and (474) give $\max_{j,t\le B_N^{\mathrm{burn}}} |h_{j,t} + \rho_{p,j}| = o_{\overline{\mathbb{P}}}(1)$. Every numerical prefix value has the same additional $o(1)$ solver error. The sample median lies between its extremes, so its frozen center differs from $-\rho_{p,j}$ by $o_{\overline{\mathbb{P}}}(1)$.

For all three modes, adding and subtracting the exact indexed eigenvalue and its analytic center gives

$$\max_{(j,t)\in\mathfrak{Q}_N(\widehat{S})} |\widehat{S}_{j,t} - F_{j,t}| \le \varepsilon_N^{\mathrm{H}} + \frac{\eta_N^{\mathrm{eig}} + \eta_N^{\mathrm{cal}}}{p\lambda_-} = o_{\overline{\mathbb{P}}}(1). \tag{478}$$

In particular the score event (467) holds with probability tending to one. For burn-in, $a_N < B_N^{\mathrm{burn}} + w_N = o(S_N)$ by (29) and (470). On $\mathcal{D}_N$, $\tau_1^B \ge c_d S_N$, so $a_N < \tau_1^B - L_N$ eventually; for the other modes $a_N = 0$. Proposition D.1 gives $\inf_{x\in\mathcal{C}_N} \mathbb{P}_N^x(\mathcal{D}_N) \to 1$; this probability is unchanged under the numerical product law. Apply Proposition D.3 and then (469)–(470). All bounds are uniform in the allowed deterministic initialization. □

Microscopic calibration prefix. Under the stage assumptions, the prefix condition (474) follows from the path event $\mathcal{P}_N$ in the proof of Proposition D.1. On that event, $T_{\mathrm{sel},1} \ge \underline{c}_1 S_N$. Since $B_N^{\mathrm{burn}} = o(S_N)$, the calibration prefix lies before the first selection time for all sufficiently large N. The Regime-I guards therefore give

$$\max_{i,j\le r} \max_{1\le t\le B_N^{\mathrm{burn}}} |m_{ij}(X_{t-1})| \le \gamma_{+,1} N^{-1/2}.$$

The complement of $\mathcal{P}_N$ has probability tending to zero uniformly over $x \in \mathcal{C}_N$, and the numerical extension preserves the law of the original trajectory. Thus the prefix condition is automatic in this regime; it is displayed in the numerical theorem to make the calibration requirement explicit. The separate numerical-solver conditions remain as stated.

Role of Theorem D.5. It proves the asymptotic clock transfer for a correctly indexed numerical Hessian score, after including solver calls and the chosen center's actual calibration samples. It does not turn the finite-dimensional numerical experiments into tests of the certified-stage intervals.