# Beyond the Hessian Edge: The Stochastic Stability Cocycle of Mini-Batch SGD

**Anonymous Authors**[1]

## Abstract

We study the training-time stochastic stability cocycle of multi-pass mini-batch $\mathrm{SGD}$ in the Gaussian-design / affine-row-regularizer specialization of fixed-rank proportional-asymptotic multi-index models. Our main result is a finite-horizon replicated tangent $\mathrm{DMFT}$: by transporting finitely many common-noise tangent replicas through the same batch sequence, we obtain a closed deterministic kernel law for replica–replica, replica–iterate, and replica–teacher overlaps. This yields an exact deterministic law for the characteristic $q$-volume growth, defined by the log-determinant growth of the tangent Gram matrix and equivalently interpretable as exterior-power growth or accumulated QR growth of the Jacobian cocycle. Under an asymptotically stationary regime, the finite-time law converges to a top stationary exponent whose zero-crossing defines the stochastic edge. We also compare the Poissonian mini-batch theory with its Brownian $\mathrm{SGF}$ reference limit and derive a finite-horizon jump-cumulant correction for the characteristic $q$-volume law, which vanishes exactly in the affine-linear learner-side regime. In the rank-one logistic model, the tangent dynamics split into an explicit signal channel and an orthogonal bulk channel, yielding a closed one-dimensional critical equation for the stochastic edge.

## 1. Introduction

Training-time instability in modern optimization is fundamentally a *multiplicative* phenomenon. If the mini-batch SGD iterate is $\bar{\Theta}_m^{(d)}$ and the current batch is $B_m$, then infinitesimal perturbations are transported by the random Jacobian product

$$\Phi_m^{(d)} = \prod_{r=0}^{m-1} \Big( I - \eta_r \nabla^2 \widehat{L}_{B_r}(\bar{\Theta}_r^{(d)}) \Big), \tag{1}$$

not by a single Hessian snapshot. Existing high-dimensional theory already gives sharp asymptotic descriptions of the *iterate* dynamics of multi-pass mini-batch $\mathrm{SGD}$ in fixed-rank multi-index models (Fan & Wang, 2026), of the corresponding Brownian SGF/SME reference limit (Nishiyama & Imaizumi, 2026), and of characteristic Lyapunov exponents for local SGD stability near fixed minima (Chemnitz & Engel, 2025). What is still missing is an asymptotically exact high-dimensional law for the *training-time stochastic stability cocycle itself*.

This paper studies that missing object in the Gaussian-design / affine-row-regularizer specialization of the fixed-rank multi-index regime. The central move is to introduce finitely many *common-noise tangent replicas* transported by the same batch sequence. This converts the matrix product $\Phi_m^{(d)}$ into a replicated overlap process with a finite deterministic closure in the high-dimensional limit. The associated observable is the *characteristic $q$-volume growth*: the logarithmic growth of the $q \times q$ tangent Gram determinant. By the deterministic identities in Appendix B, this is equivalently the growth of $q$-dimensional tangent volume, the norm growth of the exterior-power cocycle $\bigwedge^q \Phi_m^{(d)}$, or the accumulated log-determinant of the local QR factors. Thus the paper replaces scalar sharpness heuristics by a canonical basis-invariant multiplicative object.

The first main result derives a finite-horizon *replicated tangent* $\mathrm{DMFT}$: a deterministic kernel tuple governing replica–replica, replica–iterate, and replica–teacher overlaps. The second main result converts that kernel theory into an exact deterministic law for the characteristic $q$-volume growth and, under an asymptotically stationary regime, into a top stationary exponent $\lambda_1$ whose zero-crossing defines the *stochastic edge*. The third main result compares this Poissonian mini-batch law with the Brownian $\mathrm{SGF}$ reference theory and identifies an explicit jump–cumulant correction at the multiplicative level; in the affine-linear learner-side regime, the correction vanishes exactly, matching the linear deterministic-equivalence picture (Atanasov et al., 2025). Finally, in the rank-one logistic model, the tangent dynam-

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

ics split into an explicit signal-aligned channel and an orthogonal bulk channel, yielding a one-dimensional critical equation for the stochastic edge.

From the proof side, the contribution is a full architecture rather than a single estimate: an augmented cavity expansion for common-noise tangent replicas, finite-horizon martingale and concentration bounds for the batch fluctuations, a contractive replicated Volterra kernel map, and deterministic tangent-geometry identities connecting Gram determinants, exterior powers, and QR growth. Section 2 defines the cocycle and the characteristic volume-growth observable. Section 3 states the general finite-horizon and stationary theorems. Section 4 gives the Poisson–Brownian comparison and the explicit rank-one logistic specialization. Section 5 summarizes the proof strategy, with complete arguments deferred to the appendices.

## 2. Problem Setting and Stochastic Stability Cocycle

### 2.1. Kinematic Setup

We work in the fixed-rank proportional-asymptotic multi-index regime underlying recent high-dimensional analyses of multi-pass SGD and its Brownian SGF/SME reference limit (Fan & Wang, 2026; Nishiyama & Imaizumi, 2026). The full technical assumptions are collected in Appendix A; the finite-horizon proof of the main theorem is carried out under the Gaussian-design / affine-row-regularizer specialization of Appendix F, which already contains the explicit nonlinear rank-one logistic theorem of Appendix J. The purpose of this section is to isolate the new object of the paper—the *training-time stochastic stability cocycle*—and the canonical multiplicative observable attached to it.

**Assumption 2.1** (Main high-dimensional regime)**.** The learner parameter is $\Theta \in \mathbb{R}^{d\times k}$, where the latent dimensions $k$ and $k_\star$ are fixed as $n, d \to \infty$. The aspect ratio satisfies $n/d \to \phi \in (0,\infty)$, and the data follow a fixed-rank multi-index model

$$Y_i = \sigma_\star(x_i^\top \Theta_\star^{(d)}, \varepsilon_i), \qquad i \in [n].$$

For a nonempty batch $B \subseteq [n]$, the batch risk has the row-separable form

$$\widehat{L}_B(\Theta) = \frac{1}{|B|}\sum_{i\in B} L\big(\sigma(x_i^\top\Theta), Y_i\big) + \frac{1}{n}\sum_{j=1}^{d} G(\Theta_j), \quad (2)$$

where $\Theta_j \in \mathbb{R}^k$ denotes the $j$-th row. The gradient and Hessian are assumed to exist and satisfy the bounded-derivative hypotheses stated in Appendix A. Mini-batches $B_m$ are i.i.d. uniform subsets of $[n]$ of cardinality $\kappa_n$, with

$$\kappa_n \sim \bar{\kappa} n^\alpha, \qquad \alpha \in [0,1),$$

and the learning-rate schedule satisfies

$$\eta_m = n^\alpha \bar{\eta}(mn^{\alpha-1}) + o(n^\alpha)$$

uniformly on compact epoch-time windows, for a bounded Lipschitz profile $\bar{\eta} : [0,\infty) \to (0,\infty)$.

The discrete mini-batch SGD trajectory is

$$\bar{\Theta}_{m+1}^{(d)} = \bar{\Theta}_m^{(d)} - \eta_m \nabla \widehat{L}_{B_m}\big(\bar{\Theta}_m^{(d)}\big), \qquad \bar{\Theta}_0^{(d)} = \Theta_0^{(d)}, \quad (3)$$

and we pass to epoch time by the embedding

$$\Theta_t^{(d)} := \bar{\Theta}_{\lfloor tn^{1-\alpha}\rfloor}^{(d)}, \qquad t \geq 0. \quad (4)$$

**Definition 2.2** (Training-time stochastic stability cocycle)**.** For each step $m$, define the linearized mini-batch update

$$J_m^{(d)}[U] := U - \eta_m \nabla^2 \widehat{L}_{B_m}\big(\bar{\Theta}_m^{(d)}\big)[U], \qquad U \in \mathbb{R}^{d\times k}. \quad (5)$$

The *training-time stochastic stability cocycle* is the random product of linear maps

$$\Phi_0^{(d)} := I, \qquad \Phi_m^{(d)} := J_{m-1}^{(d)} \circ \cdots \circ J_0^{(d)}, \qquad m \geq 1. \quad (6)$$

**Definition 2.3** (Common-noise tangent replicas)**.** Fix $q \in \mathbb{N}$ independently of $n, d$, and choose deterministic tangent initializations $U_0^{1,(d)}, \ldots, U_0^{q,(d)} \in \mathbb{R}^{d\times k}$ satisfying the nondegeneracy condition of Assumption A.5. The associated *common-noise tangent replicas* are the processes $\bar{U}_m^{a,(d)} \in \mathbb{R}^{d\times k}$, $a \in [q]$, defined by

$$\bar{U}_{m+1}^{a,(d)} = J_m^{(d)}\big[\bar{U}_m^{a,(d)}\big], \qquad a \in [q], \quad m \geq 0. \quad (7)$$

The point of using common-noise replicas is that they encode the multiplicative growth of the same stochastic Jacobian product seen from finitely many tangent directions, while still closing onto a finite collection of overlap kernels in high dimension.

**Definition 2.4** (Characteristic $q$-volume growth)**.** Let $m_t := \lfloor tn^{1-\alpha}\rfloor$. The empirical scalar tangent Gram matrix at epoch time $t$ is

$$\mathbf{G}_t^{(d,q)} := \left[\frac{1}{d}\left\langle \bar{U}_{m_t}^{a,(d)}, \bar{U}_{m_t}^{b,(d)}\right\rangle_{\mathrm{F}}\right]_{a,b=1}^{q} \in \mathbb{R}^{q\times q}. \quad (8)$$

Whenever $\mathbf{G}_0^{(d,q)}$ and $\mathbf{G}_t^{(d,q)}$ are positive definite, the *characteristic $q$-volume growth* is

$$\Lambda_{q,t}^{(d)} := \frac{1}{2t}\log\det\Big(\mathbf{G}_t^{(d,q)}\big(\mathbf{G}_0^{(d,q)}\big)^{-1}\Big). \quad (9)$$

By Appendix B, $\Lambda_{q,t}^{(d)}$ is equivalently: (i) the log-growth of a normalized tangent Gram determinant, (ii) the growth of the

$q$-dimensional tangent volume transported by the cocycle $\Phi^{(d)}_{m_t}$, (iii) the exterior-power growth

$$t^{-1}\log\frac{\|(\bigwedge^q \Phi^{(d)}_{m_t})\,\omega_q(\mathbf{U}^{(d)}_0)\|}{\|\omega_q(\mathbf{U}^{(d)}_0)\|},$$

or (iv) the accumulated log-determinant of the local QR factors of the tangent dynamics. When $q=1$, $\Lambda^{(d)}_{1,t}$ is the finite-time characteristic growth quantity whose stationary zero-crossing defines the *stochastic edge*; near a frozen minimum, it reduces to the local Lyapunov-type stability quantity studied by Chemnitz & Engel (2025).

## 3. Main Results

Unless stated otherwise, this section works under Assumptions 2.1 and A.5, together with the Gaussian-design / affine-row-regularizer specialization of Assumption F.1. This is the regime in which the replicated tangent closure is exact and the stochastic stability cocycle admits a deterministic high-dimensional law.

For $t,s\in[0,T]$, let $m_t := \lfloor tn^{1-\alpha}\rfloor$, and define the empirical replicated tangent kernels

$$Q^{ab,(d)}_{t,s} := \frac{1}{d}\sum_{j=1}^{d}\bar U^{a,(d)}_{j,m_t}\otimes \bar U^{b,(d)}_{j,m_s}, \qquad a,b\in[q], \tag{10}$$

$$S^{a,(d)}_{t,s} := \frac{1}{d}\sum_{j=1}^{d}\bar U^{a,(d)}_{j,m_t}\otimes \bar\Theta^{(d)}_{j,m_s}, \qquad a\in[q], \tag{11}$$

$$M^{a,(d)}_{t} := \frac{1}{d}\sum_{j=1}^{d}\bar U^{a,(d)}_{j,m_t}\otimes \Theta^{(d)}_{\star,j}, \qquad a\in[q]. \tag{12}$$

Write

$$\mathfrak{X}^{(q,d)}_T = \big(\mathbf{Q}^{(d)},\mathbf{S}^{(d)},\mathbf{M}^{(d)},\mathbf{Z}^{(d)}\big), \qquad \mathbf{Z}^{(d)} := \mathbf{Q}^{(d)},$$

for the corresponding empirical replicated kernel tuple. A deterministic tuple $\mathfrak{X}=(\mathbf{Q},\mathbf{S},\mathbf{M},\mathbf{Z})$ will be called *physical* if $\mathbf{Z}=\mathbf{Q}$ and the induced time-slice block covariance matrix is positive semidefinite; see Definition F.4.

**Theorem 3.1** (Finite-horizon replicated tangent DMFT)**.** *Fix $T>0$ and $q\ge 1$. Then there exists a unique physical replicated kernel tuple*

$$\mathfrak{X}^{(q)}_{T,\star} = \big(\mathbf{Q}_\star,\mathbf{S}_\star,\mathbf{M}_\star,\mathbf{Z}_\star\big)$$

*on $[0,T]$ such that, for every $\lambda\ge\lambda_\star(T,q)$,*

$$\mathbb{E}\Big[\|\mathfrak{X}^{(q,d)}_T-\mathfrak{X}^{(q)}_{T,\star}\|_{\lambda,T}\Big]\longrightarrow 0 \qquad \text{as } n,d\to\infty, \tag{13}$$

*where $\|\cdot\|_{\lambda,T}$ is the weighted supremum norm of Appendix E.*

*Equivalently,*

$$\max_{a,b\in[q]}\sup_{t,s\in[0,T]}\|Q^{ab,(d)}_{t,s}-Q^{ab}_{\star,t,s}\|_{\mathrm{F}}\xrightarrow[L^1]{}0, \tag{14}$$

$$\max_{a\in[q]}\sup_{t,s\in[0,T]}\|S^{a,(d)}_{t,s}-S^{a}_{\star,t,s}\|_{\mathrm{F}}\xrightarrow[L^1]{}0, \tag{15}$$

$$\max_{a\in[q]}\sup_{t\in[0,T]}\|M^{a,(d)}_{t}-M^{a}_{\star,t}\|_{\mathrm{F}}\xrightarrow[L^1]{}0. \tag{16}$$

*In particular, projected tangent fields driven by the same batch noise converge to the Gaussian lift determined by $\mathfrak{X}^{(q)}_{T,\star}$.*

Theorem 3.1 is the new object in the paper: an exact finite-horizon high-dimensional law for the *multiplicative tangent sector*, not merely for the iterate sector. The iterate kernels form one block of the closure; the new blocks are the replica–replica, replica–iterate, and replica–teacher kernels $(\mathbf{Q},\mathbf{S},\mathbf{M})$.

**Corollary 3.2** (Recovery of iterate-level effective dynamics)**.** *In the formal zero-replica limit $q=0$, the replicated tangent theory collapses exactly to the imported iterate-level multi-pass* SGD DMFT *of Fan & Wang (2026). Equivalently, the present theory is a strict extension of the iterate-only high-dimensional limit, not a competing one.*

The next theorem converts the kernel law into a deterministic law for the multiplicative observable introduced in Section 2. Define

$$\mathbf{G}^{(d,q)}_t = \Big[\frac{1}{d}\Big\langle \bar U^{a,(d)}_{m_t},\,\bar U^{b,(d)}_{m_t}\Big\rangle_{\mathrm{F}}\Big]_{a,b=1}^{q},$$
$$\mathbf{G}^{(q)}_{\star,t} = \Big[\mathrm{Tr}(Q^{ab}_{\star,t,t})\Big]_{a,b=1}^{q}.$$

**Theorem 3.3** (Deterministic law for characteristic $q$-volume growth)**.** *Let $0<\tau\le T<\infty$, and assume that the interval $[\tau,T]$ is $q$-volume admissible, i.e.*

$$\inf_{t\in[\tau,T]}\lambda_{\min}\big(\mathbf{G}^{(q)}_{\star,t}\big)>0.$$

*Define*

$$\Lambda^{(d)}_{q,t} := \frac{1}{2t}\log\det\Big(\mathbf{G}^{(d,q)}_t\big(\mathbf{G}^{(d,q)}_0\big)^{-1}\Big), \tag{17}$$

$$\Lambda_{q,t} := \frac{1}{2t}\log\det\Big(\mathbf{G}^{(q)}_{\star,t}\big(\mathbf{G}^{(q)}_{\star,0}\big)^{-1}\Big). \tag{18}$$

*Then there exists an event $\mathcal{A}^{(d)}_{\tau,T}$ with $\mathbb{P}(\mathcal{A}^{(d)}_{\tau,T})\to 1$ such that all empirical Gram matrices on $[\tau,T]$ are positive definite on $\mathcal{A}^{(d)}_{\tau,T}$, and*

$$\sup_{t\in[\tau,T]}\big|\Lambda^{(d)}_{q,t}-\Lambda_{q,t}\big|\xrightarrow[\mathbb{P}]{}0 \qquad \text{on } \mathcal{A}^{(d)}_{\tau,T}. \tag{19}$$

*By Appendix B, the same limit governs the exterior-power growth of $\bigwedge^q\Phi^{(d)}_{m_t}$ and the accumulated QR growth of the tangent cocycle.*

Theorem 3.3 identifies a deterministic limit for a genuinely multiplicative stability observable. This is stronger than tracking Hessian snapshots or scalar sharpness surrogates: the limiting object is the transported $q$-dimensional tangent volume itself.

**Theorem 3.4** (Stationary stochastic edge)**.** *Assume $q = 1$, and let $\mathbf{G}^{(1)}_{\star,t}$ be the limiting scalar Gram path from Theorem 3.3. Suppose that after some finite transient $t_{\mathrm{st}}$, the path is $C^1$, strictly positive, and the instantaneous log-growth rate*

$$\mathfrak{g}_1(t) := \frac{1}{2}\,\mathrm{Tr}\Big((\mathbf{G}^{(1)}_{\star,t})^{-1}\dot{\mathbf{G}}^{(1)}_{\star,t}\Big) \tag{20}$$

*admits a finite limit*

$$\lambda_1 := \lim_{t\to\infty} \mathfrak{g}_1(t).$$

*Then*

$$\Lambda_{1,t} \longrightarrow \lambda_1 \qquad \text{as } t \to \infty. \tag{21}$$

*Now let the regime depend on a control parameter $p \in [p_-, p_+]$, and suppose that $p \mapsto \lambda_1(p)$ is continuous, strictly monotone, and satisfies*

$$\lambda_1(p_-)\lambda_1(p_+) < 0.$$

*Then there exists a unique $p_c \in [p_-, p_+]$ such that*

$$\lambda_1(p_c) = 0. \tag{22}$$

*We call $p_c$ the* stochastic edge*. In the frozen-minimum regime, this top stationary exponent reduces exactly to the characteristic Lyapunov exponent of the local stochastic-gradient cocycle, consistent with the local theory of Chemnitz & Engel (2025).*

## 4. Jump–Diffusion Separation and Explicit Nonlinear Specialization

The general kernel theory admits two sharpenings. First, it can be compared with the Brownian SGF/SME reference limit of Nishiyama & Imaizumi (2026), yielding a *multiplicative* Poisson–Brownian correction that is invisible at the level of scalar iterate observables. Second, in the rank-one logistic model the tangent sector splits into an explicit signal channel and an orthogonal bulk channel, which makes the stochastic edge a one-dimensional deterministic equation.

**Theorem 4.1** (Jump–diffusion separation at the multiplicative level)**.** *Fix $0 < \tau \le T < T^\star_{\mathrm{SGF}}$, and let $\mathbf{G}^{(q),\mathrm{SGD}}_{\star,t}$ and $\mathbf{G}^{(q),\mathrm{SGF}}_{\star,t}$ denote the Poissonian and Brownian limiting scalar tangent Gram matrices. Assume both are positive definite on $[\tau, T]$, and define*

$$\begin{aligned}\mathcal{J}_{q,t} &:= \Lambda^{\mathrm{SGD}}_{q,t} - \Lambda^{\mathrm{SGF}}_{q,t} \\ &= \frac{1}{2t}\log\det\Big(\mathbf{G}^{(q),\mathrm{SGD}}_{\star,t}\big(\mathbf{G}^{(q),\mathrm{SGF}}_{\star,t}\big)^{-1}\Big), \\ &\qquad t \in [\tau, T].\end{aligned} \tag{23}$$

*If*

$$\gamma^{(q),\mathrm{com}}_{\tau,T} := \frac{1}{2}\min\Big\{\inf_{t\in[\tau,T]} \lambda_{\min}\big(\mathbf{G}^{(q),\mathrm{SGD}}_{\star,t}\big), \inf_{t\in[\tau,T]} \lambda_{\min}\big(\mathbf{G}^{(q),\mathrm{SGF}}_{\star,t}\big)\Big\} > 0.$$

*then*

$$|\mathcal{J}_{q,t}| \le \frac{\sqrt{q}}{2\tau\,\gamma^{(q),\mathrm{com}}_{\tau,T}}\big\|\mathbf{G}^{(q),\mathrm{SGD}}_{\star,t} - \mathbf{G}^{(q),\mathrm{SGF}}_{\star,t}\big\|_{\mathrm{F}}, \qquad t \in [\tau, T]. \tag{24}$$

*Moreover, there exists $C_{q,T} < \infty$ such that*

$$\sup_{t\in[\tau,T]} |\mathcal{J}_{q,t}| \le C_{q,T}\Bigg(\sup_{t\in[0,T]} \|\Delta C^{\vartheta,\star}_t\|_{\mathrm{F}} + \sup_{t,s\in[0,T]} \|\Delta C^{\vartheta,\vartheta}_{t,s}\|_{\mathrm{F}}\Bigg). \tag{25}$$

*where $\Delta C^{\vartheta,\star}$ and $\Delta C^{\vartheta,\vartheta}$ are the Poissonian-minus-Brownian iterate kernel differences. In the affine-linear learner-side regime, if these iterate kernels coincide, then $\mathcal{J}_{q,t} \equiv 0$ on $(0, T]$.*

The point of Theorem 4.1 is that Brownian surrogates may be accurate for iterate-level statistics and still miss the true multiplicative stability law. The discrepancy is controlled entirely by the mismatch of the two iterate effective environments, but it is detected only after transporting the tangent cocycle through a determinant-level observable.

We now specialize to the rank-one logistic model of Assumption A.7. Let

$$\begin{aligned}\ell(u; y) &= \log(1 + e^{-yu}), \\ \ell_r(u; y) &= \partial_u^r \ell(u; y), \qquad r = 1, 2, 3.\end{aligned}$$

let $W \sim \mathcal{N}(0, 1)$, and let $Y_\beta(W, \varepsilon) \in \{-1, 1\}$ be the logistic teacher label with inverse temperature $\beta$. Define

$$\begin{aligned}\chi_\perp(m) &:= \mathbb{E}[\ell_2(mW; Y_\beta)], \\ \tau_\beta(m) &:= \mathbb{E}[W\ell_3(mW; Y_\beta)], \\ \upsilon_\parallel(m) &:= \mathbb{E}[W^2\ell_2(mW; Y_\beta)^2], \\ \upsilon_\perp(m) &:= \mathbb{E}[\ell_2(mW; Y_\beta)^2].\end{aligned}$$

**Proposition 4.2** (Signal–bulk splitting in the rank-one logistic model)**.** *Assume the signal-aligned iterate regime $\xi_t = m_t W$, where $m_t$ is the scalar alignment path. Then the one-time limiting replica Gram entries admit the decomposition*

$$Q^{ab}_{\star,t,t} = \Sigma^{ab}_t + R^{ab}_t, \qquad a, b \in [q],\ t \ge 0, \tag{26}$$

*where the signal and bulk sectors solve the scalar linear systems*

$$\dot{\Sigma}_t^{ab} = \Gamma_{\parallel}(t)\Sigma_t^{ab},$$
$$\Gamma_{\parallel}(t) = -2\bar{\eta}(t)\Big[\lambda + \phi\chi_{\perp}(m_t) + \phi m_t \tau_\beta(m_t)\Big] + \bar{\eta}(t)^2\phi v_{\parallel}(m_t), \tag{27}$$
$$\dot{R}_t^{ab} = \Gamma_{\perp}(t)R_t^{ab},$$
$$\Gamma_{\perp}(t) = -2\bar{\eta}(t)\Big[\lambda + \phi\chi_{\perp}(m_t)\Big] + \bar{\eta}(t)^2\phi v_{\perp}(m_t). \tag{28}$$

*If $\Gamma_{\perp}(t) \le -c_{\perp} < 0$ for all $t \ge t_0$, then the orthogonal bulk contracts exponentially:*

$$\|R_t\|_{\mathrm{F}} \le e^{-c_{\perp}(t-t_0)}\|R_{t_0}\|_{\mathrm{F}}, \qquad t \ge t_0. \tag{29}$$

Proposition 4.2 shows that instability is carried by a finite-dimensional signal channel riding on top of a potentially contracting bulk sector. In particular, once the bulk is strictly stable, the stochastic edge is controlled by a scalar signal equation.

**Theorem 4.3** (Explicit critical surface for the stochastic edge)**.** *Assume constant step size $\bar{\eta}(t) \equiv \eta > 0$ and convergence of the iterate alignment path $m_t \to m_\infty$, where $m_\infty$ is the unique root of*

$$\lambda m + \phi\Psi_\beta(m) = 0, \qquad \Psi_\beta(m) := \mathbb{E}[W\ell_1(mW;Y_\beta)].$$

*For $q = 1$, define the asymptotic sector exponents*

$$\lambda_{\parallel}(\eta) = -\eta\Big[\lambda + \phi\chi_{\perp}(m_\infty) + \phi m_\infty \tau_\beta(m_\infty)\Big] + \frac{\eta^2}{2}\phi v_{\parallel}(m_\infty), \tag{30}$$
$$\lambda_{\perp}(\eta) = -\eta\Big[\lambda + \phi\chi_{\perp}(m_\infty)\Big] + \frac{\eta^2}{2}\phi v_{\perp}(m_\infty). \tag{31}$$

*If $\Sigma_0$ and $R_0$ denote the initial signal and bulk weights, then the top stationary exponent is*

$$\lambda_1(\eta) = \max\big\{\lambda_{\parallel}(\eta)\mathbf{1}_{\{\Sigma_0>0\}}, \lambda_{\perp}(\eta)\mathbf{1}_{\{R_0>0\}}\big\}. \tag{32}$$

*Hence the unique positive stochastic edge is the smallest active sector root:*

$$\eta_c = \min\left\{\begin{matrix}\eta_{\parallel,c}\mathbf{1}_{\{\Sigma_0>0\}} + \infty\,\mathbf{1}_{\{\Sigma_0=0\}},\\ \eta_{\perp,c}\mathbf{1}_{\{R_0>0\}} + \infty\,\mathbf{1}_{\{R_0=0\}}\end{matrix}\right\}. \tag{33}$$

*where*

$$\Lambda_{\parallel,\infty} := \lambda + \phi\chi_{\perp}(m_\infty) + \phi m_\infty\tau_\beta(m_\infty),$$
$$\Lambda_{\perp,\infty} := \lambda + \phi\chi_{\perp}(m_\infty),$$
$$\eta_{\parallel,c} = \frac{2\Lambda_{\parallel,\infty}}{\phi v_{\parallel}(m_\infty)},$$
$$\eta_{\perp,c} = \frac{2\Lambda_{\perp,\infty}}{\phi v_{\perp}(m_\infty)}.$$

*Equivalently, $\eta_c$ is the unique positive solution of the scalar edge equation*

$$\max\big\{\lambda_{\parallel}(\eta)\mathbf{1}_{\{\Sigma_0>0\}}, \lambda_{\perp}(\eta)\mathbf{1}_{\{R_0>0\}}\big\} = 0. \tag{34}$$

**Corollary 4.4** (Exact reductions and consistency checks)**.** *The theory satisfies three exact reductions.*

1. *In the formal zero-replica limit $q = 0$, it reduces exactly to the iterate-level multi-pass* SGD DMFT *of Fan & Wang (2026).*
2. *In the affine-linear learner-side regime, the replicated tangent fixed point is a closed linear Volterra/Lyapunov system, matching the linear deterministic-equivalence picture of Atanasov et al. (2025); the Poissonian and Brownian multiplicative laws then coincide whenever their iterate kernels do.*
3. *In the frozen-minimum regime, $\lambda_1$ is exactly the characteristic Lyapunov exponent of the local stochastic-gradient cocycle after the discrete-step/epoch-time normalization (Chemnitz & Engel, 2025).*

## 5. Proof Roadmap and Technical Contributions

The proof centers on a causal Volterra map on replicated kernels. The iterate-only multi-pass SGD theory supplies the iterate block (Fan & Wang, 2026); the new step is to lift it to common-noise Hessian-transported tangent replicas. The new algebraic term is the bilinear cavity operator $\mathcal{B}^{a,(-j)}$, coupling the current learner coordinate to the cavity tangent field.

**Proposition 5.1** (Contractive replicated kernel map)**.** *Under Theorem 3.1's hypotheses, the augmented-cavity coefficient family is admissible in the sense of Appendix E. Thus, for every finite horizon $T > 0$ and $q \ge 1$, there are $R < \infty$ and $\lambda_\star \ge 1$ such that, whenever $\lambda \ge \lambda_\star$, the replicated Volterra map $\mathcal{T}_T^{(q)}$ is a strict contraction on $\mathfrak{B}_{R,\lambda,T}^{(q),+}$. Its unique physical fixed point is the limiting replicated kernel tuple.*

The proof of Theorem 3.1 has four steps. Appendix C derives row-wise cavity recursions, isolating $\mathcal{A}^{(-j)}$, $\mathcal{B}^{a,(-j)}$, and the backreaction errors. Appendix D proves row envelopes and martingale bounds, and Appendix F shows the backreaction is uniformly $o(1)$ on compact epoch intervals. The same appendix then converts the empirical kernels into

$$\mathfrak{X}_T^{(q,d)} = \mathcal{T}_T^{(q)}\big(\mathfrak{X}_T^{(q,d)}\big) + \mathfrak{R}_T^{(q,d)}, \qquad \|\mathfrak{R}_T^{(q,d)}\|_{\lambda,T} \to 0,$$

so Proposition 5.1 gives convergence to the deterministic fixed point. Finally, Appendices B–J supply the geometry, $q$-volume law, stationary edge, Poisson–Brownian comparison, and rank-one logistic specialization.

## Impact Statement

This paper is a theoretical study of high-dimensional learning dynamics. It does not introduce a new deployed system, dataset, or application pipeline. Its direct contribution is a mathematical framework for analyzing the stochastic stability cocycle of mini-batch SGD, together with exact asymptotic laws for characteristic tangent-volume growth, Poisson–Brownian separation, and the stochastic edge in a class of solvable high-dimensional models.

The most plausible positive impact of this work is scientific. A sharper understanding of when stochastic optimization is stable, metastable, or unstable can improve the reliability and interpretability of training procedures for large models. In particular, a theory that distinguishes iterate-level behavior from multiplicative stability behavior may help researchers reason more carefully about continuous-time approximations, optimizer tuning, and the role of batch stochasticity. Such understanding may also reduce wasted computation by clarifying regimes in which training dynamics are provably inefficient or structurally unstable.

The main potential negative impact is indirect. Better theoretical control of large-scale optimization may eventually make it easier to train or tune more capable machine learning systems, including systems later deployed in domains with nontrivial societal risk. In addition, asymptotic stability criteria can be misapplied if they are treated as universal prescriptions outside the assumptions under which they are proved. The present results are derived in a restricted high-dimensional regime and, in the strongest theorem, under a Gaussian-design / affine-row-regularizer specialization with an explicit rank-one logistic model. They should therefore be interpreted as foundational theory, not as a general safety certification or a drop-in recipe for deployment.

Overall, we view the immediate impact of this work as advancing the theoretical foundations of optimization and learning dynamics. Any downstream operational use should be accompanied by assumption checking, finite-dimensional validation, and application-specific risk assessment.

## References

Atanasov, A., Bordelon, B., Zavatone-Veth, J. A., Paquette, C., and Pehlevan, C. Two-point deterministic equivalence for stochastic gradient dynamics in linear models. *arXiv preprint arXiv:2502.05074*, 2025. doi: 10.48550/arXiv.2502.05074. URL https://arxiv.org/abs/2502.05074. In press at Advances in Theoretical and Mathematical Physics.

Chemnitz, D. and Engel, M. Characterizing dynamical stability of stochastic gradient descent in overparameterized learning. *Journal of Machine Learning Research*, 26 (134):1–46, 2025. URL https://www.jmlr.org/papers/v26/24-1547.html.

Fan, Z. and Wang, L. High-dimensional learning dynamics of multi-pass stochastic gradient descent in multi-index models. *arXiv preprint arXiv:2601.21093*, 2026. doi: 10.48550/arXiv.2601.21093. URL https://arxiv.org/abs/2601.21093.

Nishiyama, S. and Imaizumi, M. High-dimensional limit of stochastic gradient flow via dynamical mean-field theory. *arXiv preprint arXiv:2602.06320*, 2026. doi: 10.48550/arXiv.2602.06320. URL https://arxiv.org/abs/2602.06320.

## A. Full Assumptions and Imported Background

This appendix fixes the exact asymptotic regime used throughout the paper and records the iterate-level effective theories that will be imported as black boxes later on. The standing assumptions are chosen so that they remain fully compatible with the finite-horizon high-dimensional DMFT for multi-pass mini-batch SGD in fixed-rank multi-index models, while strengthening the regularity envelope just enough to support the Hessian-valued linearization underlying the stochastic stability cocycle (Fan & Wang, 2026). For the Brownian comparison of Appendix I, we will also import the corresponding finite-horizon SGF /SME description (Nishiyama & Imaizumi, 2026).

### A.1. Global Notation and Standing Conventions

From this appendix onward we specialize the parameter space $\mathcal{H}_d$ from Section 2 to the matrix space $\mathbb{R}^{d\times k}$, equipped with the Frobenius inner product and norm. For a matrix $A \in \mathbb{R}^{d\times m}$, its $j$-th row is denoted by $A_j \in \mathbb{R}^m$, so that

$$A = [A_1, \dots, A_d]^\top, \qquad \|A\|_{\mathrm{F}}^2 = \sum_{j=1}^{d} \|A_j\|_2^2 .$$

The teacher parameter has rank $k_\star$ and is denoted by $\Theta_\star^{(d)} \in \mathbb{R}^{d\times k_\star}$, while the learner parameter has rank $k$ and is denoted by $\Theta \in \mathbb{R}^{d\times k}$. For $q \geq 1$ fixed and matrices $A^1, \dots, A^q \in \mathbb{R}^{d\times m}$, we write

$$\mu^{(d)}_{A^{1:q}} := \frac{1}{d}\sum_{j=1}^{d} \delta_{(A_j^1,\dots,A_j^q)}$$

for the empirical law of the rows. When auxiliary objects are present, we extend this notation in the obvious way, e.g.

$$\mu^{(d)}_{A^{1:q},B,C} := \frac{1}{d}\sum_{j=1}^{d} \delta_{(A_j^1,\dots,A_j^q,B_j,C_j)}.$$

For each $n$, let

$$[n] := \{1, \dots, n\}, \qquad \mathfrak{S}_{\kappa_n} := \{B \subseteq [n] : |B| = \kappa_n\}$$

denote the family of batches of cardinality $\kappa_n$. We write the discrete mini-batch SGD iterates as $(\bar{\Theta}_m^{(d)})_{m\geq 0}$ and the associated discrete tangent replicas as $(\bar{U}_m^{a,(d)})_{m\geq 0}$, $a \in [q]$, to distinguish them from the epoch-time embeddings

$$\Theta_t^{(d)} := \bar{\Theta}^{(d)}_{\lfloor tn^{1-\alpha}\rfloor}, \qquad U_t^{a,(d)} := \bar{U}^{a,(d)}_{\lfloor tn^{1-\alpha}\rfloor}, \qquad t \geq 0.$$

All limits $n, d \to \infty$ are taken jointly.

**Definition A.1** (Second-order pseudo-Lipschitz test functions)**.** Let $m \geq 1$ and $E = \mathbb{R}^m$ endowed with the Euclidean norm. A measurable function $\psi : E \to \mathbb{R}$ is called *second-order pseudo-Lipschitz* if there exists a constant $C_\psi < \infty$ such that

$$|\psi(z) - \psi(z')| \leq C_\psi\big(1 + \|z\|_2 + \|z'\|_2\big)\,\|z - z'\|_2 \qquad \text{for all } z, z' \in E.$$

We write $\psi \in \mathrm{PL}_2(E)$ in this case.

**Conditioning convention.** To avoid unnecessary bookkeeping, we state all asymptotic assumptions for deterministic sequences $\Theta_0^{(d)}$, $\Theta_\star^{(d)}$, $\varepsilon^{(n)}$, and $U_0^{1:q,(d)}$, and all almost-sure convergence statements below are with respect to the randomness of the covariates and the mini-batch sequence, conditional on these deterministic objects. The same statements therefore apply verbatim whenever these objects are random but independent of the covariates, by conditioning (Fan & Wang, 2026; Nishiyama & Imaizumi, 2026).

### A.2. General High-Dimensional Regime

We now state the standing assumptions for the general nonlinear theory.

**Assumption A.2** (Mini-batch sampling, critical stepsize scaling, and epoch time)**.** There exist constants $\alpha \in [0, 1)$ and $\bar{\kappa} \in (0, \infty)$, together with a Lipschitz function $\bar{\eta} : [0, \infty) \to (0, \infty)$, such that the following hold.

1. The batch sizes satisfy
$$\lim_{n,d\to\infty} \frac{\kappa_n}{n^\alpha} = \bar{\kappa}.$$
2. The learning-rate schedule $(\eta_m^{(n,d)})_{m\geq 0}$ obeys, for every fixed $T > 0$,
$$\lim_{n,d\to\infty} \sup_{t\in[0,T]} \left| \frac{\eta^{(n,d)}_{\lfloor tn^{1-\alpha}\rfloor}}{n^\alpha} - \bar{\eta}(t) \right| = 0.$$
3. Conditional on the training data, the mini-batches $B_0, B_1, \ldots$ are i.i.d. and uniformly distributed over $\mathfrak{S}_{\kappa_n}$.
4. The discrete SGD iteration is
$$\bar{\Theta}^{(d)}_{m+1} = \bar{\Theta}^{(d)}_m - \eta^{(n,d)}_m \nabla \widehat{L}_{B_m}(\bar{\Theta}^{(d)}_m), \qquad \bar{\Theta}^{(d)}_0 = \Theta^{(d)}_0,$$
and its epoch-time embedding is $\Theta^{(d)}_t = \bar{\Theta}^{(d)}_{\lfloor tn^{1-\alpha}\rfloor}$.
5. For each fixed $q$, the discrete tangent replicas evolve by
$$\bar{U}^{a,(d)}_{m+1} = \Big(I - \eta^{(n,d)}_m \nabla^2 \widehat{L}_{B_m}(\bar{\Theta}^{(d)}_m)\Big)\bar{U}^{a,(d)}_m, \qquad a \in [q],$$
with epoch-time embedding $U^{a,(d)}_t = \bar{U}^{a,(d)}_{\lfloor tn^{1-\alpha}\rfloor}$.

**Assumption A.3** (Finite-rank multi-index model and proportional asymptotics)**.** There exist fixed integers $k \geq 1$ and $k_\star \geq 1$, a constant $\phi \in (0, \infty)$, and deterministic sequences
$$\Theta^{(d)}_\star \in \mathbb{R}^{d\times k_\star}, \qquad \Theta^{(d)}_0 \in \mathbb{R}^{d\times k}, \qquad \varepsilon^{(n)} = (\varepsilon^{(n)}_1, \ldots, \varepsilon^{(n)}_n) \in \mathbb{R}^n,$$
such that the following hold.

1. $n/d \to \phi$ as $n, d \to \infty$.
2. The covariates $x_1, \ldots, x_n \in \mathbb{R}^d$ are i.i.d., and each $x_i = (x_{i1}, \ldots, x_{id})$ has independent coordinates satisfying
$$\mathbb{E}[x_{ij}] = 0, \qquad \mathbb{E}[x_{ij}^2] = d^{-1}, \qquad \sup_{d\geq 1} \sup_{1\leq j\leq d} \mathbb{E}\exp\big(c_x d x_{ij}^2\big) < \infty$$
for some constant $c_x > 0$.
3. There exists a measurable teacher channel $\sigma_\star : \mathbb{R}^{k_\star} \times \mathbb{R} \to \mathbb{R}$ such that the labels are generated by
$$Y_i = \sigma_\star\big(x_i^\top \Theta^{(d)}_\star,\, \varepsilon^{(n)}_i\big), \qquad i \in [n].$$
4. The empirical law of the row pairs converges weakly and in $W_p$ for every fixed $p \geq 1$:
$$\mu^{(d)}_{0,\star} := \frac{1}{d}\sum_{j=1}^d \delta_{(\Theta^{(d)}_{0,j},\, \Theta^{(d)}_{\star,j})} \Longrightarrow \mu_{0,\star},$$
where $\mu_{0,\star}$ is a probability measure on $\mathbb{R}^k \times \mathbb{R}^{k_\star}$ having finite exponential moments in a neighborhood of the origin.
5. The empirical law of the label-noise variables converges weakly and in $W_p$ for every fixed $p \geq 1$:
$$\mu^{(n)}_\varepsilon := \frac{1}{n}\sum_{i=1}^n \delta_{\varepsilon^{(n)}_i} \Longrightarrow \mu_\varepsilon,$$
where $\mu_\varepsilon$ is a probability measure on $\mathbb{R}$ having finite exponential moments in a neighborhood of the origin.

**Assumption A.4** (Loss, activation, regularizer, and Hessian action)**.** There exist functions

$$\sigma : \mathbb{R}^k \to \mathbb{R}, \qquad L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \qquad G : \mathbb{R}^k \to \mathbb{R}$$

such that the empirical batch risk is

$$\widehat{L}_B(\Theta) = \frac{1}{|B|} \sum_{i \in B} L\big(\sigma(x_i^\top \Theta),\, Y_i\big) + \frac{1}{n} \sum_{j=1}^{d} G(\Theta_j). \tag{35}$$

Define

$$f(\xi, w_\star, \varepsilon) := \partial_1 L\big(\sigma(\xi), \sigma_\star(w_\star, \varepsilon)\big) \nabla \sigma(\xi) \in \mathbb{R}^k, \qquad g(u) := \nabla G(u) \in \mathbb{R}^k. \tag{36}$$

We assume the following.

1. The map $f : \mathbb{R}^k \times \mathbb{R}^{k_\star} \times \mathbb{R} \to \mathbb{R}^k$ belongs to $C^3$ in its first argument $\xi$, and each derivative $D_1^r f$, $r = 0, 1, 2, 3$, is globally bounded:
$$\sup_{(\xi, w_\star, \varepsilon) \in \mathbb{R}^k \times \mathbb{R}^{k_\star} \times \mathbb{R}} \|D_1^r f(\xi, w_\star, \varepsilon)\|_{\mathrm{op}} < \infty, \qquad r = 0, 1, 2, 3.$$

2. The regularizer gradient $g : \mathbb{R}^k \to \mathbb{R}^k$ belongs to $C^3$, its derivatives of orders $1, 2, 3$ are globally bounded, and $g$ has at most linear growth:
$$\|g(u)\|_2 \le C_g(1 + \|u\|_2) \qquad \text{for all } u \in \mathbb{R}^k.$$

3. The gradient of the batch risk has the row-factorized form
$$\nabla \widehat{L}_B(\Theta) = \frac{1}{|B|} \sum_{i \in B} x_i \otimes f\big(x_i^\top \Theta, x_i^\top \Theta_\star^{(d)}, \varepsilon_i^{(n)}\big) + \frac{1}{n} g(\Theta), \tag{37}$$
where $g(\Theta) \in \mathbb{R}^{d \times k}$ is understood row-wise:
$$g(\Theta) := [g(\Theta_1), \ldots, g(\Theta_d)]^\top.$$

4. The Fréchet Hessian of $\widehat{L}_B$ acts on a tangent matrix $U \in \mathbb{R}^{d \times k}$ by
$$\nabla^2 \widehat{L}_B(\Theta)[U] = \frac{1}{|B|} \sum_{i \in B} x_i \otimes \Big(D_1 f\big(x_i^\top \Theta, x_i^\top \Theta_\star^{(d)}, \varepsilon_i^{(n)}\big)[x_i^\top U]\Big) + \frac{1}{n} Dg(\Theta)[U], \tag{38}$$
where $Dg(\Theta)[U] \in \mathbb{R}^{d \times k}$ is the row-wise linearization
$$Dg(\Theta)[U] := [Dg(\Theta_1)U_1, \ldots, Dg(\Theta_d)U_d]^\top.$$

**Assumption A.5** (Tangent initialization and nondegenerate initial $q$-volume)**.** Fix $q \ge 1$ independently of $n, d$. For each $a \in [q]$, let $U_0^{a,(d)} \in \mathbb{R}^{d \times k}$ be a deterministic tangent initialization. Assume that the joint row empirical law

$$\nu_{0,\star,q}^{(d)} := \frac{1}{d} \sum_{j=1}^{d} \delta_{\big(U_{0,j}^{1,(d)}, \ldots, U_{0,j}^{q,(d)}, \Theta_{0,j}^{(d)}, \Theta_{\star,j}^{(d)}\big)}$$

converges weakly and in $W_p$ for every fixed $p \ge 1$ to a probability measure $\nu_{0,\star,q}$ on $(\mathbb{R}^k)^q \times \mathbb{R}^k \times \mathbb{R}^{k_\star}$ having finite exponential moments in a neighborhood of the origin. Moreover, the limiting initial tangent Gram matrix

$$\mathbf{Q}_0^{(q)} := \Big[\int \langle u^a,\, u^b \rangle\, \nu_{0,\star,q}(du^1, \ldots, du^q, d\vartheta_0, d\vartheta_\star)\Big]_{a,b=1}^{q} \tag{39}$$

is strictly positive definite.

*Remark* A.6 (Relation to the iterate-level literature). Assumptions A.2–A.4 are a notationally compatible and slightly strengthened version of the standing hypotheses used for the iterate-level high-dimensional multi-pass SGD DMFT of Fan & Wang (2026) and the SGF DMFT of Nishiyama & Imaizumi (2026). Relative to those works, the only substantive strengthening is that we assume one additional bounded derivative in the learner-side nonlinearity, because the present paper studies the Hessian-valued linearization rather than the iterate dynamics alone. Assumption A.5 is entirely new and exists solely to make the tangent-volume observable of Definition 2.4 mathematically well posed at time 0.

### A.3. Model Specialization for the Explicit Rank-One Theorem

The explicit nonlinear theorem of Section 4 will be proved under the following rank-one specialization.

**Assumption A.7** (Rank-one logistic specialization)**.** For the results in Appendix J and Theorem 4.3, we impose the following additional assumptions.

1. $k = k_\star = 1$, and we write
$$\theta_\star^{(d)} \in \mathbb{R}^d, \qquad \theta_0^{(d)} \in \mathbb{R}^d, \qquad u_0^{a,(d)} \in \mathbb{R}^d, \quad a \in [q],$$
for the teacher, learner initialization, and tangent initial states, respectively.

2. The covariates are exactly Gaussian:
$$x_i \sim \mathcal{N}(0, I_d/d) \qquad \text{i.i.d. for } i \in [n].$$

3. The teacher channel is binary logistic with inverse temperature $\beta > 0$: there exist i.i.d. $\varepsilon_i^{(n)} \sim \mathrm{Unif}[0,1]$ such that
$$Y_i = \sigma_\star(x_i^\top \theta_\star^{(d)}, \varepsilon_i^{(n)}) := 2\mathbf{1}\left\{\varepsilon_i^{(n)} \le \frac{1}{1 + e^{-\beta x_i^\top \theta_\star^{(d)}}}\right\} - 1.$$

4. The learner uses the identity activation, logistic loss, and ridge regularization:
$$\sigma(\xi) = \xi, \qquad L(\hat{y}, y) = \log(1 + e^{-y\hat{y}}), \qquad G(u) = \frac{\lambda}{2}u^2 \quad (\lambda > 0).$$

5. The signal is normalized and the learner has nontrivial initial alignment:
$$\frac{1}{d}\left\|\theta_\star^{(d)}\right\|_2^2 \to 1, \qquad \frac{1}{d}\left\langle \theta_0^{(d)}, \theta_\star^{(d)} \right\rangle \to m_0 \quad \text{for some } m_0 \neq 0.$$

6. For every fixed $q$, the tangent initializations satisfy Assumption A.5, and the corresponding $\mathbf{Q}_0^{(q)}$ is strictly positive definite.

*Remark* A.8 (Why the rank-one logistic specialization is technically clean). Under Assumption A.7, the learner-side nonlinearity satisfies the regularity requirements of Assumption A.4: for logistic loss with identity activation, the function $f(\xi, w_\star, \varepsilon)$ is scalar-valued, bounded, and has bounded derivatives of every order in $\xi$, while the ridge term gives the linear regularizer gradient $g(u) = \lambda u$. The Gaussian design furthermore induces an exact signal–orthogonal decomposition for the tangent dynamics that will later allow us to isolate a single signal-carrying cocycle from a strictly contracting bulk sector in Appendix J.

### A.4. Imported Iterate-Level Effective Theory

We now record the two iterate-level results that will be imported later as black boxes. The first is the finite-horizon DMFT for multi-pass mini-batch SGD in the fixed-rank multi-index setting; the second is the corresponding Brownian SGF reference limit. Neither statement contains any tangent-replica, Jacobian-cocycle, or multiplicative volume-growth information.

**Proposition A.9** (Imported finite-horizon iterate-level DMFT)**.** *Assume Assumptions A.2–A.4. Then for every finite horizon $T > 0$ there exists a unique admissible deterministic kernel tuple*
$$\mathfrak{K}_T^{\mathrm{it}} = \left(C^\vartheta,\, R^\vartheta,\, C^f,\, R^f,\, R^{f,\star},\, \Gamma\right)$$
*and associated effective processes*
$$(\vartheta_t)_{t\in[0,T]} \in \mathbb{R}^k, \qquad (\xi_t)_{t\in[0,T]} \in \mathbb{R}^k, \qquad \vartheta_\star \in \mathbb{R}^{k_\star}, \qquad w_\star \in \mathbb{R}^{k_\star},$$
*such that the following hold for the epoch-embedded* SGD *iterate* $(\Theta_t^{(d)})_{t\in[0,T]}$.

*For every $m \geq 1$, every finite time grid $0 \leq t_1 \leq \cdots \leq t_m \leq T$, every $\psi \in \mathrm{PL}_2((\mathbb{R}^k)^m \times \mathbb{R}^{k_\star})$, and every $\varphi \in \mathrm{PL}_2((\mathbb{R}^k)^m \times \mathbb{R}^{k_\star} \times \mathbb{R})$,*

$$\frac{1}{d}\sum_{j=1}^{d}\psi\big(\Theta^{(d)}_{j,t_1},\ldots,\Theta^{(d)}_{j,t_m},\Theta^{(d)}_{\star,j}\big) \xrightarrow[n,d\to\infty]{\text{a.s.}} \mathbb{E}\big[\psi\big(\vartheta_{t_1},\ldots,\vartheta_{t_m},\vartheta_\star\big)\big]\,, \tag{40}$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi\big(x_i^\top\Theta^{(d)}_{t_1},\ldots,x_i^\top\Theta^{(d)}_{t_m},x_i^\top\Theta^{(d)}_{\star},\varepsilon^{(n)}_i\big) \xrightarrow[n,d\to\infty]{\text{a.s.}} \mathbb{E}\big[\varphi\big(\xi_{t_1},\ldots,\xi_{t_m},w_\star,\varepsilon\big)\big]\,. \tag{41}$$

*Equivalently, the corresponding empirical measures converge weakly and in $W_2$ on the relevant finite-dimensional state spaces. Moreover, the limiting law depends on the batch-size exponent $\alpha$ only through $(\bar\kappa,\bar\eta)$; in particular, once $\bar\kappa$ and $\bar\eta$ are fixed, the high-dimensional limit is identical for all $\alpha \in [0,1)$.*

*This is the finite-horizon multi-pass* SGD DMFT *of Fan & Wang (2026, Theorems 2.5–2.6 and Remarks 2.7–2.8), restated in notation compatible with the present paper.*

*Remark* A.10 (Role of Proposition A.9). Proposition A.9 is the only iterate-level statement imported from the multi-pass SGD literature. In the later tangent theory, its kernel tuple $\mathfrak{K}^{\mathrm{it}}_T$ will reappear as the iterate sub-block of a strictly larger replicated kernel system, but no multiplicative information is borrowed from prior work.

**Proposition A.11** (Imported Brownian reference limit for SGF)**.** *Assume Assumptions A.2–A.4, and consider the Brownian reference process*

$$\begin{aligned} d\Theta^{\mathrm{SGF}}_t &= -\bar\eta(t)\left[\sum_{i=1}^{n} x_i \otimes f\big(x_i^\top\Theta^{\mathrm{SGF}}_t, x_i^\top\Theta^{(d)}_\star, \varepsilon^{(n)}_i\big) + g(\Theta^{\mathrm{SGF}}_t)\right]dt \\ &\quad + \frac{\bar\eta(t)}{\sqrt{\bar\kappa}}\sum_{i=1}^{n}\Big[x_i \otimes f\big(x_i^\top\Theta^{\mathrm{SGF}}_t, x_i^\top\Theta^{(d)}_\star, \varepsilon^{(n)}_i\big)\Big]\,dB_i(t), \qquad \Theta^{\mathrm{SGF}}_0 = \Theta^{(d)}_0, \end{aligned} \tag{42}$$

*where $B_1,\ldots,B_n$ are independent standard Brownian motions. Then there exists a deterministic horizon $T^\star_{\mathrm{SGF}} \in (0,\infty]$ such that for every $T < T^\star_{\mathrm{SGF}}$ there exists a unique admissible Brownian kernel tuple and associated effective processes $(\vartheta^{\mathrm{SGF}}_t, \xi^{\mathrm{SGF}}_t)_{t\in[0,T]}$ for which the analogues of (40)–(41) hold with convergence in probability in $W_2$. In the linear case, one may take $T^\star_{\mathrm{SGF}} = \infty$.*

*This is the finite-horizon* SGF /SME DMFT *characterization of Nishiyama & Imaizumi (2026, Theorems 3.1–3.2), again rewritten in notation compatible with the present paper.*

*Remark* A.12 (External consistency targets). Later exact reductions will be formulated so as to recover two benchmark results in their native regimes: the characteristic Lyapunov exponent governing dynamical stability of SGD near a fixed minimum (Chemnitz & Engel, 2025), and the two-point deterministic equivalence for high-dimensional linear SGD (Atanasov et al., 2025). These results are not used as black boxes for the nonlinear cocycle theory developed in this paper; they serve as exact consistency checks in Appendix K.

## B. Deterministic Tangent-Geometry Identities

This appendix is completely deterministic. Its sole purpose is to isolate the linear-algebraic and geometric identities that underlie Definitions 2.2–2.4. In particular, we show that the characteristic $q$-volume growth is a basis-invariant $q$-dimensional volume-growth observable, that it admits an exact reformulation on the exterior-power space, and that it can be computed through a sequential QR factorization of the tangent dynamics. No stochastic assumption enters this appendix.

Throughout, $\mathcal{H}$ denotes a finite-dimensional real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. We fix $q \in \mathbb{N}$. For a $q$-tuple $\mathbf{u} = (u^1,\ldots,u^q) \in \mathcal{H}^q$, we write

$$\mathrm{span}(\mathbf{u}) := \mathrm{span}\{u^1,\ldots,u^q\} \subseteq \mathcal{H}.$$

For $A \in \mathbb{R}^{q\times q}$, we use the shorthand

$$\mathbf{u}A := (v^1,\ldots,v^q), \qquad v^a := \sum_{b=1}^{q} u^b A_{ba}, \qquad a \in [q].$$

Thus, $\mathbf{u}A$ is the $q$-tuple obtained by taking linear combinations of the vectors in $\mathbf{u}$ with coefficient matrix $A$.

### B.1. Gram Determinants and $q$-Dimensional Volume

**Definition B.1** (Synthesis operator and Gram matrix)**.** For $\mathbf{u} = (u^1, \dots, u^q) \in \mathcal{H}^q$, define the synthesis operator

$$\mathcal{U}_{\mathbf{u}} : \mathbb{R}^q \to \mathcal{H}, \qquad \mathcal{U}_{\mathbf{u}} c := \sum_{a=1}^{q} c_a u^a.$$

Its associated Gram matrix is

$$\mathbf{G}(\mathbf{u}) := \mathcal{U}_{\mathbf{u}}^* \mathcal{U}_{\mathbf{u}} = \left[\langle u^a, \, u^b \rangle\right]_{a,b=1}^{q} \in \mathbb{R}^{q \times q}. \tag{43}$$

Whenever $\mathbf{G}(\mathbf{u})$ is positive definite, we define the $q$-dimensional volume of $\mathbf{u}$ by

$$\mathrm{vol}_q(\mathbf{u}) := \sqrt{\det \mathbf{G}(\mathbf{u})}. \tag{44}$$

If $\mathbf{G}(\mathbf{u})$ is singular, we set $\mathrm{vol}_q(\mathbf{u}) := 0$.

**Lemma B.2** (Positivity and linear independence)**.** *For every $\mathbf{u} \in \mathcal{H}^q$, the Gram matrix $\mathbf{G}(\mathbf{u})$ is symmetric positive semidefinite. Moreover, the following are equivalent:*

1. $\mathbf{G}(\mathbf{u})$ *is positive definite;*
2. $u^1, \dots, u^q$ *are linearly independent;*
3. $\dim \mathrm{span}(\mathbf{u}) = q$*;*
4. $\mathrm{vol}_q(\mathbf{u}) > 0$*.*

*Proof.* Symmetry is immediate from the definition. For $c = (c_1, \dots, c_q)^\top \in \mathbb{R}^q$,

$$c^\top \mathbf{G}(\mathbf{u}) c = \sum_{a,b=1}^{q} c_a c_b \, \langle u^a, \, u^b \rangle = \left\langle \sum_{a=1}^{q} c_a u^a, \, \sum_{b=1}^{q} c_b u^b \right\rangle = \left\| \sum_{a=1}^{q} c_a u^a \right\|^2 \geq 0.$$

Hence $\mathbf{G}(\mathbf{u})$ is positive semidefinite. Equality $c^\top \mathbf{G}(\mathbf{u}) c = 0$ holds if and only if $\sum_{a=1}^{q} c_a u^a = 0$. Therefore, $\mathbf{G}(\mathbf{u})$ is positive definite if and only if the family $(u^1, \dots, u^q)$ is linearly independent, which is equivalent to $\dim \mathrm{span}(\mathbf{u}) = q$. The final equivalence follows because a symmetric positive semidefinite matrix has strictly positive determinant if and only if it is positive definite. $\square$

**Lemma B.3** (Change of basis inside the tangent span)**.** *Let $\mathbf{u} \in \mathcal{H}^q$ and $A \in \mathbb{R}^{q \times q}$. Then*

$$\mathbf{G}(\mathbf{u}A) = A^\top \mathbf{G}(\mathbf{u}) A. \tag{45}$$

*Consequently,*

$$\det \mathbf{G}(\mathbf{u}A) = \det(A)^2 \det \mathbf{G}(\mathbf{u}). \tag{46}$$

*In particular, if $A$ is orthogonal, then*

$$\mathbf{G}(\mathbf{u}A) = A^\top \mathbf{G}(\mathbf{u}) A, \qquad \mathrm{vol}_q(\mathbf{u}A) = \mathrm{vol}_q(\mathbf{u}). \tag{47}$$

*Proof.* For $c \in \mathbb{R}^q$,

$$\mathcal{U}_{\mathbf{u}A} c = \sum_{a=1}^{q} c_a \sum_{b=1}^{q} u^b A_{ba} = \sum_{b=1}^{q} (Ac)_b u^b = \mathcal{U}_{\mathbf{u}}(Ac).$$

Hence $\mathcal{U}_{\mathbf{u}A} = \mathcal{U}_{\mathbf{u}} A$. Therefore,

$$\mathbf{G}(\mathbf{u}A) = \mathcal{U}_{\mathbf{u}A}^* \mathcal{U}_{\mathbf{u}A} = A^\top \mathcal{U}_{\mathbf{u}}^* \mathcal{U}_{\mathbf{u}} A = A^\top \mathbf{G}(\mathbf{u}) A,$$

which proves (45). Taking determinants yields (46). If $A$ is orthogonal, then $|\det A| = 1$, so $\det \mathbf{G}(\mathbf{u}A) = \det \mathbf{G}(\mathbf{u})$, and hence $\mathrm{vol}_q(\mathbf{u}A) = \mathrm{vol}_q(\mathbf{u})$. $\square$

**Corollary B.4** (Normalization invariance of finite-time volume-growth ratios)**.** *Let $\mathbf{G}_0, \mathbf{G}_M \in \mathbb{R}^{q\times q}$ be positive definite, and let $c > 0$ be a scalar. Then*

$$\det\Big((c\mathbf{G}_M)(c\mathbf{G}_0)^{-1}\Big) = \det\Big(\mathbf{G}_M\mathbf{G}_0^{-1}\Big). \tag{48}$$

*Consequently, any finite-time volume-growth observable defined through a ratio of the form $\det(\mathbf{G}_M\mathbf{G}_0^{-1})$ is unchanged by multiplying all Gram matrices by the same positive scalar.*

*Proof.* Since $(c\mathbf{G}_0)^{-1} = c^{-1}\mathbf{G}_0^{-1}$,

$$(c\mathbf{G}_M)(c\mathbf{G}_0)^{-1} = (c\mathbf{G}_M)(c^{-1}\mathbf{G}_0^{-1}) = \mathbf{G}_M\mathbf{G}_0^{-1}.$$

Taking determinants proves the claim. $\square$

**Corollary B.5** (Equivalence of ambient-dimension and row-wise normalizations)**.** *In the matrix setting of Appendix A, where $\mathcal{H}_d = \mathbb{R}^{d\times k}$ and $p_d = \dim(\mathcal{H}_d) = dk$, define*

$$\mathbf{Q}_m^{(d,q)} := \left[\frac{1}{p_d}\left\langle U_m^{a,(d)}, U_m^{b,(d)}\right\rangle_{\mathrm{F}}\right]_{a,b=1}^{q}$$

*as in Definition 2.4, and define the row-normalized Gram matrix*

$$\widetilde{\mathbf{Q}}_m^{(d,q)} := \left[\frac{1}{d}\left\langle U_m^{a,(d)}, U_m^{b,(d)}\right\rangle_{\mathrm{F}}\right]_{a,b=1}^{q}.$$

*Then*

$$\widetilde{\mathbf{Q}}_m^{(d,q)} = k\,\mathbf{Q}_m^{(d,q)} \qquad \textit{for every } m, \tag{49}$$

*and therefore, whenever the endpoint Gram matrices are positive definite,*

$$\det\Big(\widetilde{\mathbf{Q}}_M^{(d,q)}\big(\widetilde{\mathbf{Q}}_0^{(d,q)}\big)^{-1}\Big) = \det\Big(\mathbf{Q}_M^{(d,q)}\big(\mathbf{Q}_0^{(d,q)}\big)^{-1}\Big). \tag{50}$$

*Proof.* Because $p_d = dk$,

$$\frac{1}{d}\left\langle U_m^{a,(d)}, U_m^{b,(d)}\right\rangle_{\mathrm{F}} = k\cdot\frac{1}{p_d}\left\langle U_m^{a,(d)}, U_m^{b,(d)}\right\rangle_{\mathrm{F}},$$

which proves (49). Then (50) follows from Corollary B.4 with $c = k$. $\square$

**Lemma B.6** (Rank monotonicity under linear transport)**.** *Let $T_0, \ldots, T_{M-1} : \mathcal{H} \to \mathcal{H}$ be linear maps, and define*

$$u_{m+1}^a := T_m u_m^a, \qquad a \in [q],\ m = 0, \ldots, M-1.$$

*Then the dimensions $\dim\operatorname{span}(u_m^1, \ldots, u_m^q)$ form a nonincreasing sequence in $m$. Equivalently, if $\mathbf{G}(\mathbf{u}_m)$ is singular for some $m$, then $\mathbf{G}(\mathbf{u}_{m'})$ is singular for all $m' \geq m$. In particular, if $\mathbf{G}(\mathbf{u}_M)$ is positive definite, then $\mathbf{G}(\mathbf{u}_m)$ is positive definite for every $m \leq M$.*

*Proof.* Let $V_m := \operatorname{span}(u_m^1, \ldots, u_m^q)$. Since each $u_{m+1}^a = T_m u_m^a$,

$$V_{m+1} = \operatorname{span}(T_m u_m^1, \ldots, T_m u_m^q) \subseteq T_m(V_m).$$

Hence

$$\dim V_{m+1} \leq \dim T_m(V_m) \leq \dim V_m.$$

Thus $m \mapsto \dim V_m$ is nonincreasing. By Lemma B.2, $\mathbf{G}(\mathbf{u}_m)$ is singular if and only if $\dim V_m < q$. The final statement follows immediately. $\square$

### B.2. Wedge-Product and QR Reformulations

We now identify the same $q$-volume observable in two equivalent ways: first as the norm of a wedge product in the exterior-power space, and second as the product of local QR factors along the tangent trajectory.

**Definition B.7** (Exterior-power representative)**.** For $\mathbf{u} = (u^1, \dots, u^q) \in \mathcal{H}^q$, define

$$\omega_q(\mathbf{u}) := u^1 \wedge \cdots \wedge u^q \in \bigwedge^q \mathcal{H}.$$

The space $\bigwedge^q \mathcal{H}$ is endowed with its canonical Hilbert structure, i.e. the one for which, given any orthonormal basis $(e_1, \dots, e_p)$ of $\mathcal{H}$, the wedge basis

$$\{e_{i_1} \wedge \cdots \wedge e_{i_q} : 1 \le i_1 < \cdots < i_q \le p\}$$

is orthonormal.

**Proposition B.8** (Gram determinant equals squared wedge norm)**.** *For every* $\mathbf{u} = (u^1, \dots, u^q) \in \mathcal{H}^q$*,*

$$\|\omega_q(\mathbf{u})\|^2 = \det \mathbf{G}(\mathbf{u}). \tag{51}$$

*Equivalently,*

$$\mathrm{vol}_q(\mathbf{u}) = \|\omega_q(\mathbf{u})\|. \tag{52}$$

*Proof.* Let $p := \dim(\mathcal{H})$, and choose an orthonormal basis $(e_1, \dots, e_p)$ of $\mathcal{H}$. Write

$$u^a = \sum_{i=1}^{p} U_{ia} e_i, \qquad a \in [q],$$

and let $U \in \mathbb{R}^{p \times q}$ be the coordinate matrix whose $a$-th column is $(U_{1a}, \dots, U_{pa})^\top$. Then

$$\mathbf{G}(\mathbf{u}) = U^\top U.$$

By the Cauchy–Binet formula,

$$\det(U^\top U) = \sum_{I \subseteq [p],\, |I|=q} \det(U_I)^2, \tag{53}$$

where $U_I$ denotes the $q \times q$ submatrix obtained by restricting $U$ to the rows indexed by $I$.

On the other hand, multilinearity and antisymmetry of the wedge product give

$$u^1 \wedge \cdots \wedge u^q = \sum_{I = \{i_1 < \cdots < i_q\} \subseteq [p]} \det(U_I)\, e_{i_1} \wedge \cdots \wedge e_{i_q}.$$

Since the wedge basis is orthonormal by construction,

$$\left\|u^1 \wedge \cdots \wedge u^q\right\|^2 = \sum_{I \subseteq [p],\, |I|=q} \det(U_I)^2.$$

Comparing with (53) yields (51). Taking square roots proves (52). ☐

**Corollary B.9** (Exterior-power representation of the cocycle growth)**.** *Let $T_0, \dots, T_{M-1} : \mathcal{H} \to \mathcal{H}$ be linear maps, and define*

$$\Phi_M := T_{M-1} \cdots T_1 T_0.$$

*Let* $\mathbf{u}_0 = (u_0^1, \dots, u_0^q) \in \mathcal{H}^q$*, and propagate it by*

$$u_{m+1}^a = T_m u_m^a, \qquad a \in [q], \quad m = 0, \dots, M-1.$$

*Then*

$$\omega_q(\mathbf{u}_M) = \bigwedge^q \Phi_M\, \omega_q(\mathbf{u}_0). \tag{54}$$

*If* $\mathbf{G}(\mathbf{u}_0)$ *and* $\mathbf{G}(\mathbf{u}_M)$ *are positive definite, then*

$$\frac{1}{2M} \log\det\Big(\mathbf{G}(\mathbf{u}_M)\mathbf{G}(\mathbf{u}_0)^{-1}\Big) = \frac{1}{M} \log \frac{\|(\bigwedge^q \Phi_M)\omega_q(\mathbf{u}_0)\|}{\|\omega_q(\mathbf{u}_0)\|}. \tag{55}$$

*In the matrix setting of Definition 2.4, the same identity holds with* $\mathbf{G}$ *replaced by* $\mathbf{Q}^{(d,q)}$ *or by* $\widetilde{\mathbf{Q}}^{(d,q)}$*, thanks to Corollary B.5.*

*Proof.* The identity (54) is the defining functoriality of the exterior power:

$$\bigwedge^q \Phi_M(u_0^1 \wedge \cdots \wedge u_0^q) = \Phi_M u_0^1 \wedge \cdots \wedge \Phi_M u_0^q = u_M^1 \wedge \cdots \wedge u_M^q.$$

By Proposition B.8,

$$\det \mathbf{G}(\mathbf{u}_M) = \|\omega_q(\mathbf{u}_M)\|^2, \qquad \det \mathbf{G}(\mathbf{u}_0) = \|\omega_q(\mathbf{u}_0)\|^2.$$

Therefore,

$$\frac{1}{2M} \log\det\Big(\mathbf{G}(\mathbf{u}_M)\mathbf{G}(\mathbf{u}_0)^{-1}\Big) = \frac{1}{2M} \log \frac{\det \mathbf{G}(\mathbf{u}_M)}{\det \mathbf{G}(\mathbf{u}_0)} = \frac{1}{M} \log \frac{\|\omega_q(\mathbf{u}_M)\|}{\|\omega_q(\mathbf{u}_0)\|}.$$

Using (54) proves (55). The final sentence follows from Corollary B.5. ☐

**Definition B.10** (QR factorization of a $q$-tuple)**.** Let $\mathbf{u} = (u^1, \dots, u^q) \in \mathcal{H}^q$ with $\dim \operatorname{span}(\mathbf{u}) = q$. A *QR factorization* of $\mathbf{u}$ means a representation

$$\mathbf{u} = \mathbf{v}R, \tag{56}$$

where $\mathbf{v} = (v^1, \dots, v^q) \in \mathcal{H}^q$ is an orthonormal $q$-tuple, i.e.

$$\langle v^a, v^b \rangle = \delta_{ab} \qquad \text{for all } a, b \in [q],$$

and $R \in \mathbb{R}^{q\times q}$ is upper triangular with strictly positive diagonal.

*Remark* B.11. A QR factorization in the sense of Definition B.10 always exists whenever $\mathbf{G}(\mathbf{u})$ is positive definite. Indeed, after choosing an orthonormal basis of $\mathcal{H}$, it reduces to the standard thin QR factorization of a full-column-rank matrix, with the positive diagonal condition fixing the sign ambiguity. We will use this fact repeatedly without further comment.

**Proposition B.12** (Sequential QR factorization of $q$-volume growth)**.** *Let $T_0, \dots, T_{M-1} : \mathcal{H} \to \mathcal{H}$ be linear maps, and let $\mathbf{u}_0 = (u_0^1, \dots, u_0^q) \in \mathcal{H}^q$ satisfy $\mathbf{G}(\mathbf{u}_0) \succ 0$. Define recursively*

$$u_{m+1}^a := T_m u_m^a, \qquad a \in [q], \quad m = 0, \dots, M-1.$$

*Assume that $\mathbf{G}(\mathbf{u}_M) \succ 0$. Then, by Lemma B.6, every intermediate Gram matrix is positive definite, so each $\mathbf{u}_m$ admits a QR factorization*

$$\mathbf{u}_m = \mathbf{v}_m S_m, \qquad m = 0, \dots, M,$$

*with $\mathbf{v}_m$ orthonormal and $S_m$ upper triangular with strictly positive diagonal. Moreover, there exist unique upper triangular matrices $R_0, \dots, R_{M-1} \in \mathbb{R}^{q\times q}$ with strictly positive diagonal such that*

$$T_m \mathbf{v}_m = \mathbf{v}_{m+1} R_m, \qquad m = 0, \dots, M-1, \tag{57}$$

*and the following identities hold:*

$$S_M = R_{M-1} \cdots R_1 R_0 S_0, \tag{58}$$

$$\det \mathbf{G}(\mathbf{u}_M) = \det \mathbf{G}(\mathbf{u}_0) \prod_{m=0}^{M-1} \det(R_m)^2, \tag{59}$$

$$\frac{1}{2M} \log\det\Big(\mathbf{G}(\mathbf{u}_M)\mathbf{G}(\mathbf{u}_0)^{-1}\Big) = \frac{1}{M} \sum_{m=0}^{M-1} \log\det(R_m). \tag{60}$$

*In the matrix setting of Definition 2.4, the same identity holds with $\mathbf{G}$ replaced by $\mathbf{Q}^{(d,q)}$ or by $\widetilde{\mathbf{Q}}^{(d,q)}$.*

*Proof.* By the assumption $\mathbf{G}(\mathbf{u}_M) \succ 0$ and Lemma B.6, we have $\mathbf{G}(\mathbf{u}_m) \succ 0$ for all $m \le M$. Hence each $\mathbf{u}_m$ admits a QR factorization $\mathbf{u}_m = \mathbf{v}_m S_m$ in the sense of Definition B.10, with $S_m$ upper triangular and positive diagonal.

Now fix $m \in \{0, \dots, M-1\}$. Since $\mathbf{v}_m$ is orthonormal and $\mathbf{u}_{m+1} = T_m \mathbf{u}_m$ has full rank, the $q$-tuple $T_m \mathbf{v}_m$ also has full rank. Thus $T_m \mathbf{v}_m$ admits a unique QR factorization

$$T_m \mathbf{v}_m = \mathbf{v}_{m+1} R_m,$$

with $\mathbf{v}_{m+1}$ orthonormal and $R_m$ upper triangular with strictly positive diagonal. This proves (57).

Using $\mathbf{u}_m = \mathbf{v}_m S_m$ and (57),

$$\mathbf{u}_{m+1} = T_m \mathbf{u}_m = T_m(\mathbf{v}_m S_m) = (T_m \mathbf{v}_m) S_m = \mathbf{v}_{m+1} R_m S_m.$$

By uniqueness of QR factorization with positive diagonal, we must have

$$S_{m+1} = R_m S_m.$$

Iterating this identity yields (58).

Since $\mathbf{v}_m$ is orthonormal, we have

$$\mathbf{G}(\mathbf{u}_m) = S_m^\top \mathbf{G}(\mathbf{v}_m) S_m = S_m^\top S_m,$$

and therefore

$$\det \mathbf{G}(\mathbf{u}_m) = \det(S_m)^2.$$

Applying this at times $0$ and $M$, then using (58), gives

$$\det \mathbf{G}(\mathbf{u}_M) = \det(S_M)^2 = \det(S_0)^2 \prod_{m=0}^{M-1} \det(R_m)^2 = \det \mathbf{G}(\mathbf{u}_0) \prod_{m=0}^{M-1} \det(R_m)^2,$$

which is (59). Taking logarithms and dividing by $2M$ proves (60). The final sentence follows from Corollary B.5. □

**Corollary B.13** (The case $q = 1$)**.** *Let $q = 1$. Under the assumptions of Proposition B.12, write $v_m$ for the orthonormal frame vector at step $m$, so that $\|v_m\| = 1$, and let $r_m > 0$ be the scalar defined by*

$$T_m v_m = v_{m+1} r_m.$$

*Then*

$$\frac{1}{2M} \log \frac{\|u_M\|^2}{\|u_0\|^2} = \frac{1}{M} \sum_{m=0}^{M-1} \log r_m = \frac{1}{M} \sum_{m=0}^{M-1} \log \|T_m v_m\|. \tag{61}$$

*Thus, when $q = 1$, the characteristic volume-growth observable reduces exactly to the mean log-growth of the orthonormalized one-step tangent magnitudes.*

*Proof.* For $q = 1$, the matrices $S_m$ and $R_m$ in Proposition B.12 are positive scalars. Writing $R_m = r_m$ gives

$$T_m v_m = v_{m+1} r_m, \qquad r_m = \|T_m v_m\|,$$

because $\|v_{m+1}\| = 1$. Identity (61) is then the specialization of (60) to $q = 1$. □

*Remark* B.14 (Interpretation for the stochastic stability cocycle)*.* Appendix B shows that the characteristic growth quantity in Definition 2.4 admits three exactly equivalent deterministic interpretations:

1. as the log-growth of a normalized Gram determinant;
2. as the log-growth of a $q$-dimensional tangent volume in $\mathcal{H}_d$;
3. as the log-norm growth of the exterior-power cocycle $\bigwedge^q \Phi_M$, or equivalently as the accumulated log-determinant of the local QR factors.

The point of the later probabilistic analysis is therefore not to invent a stability statistic, but to derive an asymptotically exact high-dimensional law for a canonical multiplicative object.

# C. Augmented Cavity Expansion

This appendix derives the exact row-wise cavity decomposition that will drive the proof of Theorem 3.1. The iterate-only version of this decomposition underlies the imported finite-horizon DMFT of Proposition A.9; the present section augments that algebra by adjoining $q$ common-noise tangent replicas and tracking the additional Hessian-induced bilinear couplings that they create. The identities in this appendix are exact at finite $n, d$: no limit transition, concentration argument, or stochastic approximation is used here. Compare with the iterate-level high-dimensional multi-pass SGD DMFT of Fan & Wang (2026).

For readability, we suppress the superscript $(d)$ on the discrete iterates, tangent replicas, and teacher parameter throughout this appendix. Thus

$$\bar{\Theta}_m \equiv \bar{\Theta}_m^{(d)} \in \mathbb{R}^{d\times k}, \qquad \bar{U}_m^a \equiv \bar{U}_m^{a,(d)} \in \mathbb{R}^{d\times k}, \qquad \Theta_\star \equiv \Theta_\star^{(d)} \in \mathbb{R}^{d\times k_\star},$$

and likewise $\eta_m \equiv \eta_m^{(n,d)}$.

## C.1. Representative-Coordinate Expansion

**Coordinate-removal projector.** Fix $j \in [d]$. For vectors $x \in \mathbb{R}^d$, define

$$P_j^\perp x := x - x_j e_j,$$

where $e_j \in \mathbb{R}^d$ is the $j$-th Euclidean basis vector. For a matrix $A \in \mathbb{R}^{d\times m}$, define $P_j^\perp A \in \mathbb{R}^{d\times m}$ row-wise by

$$(P_j^\perp A)_\ell = \begin{cases} 0, & \ell = j, \\ A_\ell, & \ell \neq j. \end{cases}$$

We write

$$x_i^{(-j)} := P_j^\perp x_i \in \mathbb{R}^d, \qquad A^{(-j)} := P_j^\perp A \in \mathbb{R}^{d\times m}.$$

**Definition C.1** (Coordinate-$j$ cavity iterate)**.** Fix $j \in [d]$. The *coordinate-$j$ cavity iterate* is the sequence $(\bar{\Theta}_m^{(-j)})_{m\geq 0} \subseteq \mathbb{R}^{d\times k}$ defined by

$$\bar{\Theta}_0^{(-j)} := P_j^\perp \Theta_0, \tag{62}$$

and, for all $m \geq 0$,

$$\bar{\Theta}_{m+1}^{(-j)} = P_j^\perp \left[ \bar{\Theta}_m^{(-j)} - \eta_m \left\{ \frac{1}{\kappa_n} \sum_{i\in B_m} x_i^{(-j)} \otimes f\big((x_i^{(-j)})^\top \bar{\Theta}_m^{(-j)},\, W_i^\star,\, \varepsilon_i\big) + \frac{1}{n} g\big(\bar{\Theta}_m^{(-j)}\big) \right\} \right], \tag{63}$$

where

$$W_i^\star := x_i^\top \Theta_\star \in \mathbb{R}^{k_\star}. \tag{64}$$

**Definition C.2** (Coordinate-$j$ cavity tangent replicas)**.** Fix $j \in [d]$ and $q \geq 1$. For each $a \in [q]$, the *coordinate-$j$ cavity tangent replica* is the sequence $(\bar{U}_m^{a,(-j)})_{m\geq 0} \subseteq \mathbb{R}^{d\times k}$ defined by

$$\bar{U}_0^{a,(-j)} := P_j^\perp U_0^a, \tag{65}$$

and, for all $m \geq 0$,

$$\begin{aligned} \bar{U}_{m+1}^{a,(-j)} = P_j^\perp \Bigg[ \bar{U}_m^{a,(-j)} - \eta_m \Bigg\{ &\frac{1}{\kappa_n} \sum_{i\in B_m} x_i^{(-j)} \otimes \\ &D_1 f\big((x_i^{(-j)})^\top \bar{\Theta}_m^{(-j)},\, W_i^\star,\, \varepsilon_i\big) \left[(x_i^{(-j)})^\top \bar{U}_m^{a,(-j)}\right] \\ &+ \frac{1}{n} Dg\big(\bar{\Theta}_m^{(-j)}\big) \left[\bar{U}_m^{a,(-j)}\right] \Bigg\} \Bigg]. \end{aligned} \tag{66}$$

*Remark* C.3 (Why the full teacher field is retained). The cavity system removes the learner coordinate $j$ from the dynamics but retains the full teacher projection $W_i^\star = x_i^\top \Theta_\star$. This is deliberate. Our standing hypotheses in Appendix A assume smoothness of $f$ only in its *first* argument, corresponding to the learner projection. No differentiability in the teacher argument is assumed, so removing the teacher coordinate and Taylor-expanding in $W_i^\star$ would introduce an artificial smoothness requirement that is neither needed nor justified.

**Definition C.4** (Cavity projections and back-reaction errors)**.** For each $i \in [n]$, $m \geq 0$, $a \in [q]$, and fixed $j \in [d]$, define the cavity sample projections

$$\Xi_{i,m}^{(-j)} := (x_i^{(-j)})^\top \bar{\Theta}_m^{(-j)} \in \mathbb{R}^k, \qquad \zeta_{i,m}^{a,(-j)} := (x_i^{(-j)})^\top \bar{U}_m^{a,(-j)} \in \mathbb{R}^k. \tag{67}$$

Define the back-reaction matrices

$$\Delta_m^{(-j)} := \bar{\Theta}_m - \bar{\Theta}_m^{(-j)} - e_j \otimes \bar{\Theta}_{j,m} \in \mathbb{R}^{d\times k}, \qquad V_m^{a,(-j)} := \bar{U}_m^a - \bar{U}_m^{a,(-j)} - e_j \otimes \bar{U}_{j,m}^a \in \mathbb{R}^{d\times k}, \tag{68}$$

and the corresponding projected back-reaction errors

$$\delta_{i,m}^{(-j)} := x_i^\top \Delta_m^{(-j)} \in \mathbb{R}^k, \qquad \nu_{i,m}^{a,(-j)} := x_i^\top V_m^{a,(-j)} \in \mathbb{R}^k. \tag{69}$$

**Lemma C.5** (Exact decomposition of sample projections)**.** *For every $i \in [n]$, $m \geq 0$, $a \in [q]$, and $j \in [d]$,*

$$x_i^\top \bar{\Theta}_m = \Xi_{i,m}^{(-j)} + x_{ij}\bar{\Theta}_{j,m} + \delta_{i,m}^{(-j)}, \tag{70}$$

$$x_i^\top \bar{U}_m^a = \zeta_{i,m}^{a,(-j)} + x_{ij}\bar{U}_{j,m}^a + \nu_{i,m}^{a,(-j)}. \tag{71}$$

*Proof.* By definition,

$$\bar{\Theta}_m = \bar{\Theta}_m^{(-j)} + e_j \otimes \bar{\Theta}_{j,m} + \Delta_m^{(-j)}.$$

Multiplying by $x_i^\top$ and using $(x_i^{(-j)})^\top \bar{\Theta}_m^{(-j)} = x_i^\top \bar{\Theta}_m^{(-j)}$ because the $j$-th coordinate of both factors is zero, we obtain

$$x_i^\top \bar{\Theta}_m = (x_i^{(-j)})^\top \bar{\Theta}_m^{(-j)} + x_{ij}\bar{\Theta}_{j,m} + x_i^\top \Delta_m^{(-j)} = \Xi_{i,m}^{(-j)} + x_{ij}\bar{\Theta}_{j,m} + \delta_{i,m}^{(-j)}.$$

The proof of (71) is identical. □

**Lemma C.6** (Integral Taylor formulas for the nonlinear learner field)**.** *Fix $w_\star \in \mathbb{R}^{k_\star}$ and $\varepsilon \in \mathbb{R}$, and define*

$$F(\xi) := f(\xi, w_\star, \varepsilon), \qquad \xi \in \mathbb{R}^k.$$

*Then the following hold.*

1. *If $F \in C^2(\mathbb{R}^k; \mathbb{R}^k)$, then for all $\xi, h \in \mathbb{R}^k$,*

$$F(\xi + h) = F(\xi) + DF(\xi)[h] + \int_0^1 (1-s)\, D^2F(\xi + sh)[h,h]\, ds. \tag{72}$$

2. *If $F \in C^3(\mathbb{R}^k; \mathbb{R}^k)$, define*

$$\Psi(\xi, z) := DF(\xi)[z], \qquad (\xi, z) \in \mathbb{R}^k \times \mathbb{R}^k.$$

*Then for all $\xi, h, z, v \in \mathbb{R}^k$,*

$$\begin{aligned}\Psi(\xi + h, z + v) = \Psi(\xi, z) &+ DF(\xi)[v] + D^2F(\xi)[h, z] \\ &+ \int_0^1 (1-s)\Big(D^3F(\xi + sh)[h, h, z + v] + 2D^2F(\xi + sh)[h, v]\Big)\, ds.\end{aligned} \tag{73}$$

*Proof.* For (72), apply the scalar fundamental theorem of calculus to the path $\varphi(s) := F(\xi + sh)$, $s \in [0, 1]$:

$$F(\xi + h) - F(\xi) = \int_0^1 \varphi'(s)\, ds = \int_0^1 DF(\xi + sh)[h]\, ds.$$

Applying the same argument once more to $s \mapsto DF(\xi + sh)[h]$ yields

$$DF(\xi + sh)[h] = DF(\xi)[h] + \int_0^s D^2F(\xi + rh)[h, h]\, dr.$$

Integrating in $s$ and reversing the order of integration gives (72).

For (73), consider the path

$$\psi(s) := \Psi(\xi + sh,\, z + sv) = DF(\xi + sh)[z + sv].$$

Then

$$\psi'(s) = D^2F(\xi + sh)[h,\, z + sv] + DF(\xi + sh)[v].$$

Differentiating once more,

$$\psi''(s) = D^3F(\xi + sh)[h, h,\, z + sv] + 2D^2F(\xi + sh)[h, v].$$

Applying the second-order scalar integral Taylor formula to $\psi$ at $s = 0$ and $s = 1$ gives (73). □

**Definition C.7** (Cavity nonlinear fields and local operators)**.** Fix $j \in [d]$, $i \in [n]$, $m \geq 0$, and $a \in [q]$. Define

$$F_{i,m}^{(-j)} := f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big) \in \mathbb{R}^k, \tag{74}$$

$$L_{i,m}^{(-j)} := D_1 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big) \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^k), \tag{75}$$

$$H_{i,m}^{a,(-j)} := L_{i,m}^{(-j)}\big[\zeta_{i,m}^{a,(-j)}\big] = D_1 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big)\big[\zeta_{i,m}^{a,(-j)}\big] \in \mathbb{R}^k. \tag{76}$$

For $h \in \mathbb{R}^k$, define the local self-interaction operator

$$\mathcal{A}_{j,m}^{(-j)}[h] := \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij}^2 L_{i,m}^{(-j)}[h] \in \mathbb{R}^k, \tag{77}$$

and, for each $a \in [q]$, the mixed bilinear cavity operator

$$\mathcal{B}_{j,m}^{a,(-j)}[h] := \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij}^2 D_1^2 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big)\big[h, \zeta_{i,m}^{a,(-j)}\big] \in \mathbb{R}^k. \tag{78}$$

**Proposition C.8** (Exact cavity expansion for the iterate row)**.** *Fix $j \in [d]$ and $m \geq 0$. Define the row-level cavity forcing*

$$\mathcal{F}_{j,m}^{(-j)} := \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} F_{i,m}^{(-j)} \in \mathbb{R}^k, \tag{79}$$

*the total cavity perturbation*

$$h_{i,j,m}^{\vartheta,(-j)} := x_{ij}\bar{\Theta}_{j,m} + \delta_{i,m}^{(-j)} \in \mathbb{R}^k, \tag{80}$$

*and the iterate remainder*

$$\begin{aligned}\mathcal{E}_{j,m}^{\vartheta,(-j)} := &\frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} L_{i,m}^{(-j)}\big[\delta_{i,m}^{(-j)}\big] \\ &+ \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} \int_0^1 (1 - s)\, D_1^2 f\big(\Xi_{i,m}^{(-j)} + sh_{i,j,m}^{\vartheta,(-j)}, W_i^\star, \varepsilon_i\big)\big[h_{i,j,m}^{\vartheta,(-j)}, h_{i,j,m}^{\vartheta,(-j)}\big]\, ds.\end{aligned} \tag{81}$$

*Then the $j$-th learner row obeys the exact recursion*

$$\bar{\Theta}_{j,m+1} = \bar{\Theta}_{j,m} - \eta_m\Big(\mathcal{F}_{j,m}^{(-j)} + \mathcal{A}_{j,m}^{(-j)}[\bar{\Theta}_{j,m}] + \mathcal{E}_{j,m}^{\vartheta,(-j)} + \frac{1}{n}g(\bar{\Theta}_{j,m})\Big). \tag{82}$$

*Proof.* By the row-wise gradient formula (37),

$$\bar{\Theta}_{j,m+1} = \bar{\Theta}_{j,m} - \eta_m \left[ \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} f\big(x_i^\top \bar{\Theta}_m, W_i^\star, \varepsilon_i\big) + \frac{1}{n} g(\bar{\Theta}_{j,m}) \right].$$

By Lemma C.5,

$$x_i^\top \bar{\Theta}_m = \Xi_{i,m}^{(-j)} + h_{i,j,m}^{\vartheta,(-j)}.$$

Applying (72) from Lemma C.6 with

$$F(\xi) = f(\xi, W_i^\star, \varepsilon_i), \qquad \xi = \Xi_{i,m}^{(-j)}, \qquad h = h_{i,j,m}^{\vartheta,(-j)},$$

we obtain

$$\begin{aligned} f\big(x_i^\top \bar{\Theta}_m, W_i^\star, \varepsilon_i\big) &= F_{i,m}^{(-j)} + L_{i,m}^{(-j)}\big[h_{i,j,m}^{\vartheta,(-j)}\big] \\ &\quad + \int_0^1 (1-s)\, D_1^2 f\big(\Xi_{i,m}^{(-j)} + s h_{i,j,m}^{\vartheta,(-j)}, W_i^\star, \varepsilon_i\big)\big[h_{i,j,m}^{\vartheta,(-j)}, h_{i,j,m}^{\vartheta,(-j)}\big]\, ds. \end{aligned} \tag{83}$$

Now

$$L_{i,m}^{(-j)}\big[h_{i,j,m}^{\vartheta,(-j)}\big] = x_{ij} L_{i,m}^{(-j)}[\bar{\Theta}_{j,m}] + L_{i,m}^{(-j)}[\delta_{i,m}^{(-j)}].$$

Multiplying (83) by $x_{ij}$, averaging over $i \in B_m$, and substituting into the row update yields (82). □

**Proposition C.9** (Exact cavity expansion for the tangent rows)**.** *Fix $j \in [d]$, $m \geq 0$, and $a \in [q]$. Define the tangent forcing*

$$\mathcal{H}_{j,m}^{a,(-j)} := \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} H_{i,m}^{a,(-j)} \in \mathbb{R}^k, \tag{84}$$

*the total tangent perturbation*

$$h_{i,j,m}^{u,a,(-j)} := x_{ij} \bar{U}_{j,m}^a + \nu_{i,m}^{a,(-j)} \in \mathbb{R}^k, \tag{85}$$

*and the tangent remainder*

$$\begin{aligned} \mathcal{E}_{j,m}^{u,a,(-j)} &:= \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} L_{i,m}^{(-j)}\big[\nu_{i,m}^{a,(-j)}\big] \\ &\quad + \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} D_1^2 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big)\big[\delta_{i,m}^{(-j)}, \zeta_{i,m}^{a,(-j)}\big] \\ &\quad + \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} \int_0^1 (1-s)\Big[D_1^3 f\big(\Xi_{i,m}^{(-j)} + s h_{i,j,m}^{\vartheta,(-j)}, W_i^\star, \varepsilon_i\big)\big[h_{i,j,m}^{\vartheta,(-j)}, h_{i,j,m}^{\vartheta,(-j)}, \zeta_{i,m}^{a,(-j)} + h_{i,j,m}^{u,a,(-j)}\big] \\ &\qquad + 2 D_1^2 f\big(\Xi_{i,m}^{(-j)} + s h_{i,j,m}^{\vartheta,(-j)}, W_i^\star, \varepsilon_i\big)\big[h_{i,j,m}^{\vartheta,(-j)}, h_{i,j,m}^{u,a,(-j)}\big]\Big]\, ds. \end{aligned} \tag{86}$$

*Then the $j$-th tangent row obeys the exact recursion*

$$\bar{U}_{j,m+1}^a = \bar{U}_{j,m}^a - \eta_m \Big(\mathcal{H}_{j,m}^{a,(-j)} + \mathcal{A}_{j,m}^{(-j)}[\bar{U}_{j,m}^a] + \mathcal{B}_{j,m}^{a,(-j)}[\bar{\Theta}_{j,m}] + \mathcal{E}_{j,m}^{u,a,(-j)} + \frac{1}{n} Dg(\bar{\Theta}_{j,m})[\bar{U}_{j,m}^a]\Big). \tag{87}$$

*Proof.* By the row-wise Hessian action formula (38),

$$\bar{U}_{j,m+1}^a = \bar{U}_{j,m}^a - \eta_m \left[ \frac{1}{\kappa_n} \sum_{i \in B_m} x_{ij} D_1 f\big(x_i^\top \bar{\Theta}_m, W_i^\star, \varepsilon_i\big)\big[x_i^\top \bar{U}_m^a\big] + \frac{1}{n} Dg(\bar{\Theta}_{j,m})[\bar{U}_{j,m}^a] \right]. \tag{88}$$

By Lemma C.5,

$$x_i^\top \bar{\Theta}_m = \Xi_{i,m}^{(-j)} + h_{i,j,m}^{\vartheta,(-j)}, \qquad x_i^\top \bar{U}_m^a = \zeta_{i,m}^{a,(-j)} + h_{i,j,m}^{u,a,(-j)}.$$

Apply (73) from Lemma C.6 with

$$F(\xi) = f(\xi, W_i^\star, \varepsilon_i), \qquad \xi = \Xi_{i,m}^{(-j)}, \qquad h = h_{i,j,m}^{\vartheta,(-j)}, \qquad z = \zeta_{i,m}^{a,(-j)}, \qquad v = h_{i,j,m}^{u,a,(-j)}.$$

This yields

$$\begin{aligned} &D_1 f\big(x_i^\top \bar{\Theta}_m, W_i^\star, \varepsilon_i\big)\big[x_i^\top \bar{U}_m^a\big] \\ &\quad = H_{i,m}^{a,(-j)} + L_{i,m}^{(-j)}\big[h_{i,j,m}^{u,a,(-j)}\big] + D_1^2 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big)\big[h_{i,j,m}^{\vartheta,(-j)}, \zeta_{i,m}^{a,(-j)}\big] \\ &\qquad + \int_0^1 (1-s)\Big[D_1^3 f\big(\Xi_{i,m}^{(-j)} + s h_{i,j,m}^{\vartheta,(-j)}, W_i^\star, \varepsilon_i\big)\big[h_{i,j,m}^{\vartheta,(-j)}, h_{i,j,m}^{\vartheta,(-j)}, \zeta_{i,m}^{a,(-j)} + h_{i,j,m}^{u,a,(-j)}\big] \\ &\qquad\qquad + 2D_1^2 f\big(\Xi_{i,m}^{(-j)} + s h_{i,j,m}^{\vartheta,(-j)}, W_i^\star, \varepsilon_i\big)\big[h_{i,j,m}^{\vartheta,(-j)}, h_{i,j,m}^{u,a,(-j)}\big]\Big]\, ds. \end{aligned} \tag{89}$$

Now decompose the linear terms:

$$L_{i,m}^{(-j)}\big[h_{i,j,m}^{u,a,(-j)}\big] = x_{ij} L_{i,m}^{(-j)}[\bar{U}_{j,m}^a] + L_{i,m}^{(-j)}[\nu_{i,m}^{a,(-j)}],$$

and

$$\begin{aligned} D_1^2 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big)\big[h_{i,j,m}^{\vartheta,(-j)}, \zeta_{i,m}^{a,(-j)}\big] &= x_{ij} D_1^2 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big)\big[\bar{\Theta}_{j,m}, \zeta_{i,m}^{a,(-j)}\big] \\ &\quad + D_1^2 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big)\big[\delta_{i,m}^{(-j)}, \zeta_{i,m}^{a,(-j)}\big]. \end{aligned}$$

Multiply (89) by $x_{ij}$, average over $i \in B_m$, and substitute into (88). The terms proportional to $x_{ij} H_{i,m}^{a,(-j)}$, $x_{ij}^2 L_{i,m}^{(-j)}$, and $x_{ij}^2 D_1^2 f(\cdot)[\bar{\Theta}_{j,m}, \zeta_{i,m}^{a,(-j)}]$ produce $\mathcal{H}_{j,m}^{a,(-j)}$, $\mathcal{A}_{j,m}^{(-j)}[\bar{U}_{j,m}^a]$, and $\mathcal{B}_{j,m}^{a,(-j)}[\bar{\Theta}_{j,m}]$, respectively. The remaining terms are exactly $\mathcal{E}_{j,m}^{u,a,(-j)}$, proving (87). ☐

*Remark* C.10 (What is new relative to the iterate-only cavity algebra). Proposition C.8 is the iterate-level cavity decomposition, now written in the notation of the present paper. The genuinely new structure appears in Proposition C.9: the same self-interaction operator $\mathcal{A}_{j,m}^{(-j)}$ reappears on the tangent row, but there is an additional bilinear operator $\mathcal{B}_{j,m}^{a,(-j)}$ coupling the current learner coordinate $\bar{\Theta}_{j,m}$ to the cavity tangent field $\zeta_{i,m}^{a,(-j)}$. This operator is the deterministic seed of the multiplicative theory developed later in the paper.

### C.2. Identification of Candidate Effective Kernels

The previous subsection isolates, at the exact finite-$n, d$ level, the four objects that control the representative coordinate:

1. the cavity learner fields $\Xi_{i,m}^{(-j)}$,
2. the cavity tangent fields $\zeta_{i,m}^{a,(-j)}$,
3. the local operators $\mathcal{A}_{j,m}^{(-j)}$ and $\mathcal{B}_{j,m}^{a,(-j)}$,
4. the back-reaction remainders $\mathcal{E}_{j,m}^{\vartheta,(-j)}$ and $\mathcal{E}_{j,m}^{u,a,(-j)}$.

The later probabilistic analysis will show that, on fixed horizons, the remainder terms are negligible and the first three classes self-average onto a deterministic kernel system. We now record the exact prelimit arrays from which those kernels will be extracted.

**Definition C.11** (Discrete replicated state kernels)**.** Fix $M \in \mathbb{N}$. For $0 \le r, m \le M$ and $a, b \in [q]$, define

$$C^{\vartheta,\vartheta,(d)}_{m,r} := \frac{1}{d}\sum_{\ell=1}^{d} \bar{\Theta}_{\ell,m} \otimes \bar{\Theta}_{\ell,r} \in \mathbb{R}^{k\times k}, \tag{90}$$

$$C^{\vartheta,\star,(d)}_{m} := \frac{1}{d}\sum_{\ell=1}^{d} \bar{\Theta}_{\ell,m} \otimes \Theta_{\star,\ell} \in \mathbb{R}^{k\times k_\star}, \tag{91}$$

$$C^{u^a,\vartheta,(d)}_{m,r} := \frac{1}{d}\sum_{\ell=1}^{d} \bar{U}^a_{\ell,m} \otimes \bar{\Theta}_{\ell,r} \in \mathbb{R}^{k\times k}, \tag{92}$$

$$C^{u^a,u^b,(d)}_{m,r} := \frac{1}{d}\sum_{\ell=1}^{d} \bar{U}^a_{\ell,m} \otimes \bar{U}^b_{\ell,r} \in \mathbb{R}^{k\times k}, \tag{93}$$

$$C^{u^a,\star,(d)}_{m} := \frac{1}{d}\sum_{\ell=1}^{d} \bar{U}^a_{\ell,m} \otimes \Theta_{\star,\ell} \in \mathbb{R}^{k\times k_\star}. \tag{94}$$

**Definition C.12** (Discrete replicated field kernels)**.** Fix $M \in \mathbb{N}$, $j \in [d]$, and define the cavity nonlinear fields $F^{(-j)}_{i,m}$ and $H^{a,(-j)}_{i,m}$ as in (74)–(76). For $0 \le r, m \le M$ and $a, b \in [q]$, define

$$C^{f,f,(-j,d)}_{m,r} := \frac{1}{n}\sum_{i=1}^{n} F^{(-j)}_{i,m} \otimes F^{(-j)}_{i,r} \in \mathbb{R}^{k\times k}, \tag{95}$$

$$C^{h^a,f,(-j,d)}_{m,r} := \frac{1}{n}\sum_{i=1}^{n} H^{a,(-j)}_{i,m} \otimes F^{(-j)}_{i,r} \in \mathbb{R}^{k\times k}, \tag{96}$$

$$C^{h^a,h^b,(-j,d)}_{m,r} := \frac{1}{n}\sum_{i=1}^{n} H^{a,(-j)}_{i,m} \otimes H^{b,(-j)}_{i,r} \in \mathbb{R}^{k\times k}. \tag{97}$$

**Definition C.13** (Discrete local propagators)**.** Fix $j \in [d]$ and $M \in \mathbb{N}$. For integers $0 \le r \le m \le M$, define the discrete local propagator

$$\mathcal{P}^{(-j)}_{m\leftarrow r} := \begin{cases} I_k, & m = r, \\ \big(I_k - \eta_{m-1}\mathcal{A}^{(-j)}_{j,m-1}\big)\cdots\big(I_k - \eta_r\mathcal{A}^{(-j)}_{j,r}\big), & m > r, \end{cases} \tag{98}$$

where the factors are ordered from right to left in increasing time. Thus $\mathcal{P}^{(-j)}_{m\leftarrow r}$ is the exact propagator of the homogeneous linear recursion generated by the local self-interaction operators $\mathcal{A}^{(-j)}_{j,\ell}$.

**Proposition C.14** (Variation-of-constants representation)**.** *Fix $j \in [d]$ and $M \in \mathbb{N}$. Then for every $m \in \{0, \dots, M\}$,*

$$\bar{\Theta}_{j,m} = \mathcal{P}^{(-j)}_{m\leftarrow 0}\Theta_{0,j} - \sum_{r=0}^{m-1} \eta_r\,\mathcal{P}^{(-j)}_{m\leftarrow r+1}\Big(\mathcal{F}^{(-j)}_{j,r} + \mathcal{E}^{\vartheta,(-j)}_{j,r} + \frac{1}{n}g(\bar{\Theta}_{j,r})\Big), \tag{99}$$

$$\bar{U}^a_{j,m} = \mathcal{P}^{(-j)}_{m\leftarrow 0}U^a_{0,j} - \sum_{r=0}^{m-1} \eta_r\,\mathcal{P}^{(-j)}_{m\leftarrow r+1}\Big(\mathcal{H}^{a,(-j)}_{j,r} + \mathcal{B}^{a,(-j)}_{j,r}[\bar{\Theta}_{j,r}] + \mathcal{E}^{u,a,(-j)}_{j,r} + \frac{1}{n}Dg(\bar{\Theta}_{j,r})[\bar{U}^a_{j,r}]\Big) \tag{100}$$

*for each $a \in [q]$.*

*Proof.* The recursion (82) has the form

$$\theta_{m+1} = (I_k - \eta_m\mathcal{A}^{(-j)}_{j,m})\theta_m - \eta_m b_m,$$

with

$$\theta_m := \bar{\Theta}_{j,m}, \qquad b_m := \mathcal{F}^{(-j)}_{j,m} + \mathcal{E}^{\vartheta,(-j)}_{j,m} + \frac{1}{n}g(\bar{\Theta}_{j,m}).$$

Unrolling this nonautonomous linear recursion yields (99). The argument is standard and can be verified by induction on $m$.

Likewise, (87) has the form

$$u_{m+1} = (I_k - \eta_m \mathcal{A}^{(-j)}_{j,m}) u_m - \eta_m c_m,$$

with

$$u_m := \bar{U}^a_{j,m}, \qquad c_m := \mathcal{H}^{a,(-j)}_{j,m} + \mathcal{B}^{a,(-j)}_{j,m}[\bar{\Theta}_{j,m}] + \mathcal{E}^{u,a,(-j)}_{j,m} + \frac{1}{n} Dg(\bar{\Theta}_{j,m})[\bar{U}^a_{j,m}].$$

Unrolling again proves (100). □

**Definition C.15** (Candidate replicated kernel family)**.** Fix a finite horizon $M \in \mathbb{N}$ and a replica number $q \geq 1$. The *candidate replicated kernel family up to time* $M$ is the collection

$$\begin{aligned}
\mathfrak{K}^{(q,d)}_M := \Big( &\{C^{\vartheta,\vartheta,(d)}_{m,r}\}_{0\leq r,m\leq M}, \{C^{\vartheta,\star,(d)}_{m}\}_{0\leq m\leq M}, \\
&\{C^{u^a,\vartheta,(d)}_{m,r}\}_{a\in[q],\,0\leq r,m\leq M}, \{C^{u^a,u^b,(d)}_{m,r}\}_{a,b\in[q],\,0\leq r,m\leq M}, \\
&\{C^{u^a,\star,(d)}_{m}\}_{a\in[q],\,0\leq m\leq M}, \{C^{f,f,(-j,d)}_{m,r}\}_{0\leq r,m\leq M}, \\
&\{C^{h^a,f,(-j,d)}_{m,r}\}_{a\in[q],\,0\leq r,m\leq M}, \{C^{h^a,h^b,(-j,d)}_{m,r}\}_{a,b\in[q],\,0\leq r,m\leq M}, \\
&\{\mathcal{P}^{(-j)}_{m\leftarrow r}\}_{0\leq r\leq m\leq M}, \{\mathcal{B}^{a,(-j)}_{j,m}\}_{a\in[q],\,0\leq m\leq M}\Big).
\end{aligned} \tag{101}$$

*Remark* C.16 (Closure blueprint). The exact formulas above isolate the finite-horizon closure problem. To prove Theorem 3.1, it is enough to show the following three facts, all of which are deferred to later appendices.

1. The back-reaction terms $\delta^{(-j)}_{i,m}$, $\nu^{a,(-j)}_{i,m}$, $\mathcal{E}^{\vartheta,(-j)}_{j,m}$, and $\mathcal{E}^{u,a,(-j)}_{j,m}$ are negligible uniformly on compact horizons.
2. The candidate family $\mathfrak{K}^{(q,d)}_M$ self-averages onto a deterministic limit as $n, d \to \infty$.
3. That deterministic limit is well posed and yields the effective replicated process announced in Theorem 3.1.

The rest of the technical paper is organized exactly around these three points: Appendix D handles uniform control and concentration, Appendix E constructs the limiting deterministic kernel system, and Appendix F identifies the high-dimensional limit of the joint iterate–tangent dynamics.

# D. Uniform Bounds and Concentration

This appendix provides the quantitative estimates that make the augmented cavity expansion of Appendix C usable. The main points are the following.

1. We first record deterministic norm bounds for the cavity forcing terms, local operators, propagators, and remainder terms. These bounds are exact and require only the regularity assumptions of Appendix A.
2. We then exploit the fact that, conditional on the past, each mini-batch is sampled uniformly without replacement from a deterministic finite family. This yields exact conditional variance identities for the primitive cavity averages $\mathcal{F}, \mathcal{A}, \mathcal{H}, \mathcal{B}$.
3. Finally, we convert those one-step variance identities into finite-horizon $L^2$ martingale bounds for the accumulated batch fluctuations. At the row scale, the forcing fluctuations are of order one over $O(1)$ epochs, while the operator fluctuations are smaller by one additional factor of $d^{-1/2}$. This is precisely the separation needed later for the replicated DMFT.

Throughout this appendix, fix a replica number $q \geq 1$ and a finite horizon $T > 0$, and write

$$M_T := \lfloor T n^{1-\alpha} \rfloor. \tag{102}$$

All constants below may depend on $T$, on $q$, on the fixed latent dimensions $k, k_\star$, and on the regularity constants in Assumptions A.3–A.5, but never on $n, d$. To keep formulas readable, we write

$$L_r := \sup_{(\xi, w_\star, \varepsilon)} \|D^r_1 f(\xi, w_\star, \varepsilon)\|_{\mathrm{op}}, \qquad r = 0, 1, 2, 3, \tag{103}$$

where $D_1^0 f := f$, and

$$L_g := \sup_{u\in\mathbb{R}^k} \|Dg(u)\|_{\mathrm{op}}, \qquad C_g := \sup_{u\in\mathbb{R}^k} \frac{\|g(u)\|_2}{1+\|u\|_2}. \tag{104}$$

By Assumption A.4, these constants are finite.

### D.1. Uniform Moment Bounds

We first isolate deterministic envelopes for the primitive objects appearing in Appendix C. Since $k$ is fixed, all norms on $\mathcal{L}(\mathbb{R}^k, \mathbb{R}^k)$ are equivalent; in the concentration estimates below we will use the Frobenius norm $\|\cdot\|_{\mathrm{F}}$ for matrix-valued quantities.

**Definition D.1** (Full-sample means of the primitive cavity averages)**.** Fix $j \in [d]$, $m \geq 0$, and $a \in [q]$. Define

$$\mathfrak{F}_{j,m}^{(-j)} := \frac{1}{n}\sum_{i=1}^n x_{ij} F_{i,m}^{(-j)} \in \mathbb{R}^k, \tag{105}$$

$$\mathfrak{A}_{j,m}^{(-j)} := \frac{1}{n}\sum_{i=1}^n x_{ij}^2 L_{i,m}^{(-j)} \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^k), \tag{106}$$

$$\mathfrak{H}_{j,m}^{a,(-j)} := \frac{1}{n}\sum_{i=1}^n x_{ij} H_{i,m}^{a,(-j)} \in \mathbb{R}^k, \tag{107}$$

$$\mathfrak{B}_{j,m}^{a,(-j)}[h] := \frac{1}{n}\sum_{i=1}^n x_{ij}^2 D_1^2 f\big(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i\big)\big[h, \zeta_{i,m}^{a,(-j)}\big] \in \mathbb{R}^k, \qquad h \in \mathbb{R}^k. \tag{108}$$

Thus $\mathfrak{F}, \mathfrak{A}, \mathfrak{H}, \mathfrak{B}$ are the corresponding full-data averages, whereas $\mathcal{F}, \mathcal{A}, \mathcal{H}, \mathcal{B}$ from Appendix C are their mini-batch versions.

**Lemma D.2** (Primitive deterministic norm bounds)**.** *Fix $j \in [d]$, $m \geq 0$, and $a \in [q]$. Then the following bounds hold.*

1. *The cavity nonlinear fields satisfy*

$$\|F_{i,m}^{(-j)}\|_2 \leq L_0, \tag{109}$$

$$\|L_{i,m}^{(-j)}\|_{\mathrm{op}} \leq L_1, \tag{110}$$

$$\|H_{i,m}^{a,(-j)}\|_2 \leq L_1 \|\zeta_{i,m}^{a,(-j)}\|_2. \tag{111}$$

2. *The local operators satisfy, for every $h \in \mathbb{R}^k$,*

$$\|\mathcal{F}_{j,m}^{(-j)}\|_2 \leq L_0 \,\frac{1}{\kappa_n} \sum_{i\in B_m} |x_{ij}|, \tag{112}$$

$$\|\mathcal{A}_{j,m}^{(-j)}\|_{\mathrm{op}} \leq L_1 \,\frac{1}{\kappa_n} \sum_{i\in B_m} x_{ij}^2, \tag{113}$$

$$\|\mathcal{H}_{j,m}^{a,(-j)}\|_2 \leq L_1 \,\frac{1}{\kappa_n} \sum_{i\in B_m} |x_{ij}|\, \|\zeta_{i,m}^{a,(-j)}\|_2, \tag{114}$$

$$\|\mathcal{B}_{j,m}^{a,(-j)}[h]\|_2 \leq L_2 \|h\|_2 \,\frac{1}{\kappa_n} \sum_{i\in B_m} x_{ij}^2\, \|\zeta_{i,m}^{a,(-j)}\|_2. \tag{115}$$

3. *The iterate remainder obeys*

$$\|\mathcal{E}_{j,m}^{\vartheta,(-j)}\|_2 \leq L_1 \,\frac{1}{\kappa_n} \sum_{i\in B_m} |x_{ij}|\, \|\delta_{i,m}^{(-j)}\|_2 + \frac{L_2}{2}\,\frac{1}{\kappa_n} \sum_{i\in B_m} |x_{ij}|\, \|h_{i,j,m}^{\vartheta,(-j)}\|_2^2. \tag{116}$$

*4. The tangent remainder obeys*

$$\begin{aligned}\|\mathcal{E}^{u,a,(-j)}_{j,m}\|_2 \le L_1 \frac{1}{\kappa_n} \sum_{i\in B_m} |x_{ij}|\, \|\nu^{a,(-j)}_{i,m}\|_2 + L_2 \frac{1}{\kappa_n} \sum_{i\in B_m} |x_{ij}|\, \|\delta^{(-j)}_{i,m}\|_2\, \|\zeta^{a,(-j)}_{i,m}\|_2 \\ + L_2 \frac{1}{\kappa_n} \sum_{i\in B_m} |x_{ij}|\, \|h^{\vartheta,(-j)}_{i,j,m}\|_2\, \|h^{u,a,(-j)}_{i,j,m}\|_2 \\ + \frac{L_3}{2} \frac{1}{\kappa_n} \sum_{i\in B_m} |x_{ij}|\, \|h^{\vartheta,(-j)}_{i,j,m}\|_2^2 \big(\|\zeta^{a,(-j)}_{i,m}\|_2 + \|h^{u,a,(-j)}_{i,j,m}\|_2\big).\end{aligned} \tag{117}$$

*Proof.* The bounds (109) and (110) are immediate from the definitions and (103); (111) then follows from (76). The operator bounds (112)–(115) are direct consequences of (79), (77), (84), and (78), together with the previous point and the bound

$$\|D_1^2 f(\xi, w_\star, \varepsilon)[h, z]\|_2 \le L_2 \|h\|_2 \|z\|_2.$$

For the remainder terms, use the exact formulas (81) and (86). The first term in (116) is immediate from $\|L^{(-j)}_{i,m}\|_{\mathrm{op}} \le L_1$. The integral term is bounded by

$$\int_0^1 (1-s)\, L_2\, \|h^{\vartheta,(-j)}_{i,j,m}\|_2^2\, ds = \frac{L_2}{2} \|h^{\vartheta,(-j)}_{i,j,m}\|_2^2.$$

This proves (116).

For (117), use first

$$\|L^{(-j)}_{i,m}[\nu^{a,(-j)}_{i,m}]\|_2 \le L_1 \|\nu^{a,(-j)}_{i,m}\|_2,$$

and

$$\|D_1^2 f(\Xi^{(-j)}_{i,m}, W_i^\star, \varepsilon_i)[\delta^{(-j)}_{i,m}, \zeta^{a,(-j)}_{i,m}]\|_2 \le L_2 \|\delta^{(-j)}_{i,m}\|_2\, \|\zeta^{a,(-j)}_{i,m}\|_2.$$

For the integral term in (86), the bounds

$$\|D_1^3 f(\cdot)[h,h,z]\|_2 \le L_3 \|h\|_2^2 \|z\|_2, \qquad \|D_1^2 f(\cdot)[h,v]\|_2 \le L_2 \|h\|_2 \|v\|_2$$

yield

$$\begin{aligned}\int_0^1 (1-s)\Big[\|D_1^3 f(\cdot)[h,h,z+v]\|_2 + 2\|D_1^2 f(\cdot)[h,v]\|_2\Big]\, ds \\ \le \frac{L_3}{2} \|h\|_2^2 (\|z\|_2 + \|v\|_2) + L_2 \|h\|_2 \|v\|_2.\end{aligned}$$

Applying this pointwise with

$$h = h^{\vartheta,(-j)}_{i,j,m}, \qquad z = \zeta^{a,(-j)}_{i,m}, \qquad v = h^{u,a,(-j)}_{i,j,m}$$

and averaging over $i \in B_m$ proves (117). $\square$

**Lemma D.3** (Propagator envelope)**.** *Fix $j \in [d]$, integers $0 \le r \le m$, and the discrete local propagator $\mathcal{P}^{(-j)}_{m\leftarrow r}$ from Definition C.13. Then*

$$\|\mathcal{P}^{(-j)}_{m\leftarrow r}\|_{\mathrm{op}} \le \exp\!\left(\sum_{\ell=r}^{m-1} \eta_\ell\, \|\mathcal{A}^{(-j)}_{j,\ell}\|_{\mathrm{op}}\right). \tag{118}$$

*Proof.* By submultiplicativity,

$$\|\mathcal{P}^{(-j)}_{m\leftarrow r}\|_{\mathrm{op}} \le \prod_{\ell=r}^{m-1} \|I_k - \eta_\ell \mathcal{A}^{(-j)}_{j,\ell}\|_{\mathrm{op}}.$$

Since $\|I_k - \eta_\ell \mathcal{A}^{(-j)}_{j,\ell}\|_{\mathrm{op}} \le 1 + \eta_\ell \|\mathcal{A}^{(-j)}_{j,\ell}\|_{\mathrm{op}}$, we get

$$\|\mathcal{P}^{(-j)}_{m\leftarrow r}\|_{\mathrm{op}} \le \prod_{\ell=r}^{m-1} \big(1 + \eta_\ell \|\mathcal{A}^{(-j)}_{j,\ell}\|_{\mathrm{op}}\big) \le \exp\!\left(\sum_{\ell=r}^{m-1} \eta_\ell \|\mathcal{A}^{(-j)}_{j,\ell}\|_{\mathrm{op}}\right),$$

using $1 + x \le e^x$ for $x \ge 0$. $\square$

**Proposition D.4** (Row-wise a priori envelopes)**.** *Fix $j \in [d]$ and $m \in \{0, \dots, M_T\}$. Define the accumulated local curvature*

$$\mathfrak{a}_{j,m} := \sum_{r=0}^{m-1} \eta_r \|\mathcal{A}_{j,r}^{(-j)}\|_{\mathrm{op}}, \qquad \mathfrak{a}_{j,0} := 0. \tag{119}$$

*Then the iterate row and tangent rows satisfy the deterministic bounds*

$$\|\bar{\Theta}_{j,m}\|_2 \le e^{\mathfrak{a}_{j,m}} \left[ \|\Theta_{0,j}\|_2 + \sum_{r=0}^{m-1} \eta_r \Big( \|\mathcal{F}_{j,r}^{(-j)}\|_2 + \|\mathcal{E}_{j,r}^{\vartheta,(-j)}\|_2 + \frac{1}{n}\|g(\bar{\Theta}_{j,r})\|_2 \Big) \right], \tag{120}$$

$$\|\bar{U}_{j,m}^a\|_2 \le e^{\mathfrak{a}_{j,m}} \left[ \|U_{0,j}^a\|_2 + \sum_{r=0}^{m-1} \eta_r \Big( \|\mathcal{H}_{j,r}^{a,(-j)}\|_2 + \|\mathcal{B}_{j,r}^{a,(-j)}[\bar{\Theta}_{j,r}]\|_2 + \|\mathcal{E}_{j,r}^{u,a,(-j)}\|_2 + \frac{1}{n}\|Dg(\bar{\Theta}_{j,r})[\bar{U}_{j,r}^a]\|_2 \Big) \right] \tag{121}$$

*for each $a \in [q]$.*

*Proof.* The variation-of-constants formulas (99) and (100) from Proposition C.14 combined with Lemma D.3 imply

$$\begin{aligned} \|\bar{\Theta}_{j,m}\|_2 &\le \|\mathcal{P}_{m\leftarrow 0}^{(-j)}\|_{\mathrm{op}} \|\Theta_{0,j}\|_2 + \sum_{r=0}^{m-1} \eta_r \, \|\mathcal{P}_{m\leftarrow r+1}^{(-j)}\|_{\mathrm{op}} \Big( \|\mathcal{F}_{j,r}^{(-j)}\|_2 + \|\mathcal{E}_{j,r}^{\vartheta,(-j)}\|_2 + \frac{1}{n}\|g(\bar{\Theta}_{j,r})\|_2 \Big) \\ &\le e^{\mathfrak{a}_{j,m}} \left[ \|\Theta_{0,j}\|_2 + \sum_{r=0}^{m-1} \eta_r \Big( \|\mathcal{F}_{j,r}^{(-j)}\|_2 + \|\mathcal{E}_{j,r}^{\vartheta,(-j)}\|_2 + \frac{1}{n}\|g(\bar{\Theta}_{j,r})\|_2 \Big) \right], \end{aligned}$$

which proves (120). The proof of (121) is identical. □

*Remark* D.5 (Why the envelopes are finite-horizon rather than asymptotic). Proposition D.4 is intentionally deterministic. It says that every row is controlled by three inputs only:

1. the accumulated local curvature $\mathfrak{a}_{j,m}$,
2. the mini-batch forcing terms $\mathcal{F}, \mathcal{H}, \mathcal{B}$,
3. the back-reaction remainders $\mathcal{E}^\vartheta, \mathcal{E}^u$.

The remaining task is therefore to show that these inputs are well behaved over $m \le M_T$; that is exactly the purpose of the concentration analysis in the next subsection and of the small-backreaction argument in Appendix F.

### D.2. Self-Averaging and Concentration of Overlaps

The concentration mechanism at one step is simple: conditional on the past, the current batch $B_m$ is sampled uniformly without replacement from a deterministic finite family. We exploit this using exact second-moment identities.

**Definition D.6** (Natural filtration)**.** Let

$$\mathscr{H}_m := \sigma\big(X, \Theta_0, \Theta_\star, \varepsilon, B_0, \dots, B_{m-1}\big), \qquad m \ge 0. \tag{122}$$

Then $B_m$ is independent of $\mathscr{H}_m$, and all cavity objects indexed by time $m$ and depending on past batches only through $\bar{\Theta}_m^{(-j)}$, $\bar{U}_m^{a,(-j)}$ are $\mathscr{H}_m$-measurable. In particular, for fixed $j, m, a$,

$$\Xi_{i,m}^{(-j)},\ \zeta_{i,m}^{a,(-j)},\ F_{i,m}^{(-j)},\ L_{i,m}^{(-j)},\ H_{i,m}^{a,(-j)}$$

are $\mathscr{H}_m$-measurable for every $i \in [n]$.

**Lemma D.7** (Sampling without replacement in a Hilbert space)**.** *Let $\mathscr{E}$ be a finite-dimensional real Hilbert space, and let $a_1, \dots, a_n \in \mathscr{E}$ be deterministic elements. Let $B \subseteq [n]$ be a uniformly random subset of cardinality $\kappa \in \{1, \dots, n\}$, and define*

$$\bar{a} := \frac{1}{n}\sum_{i=1}^n a_i, \qquad A_B := \frac{1}{\kappa}\sum_{i \in B} a_i.$$

*Then*

$$\mathbb{E}[A_B] = \bar{a}, \tag{123}$$

*and*

$$\mathbb{E}\|A_B - \bar{a}\|_{\mathscr{E}}^2 = \frac{n-\kappa}{\kappa(n-1)} \cdot \frac{1}{n}\sum_{i=1}^{n} \|a_i - \bar{a}\|_{\mathscr{E}}^2. \tag{124}$$

*In particular,*

$$\mathbb{E}\|A_B - \bar{a}\|_{\mathscr{E}}^2 \le \frac{n-\kappa}{\kappa(n-1)} \cdot \frac{1}{n}\sum_{i=1}^{n} \|a_i\|_{\mathscr{E}}^2. \tag{125}$$

*Proof.* Let $I_i := \mathbf{1}\{i \in B\}$. Then

$$A_B - \bar{a} = \sum_{i=1}^{n}\Big(\frac{I_i}{\kappa} - \frac{1}{n}\Big)a_i = \frac{1}{\kappa}\sum_{i=1}^{n} I_i(a_i - \bar{a}),$$

because $\sum_{i=1}^{n}(a_i - \bar{a}) = 0$. Since $B$ is uniform over $\kappa$-subsets,

$$\mathbb{E}[I_i] = \frac{\kappa}{n}, \qquad \mathrm{Var}(I_i) = \frac{\kappa(n-\kappa)}{n^2}, \qquad \mathrm{Cov}(I_i, I_j) = -\frac{\kappa(n-\kappa)}{n^2(n-1)} \quad (i \neq j).$$

Therefore

$$\begin{aligned}
\mathbb{E}\|A_B - \bar{a}\|_{\mathscr{E}}^2 &= \frac{1}{\kappa^2}\sum_{i,j=1}^{n} \mathrm{Cov}(I_i, I_j)\ \langle a_i - \bar{a},\, a_j - \bar{a}\rangle_{\mathscr{E}} \\
&= \frac{1}{\kappa^2}\left[\frac{\kappa(n-\kappa)}{n^2}\sum_{i=1}^{n}\|a_i - \bar{a}\|_{\mathscr{E}}^2 - \frac{\kappa(n-\kappa)}{n^2(n-1)}\sum_{i\neq j}\langle a_i - \bar{a},\, a_j - \bar{a}\rangle_{\mathscr{E}}\right].
\end{aligned}$$

Since $\sum_{i=1}^{n}(a_i - \bar{a}) = 0$,

$$\sum_{i\neq j}\langle a_i - \bar{a},\, a_j - \bar{a}\rangle_{\mathscr{E}} = -\sum_{i=1}^{n}\|a_i - \bar{a}\|_{\mathscr{E}}^2.$$

Substituting this gives

$$\mathbb{E}\|A_B - \bar{a}\|_{\mathscr{E}}^2 = \frac{1}{\kappa^2}\frac{\kappa(n-\kappa)}{n^2}\Big(1 + \frac{1}{n-1}\Big)\sum_{i=1}^{n}\|a_i - \bar{a}\|_{\mathscr{E}}^2 = \frac{n-\kappa}{\kappa(n-1)}\frac{1}{n}\sum_{i=1}^{n}\|a_i - \bar{a}\|_{\mathscr{E}}^2,$$

which proves (124). The bound (125) follows from

$$\frac{1}{n}\sum_{i=1}^{n}\|a_i - \bar{a}\|_{\mathscr{E}}^2 = \frac{1}{n}\sum_{i=1}^{n}\|a_i\|_{\mathscr{E}}^2 - \|\bar{a}\|_{\mathscr{E}}^2 \le \frac{1}{n}\sum_{i=1}^{n}\|a_i\|_{\mathscr{E}}^2.$$

□

**Proposition D.8** (Conditional second-moment bounds for the primitive cavity averages)**.** *Fix $j \in [d]$, $m \in \{0, \ldots, M_T - 1\}$, and $a \in [q]$. Then:*

1. *The iterate forcing is conditionally unbiased and satisfies*

$$\mathbb{E}\Big[\mathcal{F}_{j,m}^{(-j)}\ \Big|\ \mathscr{H}_m\Big] = \mathfrak{F}_{j,m}^{(-j)}, \tag{126}$$

$$\mathbb{E}\Big[\Big\|\mathcal{F}_{j,m}^{(-j)} - \mathfrak{F}_{j,m}^{(-j)}\Big\|_2^2\ \Big|\ \mathscr{H}_m\Big] \le \frac{n-\kappa_n}{\kappa_n(n-1)} \cdot \frac{L_0^2}{n}\sum_{i=1}^{n} x_{ij}^2. \tag{127}$$

2. *The local self-interaction operator is conditionally unbiased and satisfies*

$$\mathbb{E}\Big[\mathcal{A}_{j,m}^{(-j)}\ \Big|\ \mathscr{H}_m\Big] = \mathfrak{A}_{j,m}^{(-j)}, \tag{128}$$

$$\mathbb{E}\Big[\Big\|\mathcal{A}_{j,m}^{(-j)} - \mathfrak{A}_{j,m}^{(-j)}\Big\|_{\mathrm{F}}^2\ \Big|\ \mathscr{H}_m\Big] \le \frac{n-\kappa_n}{\kappa_n(n-1)} \cdot \frac{kL_1^2}{n}\sum_{i=1}^n x_{ij}^4. \tag{129}$$

3. *The tangent forcing is conditionally unbiased and satisfies*

$$\mathbb{E}\Big[\mathcal{H}_{j,m}^{a,(-j)}\ \Big|\ \mathscr{H}_m\Big] = \mathfrak{H}_{j,m}^{a,(-j)}, \tag{130}$$

$$\mathbb{E}\Big[\Big\|\mathcal{H}_{j,m}^{a,(-j)} - \mathfrak{H}_{j,m}^{a,(-j)}\Big\|_2^2\ \Big|\ \mathscr{H}_m\Big] \le \frac{n-\kappa_n}{\kappa_n(n-1)} \cdot \frac{L_1^2}{n}\sum_{i=1}^n x_{ij}^2\|\zeta_{i,m}^{a,(-j)}\|_2^2. \tag{131}$$

4. *For every $\mathscr{H}_m$-measurable $h_m \in \mathbb{R}^k$, the bilinear term is conditionally unbiased and satisfies*

$$\mathbb{E}\Big[\mathcal{B}_{j,m}^{a,(-j)}[h_m]\ \Big|\ \mathscr{H}_m\Big] = \mathfrak{B}_{j,m}^{a,(-j)}[h_m], \tag{132}$$

$$\mathbb{E}\Big[\Big\|\mathcal{B}_{j,m}^{a,(-j)}[h_m] - \mathfrak{B}_{j,m}^{a,(-j)}[h_m]\Big\|_2^2\ \Big|\ \mathscr{H}_m\Big] \le \frac{n-\kappa_n}{\kappa_n(n-1)} \cdot \frac{L_2^2\|h_m\|_2^2}{n}\sum_{i=1}^n x_{ij}^4\|\zeta_{i,m}^{a,(-j)}\|_2^2. \tag{133}$$

*Proof.* Fix $j, m$, and condition on $\mathscr{H}_m$. By Definition D.6, the arrays

$$\{x_{ij}F_{i,m}^{(-j)}\}_{i=1}^n, \qquad \{x_{ij}^2 L_{i,m}^{(-j)}\}_{i=1}^n, \qquad \{x_{ij}H_{i,m}^{a,(-j)}\}_{i=1}^n, \qquad \Big\{x_{ij}^2 D_1^2 f(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i)[h_m, \zeta_{i,m}^{a,(-j)}]\Big\}_{i=1}^n$$

are deterministic, while $B_m$ is a uniform $\kappa_n$-subset of $[n]$. Hence (126), (128), (130), and (132) follow immediately from (123) in Lemma D.7.

For the second-moment bounds, apply (125) in the appropriate Hilbert space:

- $\mathscr{E} = \mathbb{R}^k$ for $\mathcal{F}$, $\mathcal{H}$, and $\mathcal{B}[h_m]$;
- $\mathscr{E} = \mathbb{R}^{k\times k}$ with Frobenius norm for $\mathcal{A}$.

For $\mathcal{F}$, use $\|F_{i,m}^{(-j)}\|_2 \le L_0$ from Lemma D.2 to obtain

$$\|x_{ij}F_{i,m}^{(-j)}\|_2^2 \le L_0^2 x_{ij}^2,$$

which gives (127).

For $\mathcal{A}$, write $L_{i,m}^{(-j)}$ as a $k \times k$ matrix. Since $\|M\|_{\mathrm{F}} \le \sqrt{k}\|M\|_{\mathrm{op}}$ for all $M \in \mathbb{R}^{k\times k}$,

$$\|x_{ij}^2 L_{i,m}^{(-j)}\|_{\mathrm{F}}^2 \le kL_1^2 x_{ij}^4,$$

yielding (129).

For $\mathcal{H}$, use (111):

$$\|x_{ij}H_{i,m}^{a,(-j)}\|_2^2 \le L_1^2 x_{ij}^2\|\zeta_{i,m}^{a,(-j)}\|_2^2,$$

which gives (131).

Finally, for $\mathcal{B}[h_m]$,

$$\Big\|x_{ij}^2 D_1^2 f(\Xi_{i,m}^{(-j)}, W_i^\star, \varepsilon_i)[h_m, \zeta_{i,m}^{a,(-j)}]\Big\|_2^2 \le L_2^2\|h_m\|_2^2 x_{ij}^4\|\zeta_{i,m}^{a,(-j)}\|_2^2,$$

which proves (133). □

**Definition D.9** (Centered batch fluctuations)**.** For $j \in [d]$, $m \in \{0, \dots, M_T - 1\}$, and $a \in [q]$, define

$$\widetilde{\mathcal{F}}^{(-j)}_{j,m} := \mathcal{F}^{(-j)}_{j,m} - \mathfrak{F}^{(-j)}_{j,m}, \tag{134}$$
$$\widetilde{\mathcal{A}}^{(-j)}_{j,m} := \mathcal{A}^{(-j)}_{j,m} - \mathfrak{A}^{(-j)}_{j,m}, \tag{135}$$
$$\widetilde{\mathcal{H}}^{a,(-j)}_{j,m} := \mathcal{H}^{a,(-j)}_{j,m} - \mathfrak{H}^{a,(-j)}_{j,m}, \tag{136}$$
$$\widetilde{\mathcal{B}}^{a,(-j)}_{j,m}[h] := \mathcal{B}^{a,(-j)}_{j,m}[h] - \mathfrak{B}^{a,(-j)}_{j,m}[h], \qquad h \in \mathbb{R}^k. \tag{137}$$

**Corollary D.10** (Finite-horizon martingale bounds for accumulated batch fluctuations)**.** *Fix $j \in [d]$, $a \in [q]$, and $m \in \{0, \dots, M_T\}$. Define*

$$M^{\vartheta}_{j,m} := \sum_{r=0}^{m-1} \eta_r \, \widetilde{\mathcal{F}}^{(-j)}_{j,r}, \tag{138}$$
$$N^{A}_{j,m} := \sum_{r=0}^{m-1} \eta_r \, \widetilde{\mathcal{A}}^{(-j)}_{j,r}, \tag{139}$$
$$M^{u,a}_{j,m} := \sum_{r=0}^{m-1} \eta_r \, \widetilde{\mathcal{H}}^{a,(-j)}_{j,r}. \tag{140}$$

*Moreover, if $(h_r)_{r=0}^{M_T-1}$ is any $(\mathscr{H}_r)$-predictable process with values in $\mathbb{R}^k$, define*

$$N^{B,a}_{j,m}(h) := \sum_{r=0}^{m-1} \eta_r \, \widetilde{\mathcal{B}}^{a,(-j)}_{j,r}[h_r]. \tag{141}$$

*Then $(M^{\vartheta}_{j,m})_{m \le M_T}$, $(N^{A}_{j,m})_{m \le M_T}$, $(M^{u,a}_{j,m})_{m \le M_T}$, and $(N^{B,a}_{j,m}(h))_{m \le M_T}$ are martingales with respect to $(\mathscr{H}_m)_{m \ge 0}$, and they satisfy*

$$\mathbb{E}\Big[\sup_{m \le M_T} \|M^{\vartheta}_{j,m}\|_2^2\Big] \le 4 \sum_{r=0}^{M_T-1} \eta_r^2 \, \mathbb{E}\left[\frac{n-\kappa_n}{\kappa_n(n-1)} \cdot \frac{L_0^2}{n} \sum_{i=1}^n x_{ij}^2\right], \tag{142}$$
$$\mathbb{E}\Big[\sup_{m \le M_T} \|N^{A}_{j,m}\|_{\mathrm{F}}^2\Big] \le 4 \sum_{r=0}^{M_T-1} \eta_r^2 \, \mathbb{E}\left[\frac{n-\kappa_n}{\kappa_n(n-1)} \cdot \frac{kL_1^2}{n} \sum_{i=1}^n x_{ij}^4\right], \tag{143}$$
$$\mathbb{E}\Big[\sup_{m \le M_T} \|M^{u,a}_{j,m}\|_2^2\Big] \le 4 \sum_{r=0}^{M_T-1} \eta_r^2 \, \mathbb{E}\left[\frac{n-\kappa_n}{\kappa_n(n-1)} \cdot \frac{L_1^2}{n} \sum_{i=1}^n x_{ij}^2 \|\zeta^{a,(-j)}_{i,r}\|_2^2\right], \tag{144}$$
$$\mathbb{E}\Big[\sup_{m \le M_T} \|N^{B,a}_{j,m}(h)\|_2^2\Big] \le 4 \sum_{r=0}^{M_T-1} \eta_r^2 \, \mathbb{E}\left[\frac{n-\kappa_n}{\kappa_n(n-1)} \cdot \frac{L_2^2 \|h_r\|_2^2}{n} \sum_{i=1}^n x_{ij}^4 \|\zeta^{a,(-j)}_{i,r}\|_2^2\right]. \tag{145}$$

*Proof.* By (126), (128), (130), and (132), each increment has zero conditional mean given $\mathscr{H}_r$; since the increments are $\mathscr{H}_{r+1}$-measurable, the partial sums are martingales.

For the $L^2$ bounds, apply Doob's maximal inequality to each martingale:

$$\mathbb{E}\Big[\sup_{m \le M_T} \|M_m\|^2\Big] \le 4\mathbb{E}\|M_{M_T}\|^2.$$

Since the martingale increments are orthogonal in $L^2$,

$$\mathbb{E}\|M_{M_T}\|^2 = \sum_{r=0}^{M_T-1} \mathbb{E}\|M_{r+1} - M_r\|^2.$$

For $M^{\vartheta}_{j,m}$, the increment is $\eta_r \widetilde{\mathcal{F}}^{(-j)}_{j,r}$, so

$$\mathbb{E}\|M^{\vartheta}_{j,M_T}\|_2^2 = \sum_{r=0}^{M_T-1} \eta_r^2 \mathbb{E}\Big[\mathbb{E}\Big[\|\widetilde{\mathcal{F}}^{(-j)}_{j,r}\|_2^2 \,\Big|\, \mathscr{H}_r\Big]\Big].$$

Using (127) proves (142). The other three inequalities follow identically from (129), (131), and (133). □

*Remark* D.11 (Scaling consequence). Corollary D.10 already reflects the correct high-dimensional scaling. Indeed, under Assumption A.3,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2\right] = d^{-1}, \qquad \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_{ij}^4\right] \lesssim d^{-2},$$

while $\eta_r \asymp n^{\alpha}$, $\kappa_n \asymp n^{\alpha}$, and $M_T \asymp n^{1-\alpha}$. Hence the right-hand side of (142) is of order

$$M_T \cdot \frac{\eta_r^2}{\kappa_n} \cdot d^{-1} \asymp n^{1-\alpha} \cdot n^{\alpha} \cdot d^{-1} \asymp 1,$$

whereas the right-hand side of (143) is of order

$$M_T \cdot \frac{\eta_r^2}{\kappa_n} \cdot d^{-2} \asymp d^{-1} \to 0.$$

Thus, over $O(1)$ epochs, the batch fluctuations of the row-level forcing are macroscopic, while the batch fluctuations of the local linear operator are strictly smaller. This separation is the deterministic origin of the later Poissonian forcing but deterministic response structure in the effective theory.

*Remark* D.12 (Where overlap self-averaging enters). The results of this appendix do not yet prove self-averaging of the empirical overlap arrays $C^{\vartheta,\vartheta}, C^{u,\vartheta}, C^{u,u}$ from Definition C.11. What they do provide are the two quantitative ingredients needed for that later step:

1. row-wise a priori envelopes (120)–(121),
2. uniform $L^2$ control of the accumulated batch martingales (142)–(145).

Appendix F will combine these with the small-backreaction estimates from the cavity method to deduce the actual self-averaging of the replicated overlap process.

## E. Well-Posedness of the Replicated Effective System

This appendix isolates the deterministic fixed-point problem underlying the replicated tangent effective theory. The iterate-level multi-pass SGD DMFT of Fan & Wang (2026) and the corresponding SGF theory of Nishiyama & Imaizumi (2026) both rely on a finite-horizon Volterra-type closure at the level of a low-dimensional kernel tuple. The present appendix shows that, once the iterate effective environment is frozen, the replicated tangent extension admits an analogous block-triangular closure. The outcome is a unique finite-horizon replicated kernel tuple obtained as the fixed point of a contractive Volterra map.

The point of organizing the argument at the kernel level is twofold. First, it separates the abstract well-posedness problem from the high-dimensional approximation argument of Appendix F. Second, it makes transparent that the tangent sector is *causal* and *affine-linear* once the iterate sector has already been solved.

Throughout this appendix, fix a horizon $T > 0$, a replica number $q \geq 1$, and let

$$\bar{\eta}_T := \sup_{t\in[0,T]} \bar{\eta}(t) < \infty$$

be the finite-horizon stepsize envelope from Assumption A.2. We also fix once and for all the iterate-level effective process

$$(\vartheta_t, \xi_t, \vartheta_\star, w_\star, \varepsilon)_{t\in[0,T]}$$

furnished by Proposition A.9. Define the imported iterate kernels

$$C_{t,s}^{\vartheta,\vartheta} := \mathbb{E}[\vartheta_t \otimes \vartheta_s] \in \mathbb{R}^{k\times k}, \qquad C_t^{\vartheta,\star} := \mathbb{E}[\vartheta_t \otimes \vartheta_\star] \in \mathbb{R}^{k\times k_\star}, \tag{146}$$

and their finite-horizon bounds

$$C_{\vartheta,T} := \sup_{t,s\in[0,T]} \|C_{t,s}^{\vartheta,\vartheta}\|_{\mathrm{F}}, \qquad C_{\star,T} := \sup_{t\in[0,T]} \|C_t^{\vartheta,\star}\|_{\mathrm{F}}. \tag{147}$$

These quantities are finite by Proposition A.9 and the moment assumptions of Appendix A.

### E.1. Kernel Space and Metric Structure

We work with a weighted supremum metric adapted to causal Volterra operators.

**Definition E.1** (Weighted kernel norms)**.** Let $K : [0,T]^2 \to \mathbb{R}^{m\times n}$ be a bounded measurable matrix-valued kernel, and let $u : [0,T] \to \mathbb{R}^{m\times n}$ be a bounded measurable matrix-valued trajectory. For $\lambda > 0$, define

$$\|K\|_{\lambda,T}^{(2)} := \sup_{(t,s)\in[0,T]^2} e^{-\lambda\max\{t,s\}} \|K_{t,s}\|_{\mathrm{F}}, \tag{148}$$

$$\|u\|_{\lambda,T}^{(1)} := \sup_{t\in[0,T]} e^{-\lambda t} \|u_t\|_{\mathrm{F}}. \tag{149}$$

**Definition E.2** (Replicated state-kernel space)**.** Let $\mathfrak{C}_T^{(q)}$ denote the collection of all $q\times q$ block families

$$\mathbf{Q} = (Q^{ab})_{a,b\in[q]}$$

such that each $Q^{ab} : [0,T]^2 \to \mathbb{R}^{k\times k}$ is bounded measurable and satisfies the symmetry constraint

$$Q_{t,s}^{ab} = (Q_{s,t}^{ba})^\top \qquad \text{for all } a,b\in[q],\ t,s\in[0,T]. \tag{150}$$

Let $\mathfrak{S}_T^{(q)}$ be the collection of all families

$$\mathbf{S} = (S^a)_{a\in[q]}, \qquad S^a : [0,T]^2 \to \mathbb{R}^{k\times k}$$

of bounded measurable kernels, and let $\mathfrak{M}_T^{(q)}$ be the collection of all families

$$\mathbf{M} = (M^a)_{a\in[q]}, \qquad M^a : [0,T] \to \mathbb{R}^{k\times k_\star}$$

of bounded measurable trajectories. Finally, let $\mathfrak{Z}_T^{(q)}$ denote the collection of all $q\times q$ block families

$$\mathbf{Z} = (Z^{ab})_{a,b\in[q]}$$

such that each $Z^{ab} : [0,T]^2 \to \mathbb{R}^{k\times k}$ is bounded measurable and satisfies

$$Z_{t,s}^{ab} = (Z_{s,t}^{ba})^\top \qquad \text{for all } a,b\in[q],\ t,s\in[0,T]. \tag{151}$$

The *replicated state-kernel space* is

$$\mathfrak{X}_T^{(q)} := \mathfrak{C}_T^{(q)} \times \mathfrak{S}_T^{(q)} \times \mathfrak{M}_T^{(q)} \times \mathfrak{Z}_T^{(q)}. \tag{152}$$

For $\mathfrak{X} = (\mathbf{Q},\mathbf{S},\mathbf{M},\mathbf{Z}) \in \mathfrak{X}_T^{(q)}$, define

$$\|\mathbf{Q}\|_{\lambda,T} := \max_{a,b\in[q]} \|Q^{ab}\|_{\lambda,T}^{(2)}, \tag{153}$$

$$\|\mathbf{S}\|_{\lambda,T} := \max_{a\in[q]} \|S^{a}\|_{\lambda,T}^{(2)}, \tag{154}$$

$$\|\mathbf{M}\|_{\lambda,T} := \max_{a\in[q]} \|M^{a}\|_{\lambda,T}^{(1)}, \tag{155}$$

$$\|\mathbf{Z}\|_{\lambda,T} := \max_{a,b\in[q]} \|Z^{ab}\|_{\lambda,T}^{(2)}, \tag{156}$$

and

$$\|\mathfrak{X}\|_{\lambda,T} := \max\big\{\|\mathbf{Q}\|_{\lambda,T}, \|\mathbf{S}\|_{\lambda,T}, \|\mathbf{M}\|_{\lambda,T}, \|\mathbf{Z}\|_{\lambda,T}\big\}. \tag{157}$$

Then $(\mathfrak{X}_T^{(q)}, \|\cdot\|_{\lambda,T})$ is a Banach space.

*Proof.* Each component space is a closed subspace of a finite product of weighted $L^\infty$ spaces, and the norm (157) is the maximum of the corresponding component norms. Completeness follows immediately. □

**Definition E.3** (Boundary data induced by the initial tangent law)**.** Couple the initial replica row $(u_0^1, \dots, u_0^q, \vartheta_0, \vartheta_\star)$ according to $\nu_{0,\star,q}$ from Assumption A.5, and let $(\vartheta_t)_{t\in[0,T]}$ denote the imported iterate effective process started from $(\vartheta_0, \vartheta_\star)$. Define the initial covariance and cross-covariance data

$$Q_0^{ab} := \mathbb{E}[u_0^a \otimes u_0^b] \in \mathbb{R}^{k\times k}, \tag{158}$$

$$M_0^a := \mathbb{E}[u_0^a \otimes \vartheta_\star] \in \mathbb{R}^{k\times k_\star}, \tag{159}$$

$$S_{0,s}^a := \mathbb{E}[u_0^a \otimes \vartheta_s] \in \mathbb{R}^{k\times k}, \qquad s \in [0,T]. \tag{160}$$

Set

$$R_{0,T}^{(q)} := \max\Big\{ \max_{a,b\in[q]} \|Q_0^{ab}\|_{\mathrm{F}}, \max_{a\in[q]} \|M_0^a\|_{\mathrm{F}}, \max_{a\in[q]} \sup_{s\in[0,T]} \|S_{0,s}^a\|_{\mathrm{F}} \Big\}. \tag{161}$$

*Remark* E.4. The quantity $R_{0,T}^{(q)}$ is finite. Indeed, Assumption A.5 gives finite moments of all orders for the initial tangent and teacher rows, while Proposition A.9 yields finite second moments for $\vartheta_s$ uniformly on compact horizons.

The closure theorem below is abstracted around four deterministic objects: a self-interaction drift $A$, three tangent coefficient maps $B, J, H$, and a bounded linear lifting operator $\mathscr{G}_T$ that reconstructs the projected field kernels from the state kernels. The exact formulas are model-dependent and will be verified in Appendix F; the present appendix only uses their structural properties.

**Definition E.5** (Admissible coefficient family)**.** A family

$$\mathfrak{F}_T^{(q)} = \big(A,\ \{B^a\}_{a\in[q]},\ \{J^a\}_{a\in[q]},\ \{H^{ab}\}_{a,b\in[q]},\ \mathscr{G}_T\big)$$

is called *admissible on* $[0,T]$ if the following conditions hold.

1. **Drift envelope.** $A : [0,T] \to \mathbb{R}^{k\times k}$ is bounded measurable, and

$$A_T := \sup_{t\in[0,T]} \|A_t\|_{\mathrm{op}} < \infty. \tag{162}$$

2. **Causality.** For each $a, b \in [q]$, the coefficient maps

$$B^a, J^a, H^{ab} : \mathfrak{X}_T^{(q)} \to L^\infty([0,T])$$

are nonanticipative in the following sense: if $\mathfrak{X}, \widetilde{\mathfrak{X}} \in \mathfrak{X}_T^{(q)}$ agree on the time slab $[0,t]$, then the values

$$B_t^a(\mathfrak{X}) = B_t^a(\widetilde{\mathfrak{X}}), \quad J_t^a(\mathfrak{X}) = J_t^a(\widetilde{\mathfrak{X}}), \quad H_t^{ab}(\mathfrak{X}) = H_t^{ab}(\widetilde{\mathfrak{X}})$$

coincide.

3. **Uniform boundedness on balls.** For every $R > 0$, there exists a finite constant $C_{R,T}$ such that for all $\lambda \ge 1$, all $\mathfrak{X} \in \mathfrak{X}_T^{(q)}$ with $\|\mathfrak{X}\|_{\lambda,T} \le R$, and all $t \in [0,T]$,

$$\|B_t^a(\mathfrak{X})\|_{\mathrm{op}} \le C_{R,T}, \tag{163}$$

$$\|J_t^a(\mathfrak{X})\|_{\mathrm{F}} \le C_{R,T}, \tag{164}$$

$$\|H_t^{ab}(\mathfrak{X})\|_{\mathrm{F}} \le C_{R,T}. \tag{165}$$

4. **Weighted Lipschitz property on balls.** For every $R > 0$, there exists a finite constant $L_{R,T}$ such that for all $\lambda \ge 1$, all $\mathfrak{X}, \widetilde{\mathfrak{X}} \in \mathfrak{X}_T^{(q)}$ with $\max\{\|\mathfrak{X}\|_{\lambda,T}, \|\widetilde{\mathfrak{X}}\|_{\lambda,T}\} \le R$, and all $t \in [0,T]$,

$$e^{-\lambda t}\|B_t^a(\mathfrak{X}) - B_t^a(\widetilde{\mathfrak{X}})\|_{\mathrm{op}} \le L_{R,T}\, \|\mathfrak{X} - \widetilde{\mathfrak{X}}\|_{\lambda,T}, \tag{166}$$

$$e^{-\lambda t}\|J_t^a(\mathfrak{X}) - J_t^a(\widetilde{\mathfrak{X}})\|_{\mathrm{F}} \le L_{R,T}\, \|\mathfrak{X} - \widetilde{\mathfrak{X}}\|_{\lambda,T}, \tag{167}$$

$$e^{-\lambda t}\|H_t^{ab}(\mathfrak{X}) - H_t^{ab}(\widetilde{\mathfrak{X}})\|_{\mathrm{F}} \le L_{R,T}\, \|\mathfrak{X} - \widetilde{\mathfrak{X}}\|_{\lambda,T}. \tag{168}$$

5. **Symmetry of the noise kernel.** For every $\mathfrak{X} \in \mathfrak{X}_T^{(q)}$, every $t \in [0,T]$, and all $a,b \in [q]$,

$$H_t^{ab}(\mathfrak{X}) = H_t^{ba}(\mathfrak{X})^\top. \tag{169}$$

6. **Bounded linear lifting.** $\mathscr{G}_T$ is a bounded linear map

$$\mathscr{G}_T : \mathfrak{C}_T^{(q)} \times \mathfrak{S}_T^{(q)} \times \mathfrak{M}_T^{(q)} \to \mathfrak{Z}_T^{(q)} \tag{170}$$

that preserves the symmetry constraint (151). Its operator norm with respect to the weighted norms is denoted by

$$G_T := \sup_{\lambda \geq 1} \sup_{(\mathbf{Q},\mathbf{S},\mathbf{M}) \neq 0} \frac{\|\mathscr{G}_T(\mathbf{Q},\mathbf{S},\mathbf{M})\|_{\lambda,T}}{\max\{\|\mathbf{Q}\|_{\lambda,T}, \|\mathbf{S}\|_{\lambda,T}, \|\mathbf{M}\|_{\lambda,T}\}} < \infty. \tag{171}$$

*Remark* E.6 (Interpretation of the lifting operator). The map $\mathscr{G}_T$ packages the iterate-response structure inherited from the imported iterate effective theory. Concretely, it sends the tangent state-kernel tuple $(\mathbf{Q},\mathbf{S},\mathbf{M})$ to the corresponding projected-field kernels $\mathbf{Z}$. The exact formula of $\mathscr{G}_T$ is not needed for the fixed-point argument below; only linearity, boundedness, and symmetry preservation are used.

**Definition E.7** (Replicated Volterra map)**.** Fix an admissible coefficient family $\mathfrak{F}_T^{(q)}$ in the sense of Definition E.5. The associated replicated Volterra map

$$\mathcal{T}_T^{(q)} : \mathfrak{X}_T^{(q)} \to \mathfrak{X}_T^{(q)}$$

is defined as follows. For an input

$$\mathfrak{X} = (\mathbf{Q},\mathbf{S},\mathbf{M},\mathbf{Z}),$$

write

$$\mathcal{T}_T^{(q)}(\mathfrak{X}) = (\mathbf{Q}',\mathbf{S}',\mathbf{M}',\mathbf{Z}').$$

Then, for every $a,b \in [q]$ and $t,s \in [0,T]$,

$$M_t^{a,\prime} = M_0^a - \int_0^t \bar{\eta}(r)\Big(A_r M_r^a + B_r^a(\mathfrak{X})\, C_r^{\vartheta,\star}\Big)\, dr, \tag{172}$$

$$S_{t,s}^{a,\prime} = S_{0,s}^a - \int_0^t \bar{\eta}(r)\Big(A_r S_{r,s}^a + B_r^a(\mathfrak{X})\, C_{r,s}^{\vartheta,\vartheta}\Big)\, dr + \int_0^{t\wedge s} \bar{\eta}(r)^2 J_r^a(\mathfrak{X})\, dr, \tag{173}$$

$$\begin{aligned} Q_{t,s}^{ab,\prime} &= Q_0^{ab} - \int_0^t \bar{\eta}(r)\Big(A_r Q_{r,s}^{ab} + B_r^a(\mathfrak{X})\,(S_{s,r}^b)^\top\Big)\, dr \\ &\quad - \int_0^s \bar{\eta}(r)\Big(Q_{t,r}^{ab} A_r^\top + S_{t,r}^a\, B_r^b(\mathfrak{X})^\top\Big)\, dr + \int_0^{t\wedge s} \bar{\eta}(r)^2 H_r^{ab}(\mathfrak{X})\, dr, \end{aligned} \tag{174}$$

$$\mathbf{Z}' = \mathscr{G}_T(\mathbf{Q}',\mathbf{S}',\mathbf{M}'). \tag{175}$$

**Lemma E.8** (Weighted Volterra estimates)**.** *Let $\lambda > 0$. Then:*

$$\sup_{t\in[0,T]} e^{-\lambda t} \int_0^t e^{\lambda r}\, dr \leq \frac{1}{\lambda}, \tag{176}$$

$$\sup_{t,s\in[0,T]} e^{-\lambda \max\{t,s\}} \int_0^t e^{\lambda \max\{r,s\}}\, dr \leq \frac{1}{\lambda} + T, \tag{177}$$

$$\sup_{t,s\in[0,T]} e^{-\lambda \max\{t,s\}} \int_0^{t\wedge s} e^{\lambda r}\, dr \leq \frac{1}{\lambda}. \tag{178}$$

*Moreover, if $r \leq t$, then*

$$e^{-\lambda t} \leq e^{-\lambda(t-r)} e^{-\lambda r}, \qquad e^{-\lambda \max\{t,s\}} \leq e^{-\lambda(\max\{t,s\}-r)} e^{-\lambda r}. \tag{179}$$

*Proof.* The factorization (179) is immediate. For (176),

$$e^{-\lambda t}\int_0^t e^{\lambda r}\,dr=\int_0^t e^{-\lambda(t-r)}\,dr\le\int_0^\infty e^{-\lambda u}\,du=\frac{1}{\lambda}.$$

Likewise,

$$e^{-\lambda\max\{t,s\}}\int_0^{t\wedge s} e^{\lambda r}\,dr\le\int_0^{t\wedge s} e^{-\lambda(\max\{t,s\}-r)}\,dr\le\int_0^\infty e^{-\lambda u}\,du=\frac{1}{\lambda},$$

which proves (178).

For (177), split the integral at $s$. If $t\le s$, then

$$e^{-\lambda s}\int_0^t e^{\lambda s}\,dr=t\le T.$$

If $t>s$, then

$$e^{-\lambda t}\int_0^s e^{\lambda s}\,dr+e^{-\lambda t}\int_s^t e^{\lambda r}\,dr\le se^{-\lambda(t-s)}+\frac{1}{\lambda}\le T+\frac{1}{\lambda}.$$

Taking the supremum proves (177). □

**Definition E.9** (Closed weighted balls with prescribed boundary data)**.** Fix $R>0$ and $\lambda\ge 1$. Define $\mathfrak{B}^{(q)}_{R,\lambda,T}\subseteq\mathfrak{X}^{(q)}_T$ to be the set of all $\mathfrak{X}=(\mathbf{Q},\mathbf{S},\mathbf{M},\mathbf{Z})$ such that:

1. $\|\mathfrak{X}\|_{\lambda,T}\le R$,
2. $\mathbf{Q}$ and $\mathbf{Z}$ satisfy the symmetry constraints (150) and (151),
3. $Q^{ab}_{0,0}=Q^{ab}_0$ for all $a,b\in[q]$,
4. $M^a_0=M^a_0$ for all $a\in[q]$,
5. $S^a_{0,s}=S^a_{0,s}$ for all $a\in[q]$ and $s\in[0,T]$.

This is a closed subset of the Banach space $\mathfrak{X}^{(q)}_T$.

## E.2. Contraction of the Replicated Map

We now prove the central deterministic result of the paper's tangent sector.

**Proposition E.10** (Basic well-posedness of the Volterra map)**.** *For every admissible coefficient family $\mathfrak{F}^{(q)}_T$, the map $\mathcal{T}^{(q)}_T$ of Definition E.7 is well-defined from $\mathfrak{X}^{(q)}_T$ into itself. Moreover, if $\mathfrak{X}\in\mathfrak{X}^{(q)}_T$ satisfies the symmetry constraints (150) and (151), then so does $\mathcal{T}^{(q)}_T(\mathfrak{X})$.*

*Proof.* The right-hand sides of (172)–(174) are finite because the input kernels are bounded measurable, the coefficient maps are bounded on bounded sets, and $\bar\eta$ is bounded on $[0,T]$. Hence $\mathbf{M}',\mathbf{S}',\mathbf{Q}'$ are bounded measurable. Then $\mathbf{Z}'=\mathscr{G}_T(\mathbf{Q}',\mathbf{S}',\mathbf{M}')$ belongs to $\mathfrak{Z}^{(q)}_T$ by bounded linearity of $\mathscr{G}_T$.

It remains to check symmetry. The symmetry of $\mathbf{Z}'$ follows by assumption on $\mathscr{G}_T$. To prove the symmetry of $\mathbf{Q}'$, fix $a,b\in[q]$ and transpose (174) after swapping $t$ and $s$. Using the symmetry of the input $\mathbf{Q}$, the identity $(S^a_{r,t})^\top=(S^a_{t,r})^\top$, and $H^{ab}_r(\mathfrak{X})=H^{ba}_r(\mathfrak{X})^\top$, we obtain

$$(Q^{ba,\prime}_{s,t})^\top=Q^{ab,\prime}_{t,s}.$$

Hence $\mathbf{Q}'$ satisfies (150). □

**Theorem E.11** (Invariant ball and contraction)**.** *Fix $T > 0$, $q \geq 1$, and an admissible coefficient family $\mathfrak{F}_T^{(q)}$. Let $R > 2(1+G_T)R_{0,T}^{(q)}$. Then there exists a finite threshold*

$$\lambda_\star = \lambda_\star(R, T, q, \mathfrak{F}_T^{(q)}) \geq 1$$

*such that for every $\lambda \geq \lambda_\star$, the map $\mathcal{T}_T^{(q)}$ sends $\mathfrak{B}_{R,\lambda,T}^{(q)}$ into itself and is a strict contraction on that set with respect to the metric induced by $\|\cdot\|_{\lambda,T}$. Consequently, $\mathcal{T}_T^{(q)}$ has a unique fixed point*

$$\mathfrak{X}_{T,\star}^{(q)} = (\mathbf{Q}_\star, \mathbf{S}_\star, \mathbf{M}_\star, \mathbf{Z}_\star) \in \mathfrak{B}_{R,\lambda,T}^{(q)}.$$

*Proof.* Write

$$\mathfrak{X} = (\mathbf{Q}, \mathbf{S}, \mathbf{M}, \mathbf{Z}) \in \mathfrak{B}_{R,\lambda,T}^{(q)}, \qquad \mathcal{T}_T^{(q)}(\mathfrak{X}) = (\mathbf{Q}', \mathbf{S}', \mathbf{M}', \mathbf{Z}').$$

Let $C_{R,T}$ and $L_{R,T}$ be the boundedness and Lipschitz constants from Definition E.5 for the radius $R$.

**Step 1: invariance of the ball.** We first bound the image of $\mathfrak{X}$.

For $\mathbf{M}'$, using (172), (179), the bound $\|\mathbf{M}\|_{\lambda,T} \leq R$, the iterate kernel bound $\|C_r^{\vartheta,\star}\|_{\mathrm{F}} \leq C_{\star,T}$, the coefficient envelope $\|B_r^a(\mathfrak{X})\|_{\mathrm{op}} \leq C_{R,T}$, and Lemma E.8, we obtain

$$\|\mathbf{M}'\|_{\lambda,T} \leq R_{0,T}^{(q)} + \frac{\bar{\eta}_T}{\lambda}\Big(A_T R + C_{\star,T} C_{R,T}\Big). \tag{180}$$

For $\mathbf{S}'$, use (173), $\|\mathbf{S}\|_{\lambda,T} \leq R$, the iterate kernel bound $\|C_{r,s}^{\vartheta,\vartheta}\|_{\mathrm{F}} \leq C_{\vartheta,T}$, and (176)–(178). This yields

$$\|\mathbf{S}'\|_{\lambda,T} \leq R_{0,T}^{(q)} + \frac{\bar{\eta}_T}{\lambda}\Big(A_T R + C_{\vartheta,T} C_{R,T}\Big) + \frac{\bar{\eta}_T^2}{\lambda} C_{R,T}. \tag{181}$$

For $\mathbf{Q}'$, use (174), the bound $\|\mathbf{Q}\|_{\lambda,T} \leq R$, the coefficient envelopes $\|B_r^a(\mathfrak{X})\|_{\mathrm{op}} \leq C_{R,T}$ and $\|H_r^{ab}(\mathfrak{X})\|_{\mathrm{F}} \leq C_{R,T}$, and again Lemma E.8. This gives

$$\|\mathbf{Q}'\|_{\lambda,T} \leq R_{0,T}^{(q)} + \frac{2\bar{\eta}_T}{\lambda}\Big(A_T R + C_{R,T} R\Big) + \frac{\bar{\eta}_T^2}{\lambda} C_{R,T}. \tag{182}$$

Finally, bounded linearity of $\mathscr{G}_T$ gives

$$\|\mathbf{Z}'\|_{\lambda,T} \leq G_T \max\Big\{\|\mathbf{Q}'\|_{\lambda,T}, \|\mathbf{S}'\|_{\lambda,T}, \|\mathbf{M}'\|_{\lambda,T}\Big\}. \tag{183}$$

Collecting (180)–(183), there exists a finite constant $K_{R,T}$ such that

$$\|\mathcal{T}_T^{(q)}(\mathfrak{X})\|_{\lambda,T} \leq (1+G_T)R_{0,T}^{(q)} + \frac{K_{R,T}}{\lambda}. \tag{184}$$

Since $R > 2(1+G_T)R_{0,T}^{(q)}$, we may choose $\lambda$ sufficiently large so that

$$(1+G_T)R_{0,T}^{(q)} + \frac{K_{R,T}}{\lambda} \leq R.$$

Thus $\mathcal{T}_T^{(q)}$ maps $\mathfrak{B}_{R,\lambda,T}^{(q)}$ into itself for all sufficiently large $\lambda$.

**Step 2: contraction estimate.** Take now

$$\mathfrak{X} = (\mathbf{Q}, \mathbf{S}, \mathbf{M}, \mathbf{Z}), \qquad \widetilde{\mathfrak{X}} = (\widetilde{\mathbf{Q}}, \widetilde{\mathbf{S}}, \widetilde{\mathbf{M}}, \widetilde{\mathbf{Z}})$$

in $\mathfrak{B}^{(q)}_{R,\lambda,T}$, and write

$$\delta := \|\mathfrak{X} - \widetilde{\mathfrak{X}}\|_{\lambda,T}.$$

Denote

$$\mathcal{T}^{(q)}_T(\mathfrak{X}) = (\mathbf{Q}', \mathbf{S}', \mathbf{M}', \mathbf{Z}'), \qquad \mathcal{T}^{(q)}_T(\widetilde{\mathfrak{X}}) = (\widetilde{\mathbf{Q}}', \widetilde{\mathbf{S}}', \widetilde{\mathbf{M}}', \widetilde{\mathbf{Z}}').$$

For $\mathbf{M}' - \widetilde{\mathbf{M}}'$, subtract (172) for the two inputs. Using (166), $\|C^{\vartheta,\star}_r\|_{\mathrm{F}} \le C_{\star,T}$, and (176), we obtain

$$\|\mathbf{M}' - \widetilde{\mathbf{M}}'\|_{\lambda,T} \le \frac{\bar{\eta}_T}{\lambda}\Big(A_T + C_{\star,T} L_{R,T}\Big)\delta. \tag{185}$$

For $\mathbf{S}' - \widetilde{\mathbf{S}}'$, subtract (173) and use (166), (167), $\|C^{\vartheta,\vartheta}_{r,s}\|_{\mathrm{F}} \le C_{\vartheta,T}$, and Lemma E.8. This yields

$$\|\mathbf{S}' - \widetilde{\mathbf{S}}'\|_{\lambda,T} \le \frac{\bar{\eta}_T}{\lambda}\Big(A_T + C_{\vartheta,T} L_{R,T}\Big)\delta + \frac{\bar{\eta}_T^2}{\lambda} L_{R,T}\delta. \tag{186}$$

For $\mathbf{Q}' - \widetilde{\mathbf{Q}}'$, subtract (174). In the two mixed drift terms, add and subtract $B^a_r(\mathfrak{X})(\widetilde{S}^b_{s,r})^\top$ and $\widetilde{S}^a_{t,r} B^b_r(\widetilde{\mathfrak{X}})^\top$, respectively, to separate the difference into a term involving $\mathbf{S} - \widetilde{\mathbf{S}}$ and a term involving $B(\mathfrak{X}) - B(\widetilde{\mathfrak{X}})$. Using the ball bound $\|\widetilde{\mathbf{S}}\|_{\lambda,T} \le R$, the coefficient envelopes $\|B^a_r(\widetilde{\mathfrak{X}})\|_{\mathrm{op}} \le C_{R,T}$, the Lipschitz bounds (166) and (168), and Lemma E.8, we obtain

$$\|\mathbf{Q}' - \widetilde{\mathbf{Q}}'\|_{\lambda,T} \le \frac{2\bar{\eta}_T}{\lambda}\Big(A_T + C_{R,T} + R L_{R,T}\Big)\delta + \frac{\bar{\eta}_T^2}{\lambda} L_{R,T}\delta. \tag{187}$$

Finally, by bounded linearity of $\mathscr{G}_T$,

$$\|\mathbf{Z}' - \widetilde{\mathbf{Z}}'\|_{\lambda,T} \le G_T \max\Big\{\|\mathbf{Q}' - \widetilde{\mathbf{Q}}'\|_{\lambda,T}, \|\mathbf{S}' - \widetilde{\mathbf{S}}'\|_{\lambda,T}, \|\mathbf{M}' - \widetilde{\mathbf{M}}'\|_{\lambda,T}\Big\}. \tag{188}$$

Combining (185)–(188), there exists a finite constant $L^\sharp_{R,T}$ such that

$$\|\mathcal{T}^{(q)}_T(\mathfrak{X}) - \mathcal{T}^{(q)}_T(\widetilde{\mathfrak{X}})\|_{\lambda,T} \le \frac{L^\sharp_{R,T}}{\lambda}\, \|\mathfrak{X} - \widetilde{\mathfrak{X}}\|_{\lambda,T}. \tag{189}$$

Choosing $\lambda$ large enough so that

$$\frac{L^\sharp_{R,T}}{\lambda} < 1$$

proves that $\mathcal{T}^{(q)}_T$ is a strict contraction on $\mathfrak{B}^{(q)}_{R,\lambda,T}$.

**Step 3: fixed point.** The set $\mathfrak{B}^{(q)}_{R,\lambda,T}$ is closed in the Banach space $\mathfrak{X}^{(q)}_T$, and $\mathcal{T}^{(q)}_T$ is a contraction from that closed set to itself. Banach's fixed-point theorem therefore yields a unique fixed point in $\mathfrak{B}^{(q)}_{R,\lambda,T}$. □

**Corollary E.12** (Lipschitz dependence on boundary data and iterate input)**.** *Fix $R$ and $\lambda \ge \lambda_\star$ as in Theorem E.11. Then the unique fixed point $\mathfrak{X}^{(q)}_{T,\star}$ depends Lipschitz continuously on:*

1. *the boundary data $(Q^{ab}_0)_{a,b\in[q]}$, $(M^a_0)_{a\in[q]}$, and $(S^a_{0,\cdot})_{a\in[q]}$,*
2. *the imported iterate kernels $C^{\vartheta,\vartheta}$ and $C^{\vartheta,\star}$,*
3. *the admissible coefficient family $\mathfrak{F}^{(q)}_T$,*

*provided the corresponding perturbations preserve admissibility with the same radius $R$ and sufficiently close constants.*

*Proof.* The proof is the standard perturbation argument for fixed points of strict contractions. Indeed, if $\mathcal{T}$ and $\widetilde{\mathcal{T}}$ are two contractions on the same complete metric space with a common contraction factor $\rho < 1$, and $x_\star, \widetilde{x}_\star$ denote their fixed points, then

$$\|x_\star - \widetilde{x}_\star\| \le \frac{1}{1-\rho} \sup_{x \in \mathfrak{B}^{(q)}_{R,\lambda,T}} \|\mathcal{T}(x) - \widetilde{\mathcal{T}}(x)\|.$$

Applying this estimate with (172)–(175) gives the claimed Lipschitz dependence. □

*Remark* E.13 (How Appendix E is used later). Appendix E is deliberately abstract. The exact coefficient family $\mathfrak{F}^{(q)}_T$ induced by the augmented cavity expansion of Appendix C will be identified in Appendix F, and the boundedness and weighted Lipschitz estimates required in Definition E.5 will be verified there using the regularity assumptions on $f$ and $g$, together with the finite-horizon bounds of Appendix D. At that point, Theorem E.11 will upgrade from an abstract Volterra statement to the concrete well-posedness theorem for the replicated tangent effective system announced in the main text.

## F. Proof of the Finite-Horizon Replicated Tangent $\mathrm{DMFT}$

This appendix proves the finite-horizon replicated tangent $\mathrm{DMFT}$, i.e. the kernel-level theorem announced as Theorem 3.1 in the main text. The proof proceeds in three layers.

1. We first instantiate the abstract Volterra theory of Appendix E with a concrete coefficient family induced by the augmented cavity expansion of Appendix C. This is the only place where we impose two extra assumptions beyond Appendix A: exact Gaussian design and an affine row-regularizer. These hypotheses guarantee that the field-side lift is exactly Gaussian and that the tangent drift closes on second-order kernels.

2. We then prove a *small-backreaction* estimate for the cavity error terms $\delta^{(-j)}_{i,m}$ and $\nu^{a,(-j)}_{i,m}$, together with the induced remainder terms $\mathcal{E}^{\vartheta,(-j)}$ and $\mathcal{E}^{u,a,(-j)}$. This is the step that justifies replacing the exact finite-$n, d$ cavity recursion by its deterministic kernel closure.

3. Finally, we show that the empirical replicated kernel tuple is an approximate fixed point of the deterministic Volterra map, with an error vanishing uniformly on compact time intervals. Contraction of the map from Appendix E then yields convergence to the unique fixed point.

The iterate-only version of this strategy is standard in high-dimensional multi-index $\mathrm{SGD}$ $\mathrm{DMFT}$ (Fan & Wang, 2026). What is new here is that the tangent sector is carried along *on the same batch noise*, which creates the bilinear operator $\mathcal{B}^{a,(-j)}$ and the mixed kernel family $(\mathbf{Q}, \mathbf{S}, \mathbf{M})$.

### F.1. Coupling and Replacement Scheme

We now make the additional structural assumptions under which the concrete replicated closure is exact at the second-order level.

**Assumption F.1** (Gaussian design and affine row-regularizer)**.** Throughout Appendix F, we strengthen Assumptions A.3–A.4 as follows.

1. The covariates are exactly Gaussian:
$$x_i \sim \mathcal{N}(0, I_d/d), \qquad i \in [n],$$
independently across $i$.

2. The row-regularizer gradient is affine:
$$g(u) = \Gamma u + b, \qquad u \in \mathbb{R}^k,$$
for some deterministic matrix $\Gamma \in \mathbb{R}^{k\times k}$ and vector $b \in \mathbb{R}^k$. In particular,
$$Dg(u) = \Gamma \qquad \text{for all } u \in \mathbb{R}^k.$$

*Remark* F.2 (Scope of Assumption F.1). Assumption F.1 includes the ridge-regularized models used in the explicit nonlinear theorem of Appendix J. The Gaussian-design restriction is imposed only to keep the field-lift exact at the kernel level; the non-Gaussian product-design extension can be obtained by the same Lindeberg replacement used in the iterate-only theory (Fan & Wang, 2026), but that extension is orthogonal to the novel tangent-sector argument and is therefore not pursued here.

The Gaussian design implies that field-side covariances are exactly equal to the corresponding row-side second moments. This makes the lifting operator from Appendix E particularly simple.

**Definition F.3** (Teacher second moment)**.** Let

$$C^{\star,\star} := \mathbb{E}[\vartheta_\star \otimes \vartheta_\star] \in \mathbb{R}^{k_\star \times k_\star}, \tag{190}$$

where $\vartheta_\star$ is the limiting teacher row variable from Appendix A.

**Definition F.4** (Physical replicated kernel tuples)**.** Fix $T > 0$ and $q \geq 1$. A kernel tuple

$$\mathfrak{X} = (\mathbf{Q}, \mathbf{S}, \mathbf{M}, \mathbf{Z}) \in \mathfrak{X}_T^{(q)}$$

is called *physical on* $[0, T]$ if the following hold.

1. $\mathbf{Z} = \mathbf{Q}$.
2. For every $t \in [0, T]$, the block matrix

$$\Sigma_t(\mathfrak{X}) := \begin{pmatrix} \left[Q_{t,t}^{ab}\right]_{a,b\in[q]} & \left[S_{t,t}^{a}\right]_{a\in[q]} & \left[M_t^{a}\right]_{a\in[q]} \\ \left[(S_{t,t}^{a})^\top\right]_{a\in[q]} & C_{t,t}^{\vartheta,\vartheta} & C_t^{\vartheta,\star} \\ \left[(M_t^{a})^\top\right]_{a\in[q]} & (C_t^{\vartheta,\star})^\top & C^{\star,\star} \end{pmatrix} \succeq 0. \tag{191}$$

We denote by $\mathfrak{X}_{+,T}^{(q)} \subseteq \mathfrak{X}_T^{(q)}$ the closed subset of physical tuples.

*Remark* F.5. For $\mathfrak{X} \in \mathfrak{X}_{+,T}^{(q)}$ and each fixed $t \in [0, T]$, the positivity condition (191) guarantees the existence of a jointly Gaussian vector

$$\left(z_t^1, \ldots, z_t^q, \xi_t, w_\star\right) \in (\mathbb{R}^k)^q \times \mathbb{R}^k \times \mathbb{R}^{k_\star}$$

with covariance matrix $\Sigma_t(\mathfrak{X})$, where the marginal law of $(\xi_t, w_\star)$ agrees with the imported iterate effective process. This Gaussian lift is unique in law.

**Definition F.6** (Concrete coefficient family)**.** Fix $T > 0$ and $q \geq 1$. For $\mathfrak{X} = (\mathbf{Q}, \mathbf{S}, \mathbf{M}, \mathbf{Z}) \in \mathfrak{X}_{+,T}^{(q)}$, define the following deterministic coefficient objects.

1. The tangent drift matrix

$$A_t := \Gamma + \phi\, \mathbb{E}[D_1 f(\xi_t, w_\star, \varepsilon)] \in \mathbb{R}^{k\times k}, \qquad t \in [0, T]. \tag{192}$$

2. For each $a \in [q]$, define the mixed drift operator $B_t^a(\mathfrak{X}) \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^k)$ by

$$B_t^a(\mathfrak{X})[h] := \phi\, \mathbb{E}\left[D_1^2 f(\xi_t, w_\star, \varepsilon)[h, z_t^a]\right], \qquad h \in \mathbb{R}^k. \tag{193}$$

3. For each $a \in [q]$, define the iterate–tangent noise kernel

$$J_t^a(\mathfrak{X}) := \phi\, \mathbb{E}\left[\left(D_1 f(\xi_t, w_\star, \varepsilon)[z_t^a]\right) \otimes f(\xi_t, w_\star, \varepsilon)\right] \in \mathbb{R}^{k\times k}. \tag{194}$$

4. For each $a, b \in [q]$, define the tangent–tangent noise kernel

$$H_t^{ab}(\mathfrak{X}) := \phi\, \mathbb{E}\left[\left(D_1 f(\xi_t, w_\star, \varepsilon)[z_t^a]\right) \otimes \left(D_1 f(\xi_t, w_\star, \varepsilon)[z_t^b]\right)\right] \in \mathbb{R}^{k\times k}. \tag{195}$$

5. Define the lifting operator

$$\mathscr{G}_T(\mathbf{Q}, \mathbf{S}, \mathbf{M}) := \mathbf{Q}. \tag{196}$$

**Lemma F.7** (Gaussian interpolation estimate)**.** *Let $m, \ell \geq 1$, and let $\Phi : \mathbb{R}^m \to \mathbb{R}^\ell$ be $C^2$ with*

$$\sup_{x \in \mathbb{R}^m} \|D^2\Phi(x)\|_{\mathrm{op}} \leq L_\Phi < \infty.$$

*For centered Gaussian vectors $G_\Sigma \sim \mathcal{N}(0, \Sigma)$ and $G_{\widetilde{\Sigma}} \sim \mathcal{N}(0, \widetilde{\Sigma})$ in $\mathbb{R}^m$,*

$$\big\|\mathbb{E}[\Phi(G_\Sigma)] - \mathbb{E}[\Phi(G_{\widetilde{\Sigma}})]\big\|_2 \leq \frac{L_\Phi}{2}\|\Sigma - \widetilde{\Sigma}\|_{\mathrm{F}}. \tag{197}$$

*Proof.* Set $\Sigma_\tau := (1-\tau)\Sigma + \tau\widetilde{\Sigma}$, $\tau \in [0,1]$, and let $G_\tau \sim \mathcal{N}(0, \Sigma_\tau)$. By Gaussian integration by parts,

$$\frac{d}{d\tau}\mathbb{E}[\Phi(G_\tau)] = \frac{1}{2}\sum_{r,s=1}^{m}(\widetilde{\Sigma} - \Sigma)_{rs}\,\mathbb{E}[\partial^2_{rs}\Phi(G_\tau)].$$

Hence

$$\left\|\frac{d}{d\tau}\mathbb{E}[\Phi(G_\tau)]\right\|_2 \leq \frac{1}{2}\sum_{r,s=1}^{m}|(\widetilde{\Sigma} - \Sigma)_{rs}|\sup_x \|\partial^2_{rs}\Phi(x)\|_2 \leq \frac{L_\Phi}{2}\|\widetilde{\Sigma} - \Sigma\|_{\mathrm{F}}.$$

Integrating in $\tau \in [0,1]$ proves (197). $\square$

**Proposition F.8** (Admissibility of the concrete coefficient family)**.** *Under Assumptions A.2, A.3, A.4, and F.1, the coefficient family*

$$\mathfrak{F}^{(q)}_T = \Big(A,\ \{B^a\}_{a\in[q]},\ \{J^a\}_{a\in[q]},\ \{H^{ab}\}_{a,b\in[q]},\ \mathscr{G}_T\Big)$$

*defined in (192)–(196) is admissible on $[0,T]$ in the sense of Definition E.5, after restricting the map $\mathcal{T}^{(q)}_T$ of Definition E.7 to the closed complete metric space*

$$\mathfrak{B}^{(q),+}_{R,\lambda,T} := \mathfrak{B}^{(q)}_{R,\lambda,T} \cap \mathfrak{X}^{(q)}_{+,T}.$$

*Proof.* We verify the items in Definition E.5.

**Drift envelope.** By (192), boundedness of $D_1 f$, and Assumption F.1,

$$\|A_t\|_{\mathrm{op}} \leq \|\Gamma\|_{\mathrm{op}} + \phi L_1 \qquad \text{for all } t \in [0,T].$$

Hence $A_T < \infty$.

**Causality.** The coefficient maps $B^a_t, J^a_t, H^{ab}_t$ depend on the input $\mathfrak{X}$ only through the time-$t$ diagonal blocks $Q^{ab}_{t,t}$, $S^a_{t,t}$, and $M^a_t$, together with the imported iterate effective law at time $t$. Hence they are nonanticipative.

**Boundedness on balls.** Fix $R > 0$, and let $\mathfrak{X} \in \mathfrak{B}^{(q),+}_{R,\lambda,T}$. Because $\mathbf{Q}, \mathbf{S}, \mathbf{M}$ are bounded by $R$ in weighted norm, their diagonal blocks satisfy

$$\|Q^{aa}_{t,t}\|_{\mathrm{F}} \leq e^{\lambda t}R, \qquad \|S^a_{t,t}\|_{\mathrm{F}} \leq e^{\lambda t}R, \qquad \|M^a_t\|_{\mathrm{F}} \leq e^{\lambda t}R.$$

Since $t \in [0,T]$, this yields a finite horizon-dependent moment envelope. Let $z^a_t$ be the Gaussian lift from Remark F.5. Then

$$\mathbb{E}\|z^a_t\|_2^2 = \mathrm{Tr}(Q^{aa}_{t,t}) \leq \sqrt{k}\|Q^{aa}_{t,t}\|_{\mathrm{F}} \leq \sqrt{k}\,e^{\lambda T}R.$$

Therefore,

$$\mathbb{E}\|z^a_t\|_2 \leq k^{1/4}e^{\lambda T/2}R^{1/2}.$$

Using the boundedness of $D_1^2 f$, $D_1 f$, and $f$, we obtain

$$\begin{aligned}
\|B^a_t(\mathfrak{X})\|_{\mathrm{op}} &\leq \phi L_2 \mathbb{E}\|z^a_t\|_2,\\
\|J^a_t(\mathfrak{X})\|_{\mathrm{F}} &\leq \phi L_1 L_0\,\mathbb{E}\|z^a_t\|_2,\\
\|H^{ab}_t(\mathfrak{X})\|_{\mathrm{F}} &\leq \phi L_1^2\Big(\mathbb{E}\|z^a_t\|_2^2\Big)^{1/2}\Big(\mathbb{E}\|z^b_t\|_2^2\Big)^{1/2}.
\end{aligned}$$

Thus the coefficients are bounded on bounded sets.

**Weighted Lipschitz continuity.** Fix $R > 0$, and let $\mathfrak{X}, \widetilde{\mathfrak{X}} \in \mathfrak{B}^{(q),+}_{R,\lambda,T}$. At each time $t$, the corresponding Gaussian lifts have covariance matrices $\Sigma_t(\mathfrak{X})$ and $\Sigma_t(\widetilde{\mathfrak{X}})$, whose difference is bounded by a constant multiple of $\|\mathfrak{X} - \widetilde{\mathfrak{X}}\|_{\lambda,T} e^{\lambda t}$.

For $B^a_t$, fix $h \in \mathbb{R}^k$ with $\|h\|_2 = 1$ and consider

$$\Phi_B(\xi, w, z) := D_1^2 f(\xi, w, \varepsilon)[h, z].$$

As a function of $(\xi, z)$, $\Phi_B$ is $C^2$ with bounded second derivatives, because $D_1^3 f$ is bounded and $\Phi_B$ is linear in $z$. Thus Lemma F.7 yields

$$e^{-\lambda t}\|B^a_t(\mathfrak{X}) - B^a_t(\widetilde{\mathfrak{X}})\|_{\mathrm{op}} \le L_{R,T}\, \|\mathfrak{X} - \widetilde{\mathfrak{X}}\|_{\lambda,T}$$

for some finite $L_{R,T}$.

The same argument applies to $J^a_t$ and $H^{ab}_t$: their integrands are $C^2$ in the Gaussian arguments with uniformly bounded second derivatives on bounded horizon balls, because $f$, $D_1 f$, $D_1^2 f$, and $D_1^3 f$ are all bounded. Therefore (167) and (168) hold.

**Symmetry of the noise kernel.** From (195),

$$H^{ab}_t(\mathfrak{X}) = \phi \mathbb{E}\Big[(D_1 f(\xi_t, w_\star, \varepsilon)[z^a_t]) \otimes (D_1 f(\xi_t, w_\star, \varepsilon)[z^b_t])\Big],$$

which immediately gives $H^{ab}_t(\mathfrak{X}) = H^{ba}_t(\mathfrak{X})^\top$.

**Lifting operator.** The map $\mathscr{G}_T(\mathbf{Q}, \mathbf{S}, \mathbf{M}) = \mathbf{Q}$ is linear, bounded, and preserves symmetry. Hence it satisfies Definition E.5 with $G_T = 1$. $\square$

The next two statements establish the smallness of the cavity backreaction.

**Definition F.9** (Continuous-time empirical kernel interpolants)**.** For $t, s \in [0, T]$, define

$$m_t := \lfloor t n^{1-\alpha} \rfloor, \qquad m_s := \lfloor s n^{1-\alpha} \rfloor.$$

The continuous-time empirical replicated kernels are

$$Q^{ab,(d)}_{t,s} := \frac{1}{d}\sum_{j=1}^{d} \bar{U}^a_{j,m_t} \otimes \bar{U}^b_{j,m_s}, \qquad a, b \in [q], \tag{198}$$

$$S^{a,(d)}_{t,s} := \frac{1}{d}\sum_{j=1}^{d} \bar{U}^a_{j,m_t} \otimes \bar{\Theta}_{j,m_s}, \qquad a \in [q], \tag{199}$$

$$M^{a,(d)}_{t} := \frac{1}{d}\sum_{j=1}^{d} \bar{U}^a_{j,m_t} \otimes \Theta_{\star,j}, \qquad a \in [q]. \tag{200}$$

Set

$$\mathfrak{X}^{(q,d)}_T := \big(\mathbf{Q}^{(d)}, \mathbf{S}^{(d)}, \mathbf{M}^{(d)}, \mathbf{Z}^{(d)}\big), \qquad \mathbf{Z}^{(d)} := \mathbf{Q}^{(d)}. \tag{201}$$

**Lemma F.10** (Small backreaction)**.** *Under Assumptions A.2, A.3, A.4, A.5, and F.1, for every finite horizon $T > 0$ and replica number $q \ge 1$, there exists a constant $C_{T,q} < \infty$ such that*

$$\sup_{m \le M_T} \max_{j \in [d]} \left\{ \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\|\delta^{(-j)}_{i,m}\|_2^2 + \sum_{a=1}^{q} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\|\nu^{a,(-j)}_{i,m}\|_2^2 \right\} \le \frac{C_{T,q}}{d}. \tag{202}$$

*Consequently,*

$$\sup_{m \le M_T} \max_{j \in [d]} \mathbb{E}\Big[\|\mathcal{E}^{\vartheta,(-j)}_{j,m}\|_2 + \sum_{a=1}^{q} \|\mathcal{E}^{u,a,(-j)}_{j,m}\|_2\Big] \longrightarrow 0 \qquad \text{as } n, d \to \infty. \tag{203}$$

*Proof.* We prove the iterate and tangent backreaction bounds simultaneously. Fix $j \in [d]$, and abbreviate

$$\Delta_m := \Delta_m^{(-j)}, \qquad V_m^a := V_m^{a,(-j)}.$$

From the definitions (68) and the full/cavity recursions, one obtains

$$\Delta_{m+1} = P_j^{\perp} \Delta_m - \eta_m \mathcal{R}_m^{\vartheta}, \tag{204}$$
$$V_{m+1}^a = P_j^{\perp} V_m^a - \eta_m \mathcal{R}_m^{u,a}, \tag{205}$$

where $\mathcal{R}_m^{\vartheta}$ and $\mathcal{R}_m^{u,a}$ are the row-removed forcing discrepancies generated by the $j$-th coordinate contribution. Their exact formulas follow by subtracting (63) from the full update and using (66) for the tangent replicas. The key point is that every term in $\mathcal{R}_m^{\vartheta}$ and $\mathcal{R}_m^{u,a}$ contains at least one factor $x_{ij}$.

Because the design is Gaussian and $x_{ij}$ is independent of $(x_i^{(-j)}, \bar{\Theta}_m^{(-j)}, \bar{U}_m^{a,(-j)})$, the primitive bounds of Lemma D.2 and the martingale envelopes of Appendix D imply the one-step estimates

$$\mathbb{E}\|\mathcal{R}_m^{\vartheta}\|_{\mathrm{F}}^2 \le \frac{C_{T,q}}{\kappa_n d} + \frac{C_{T,q}}{d} \max_{\ell \le m} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{E}\|\delta_{i,\ell}^{(-j)}\|_2^2 \right\}, \tag{206}$$
$$\mathbb{E}\|\mathcal{R}_m^{u,a}\|_{\mathrm{F}}^2 \le \frac{C_{T,q}}{\kappa_n d} + \frac{C_{T,q}}{d} \max_{\ell \le m} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{E}\|\delta_{i,\ell}^{(-j)}\|_2^2 + \sum_{b=1}^q \frac{1}{n} \sum_{i=1}^n \mathbb{E}\|\nu_{i,\ell}^{b,(-j)}\|_2^2 \right\}. \tag{207}$$

Since $\Delta_0 = 0$ and $V_0^a = 0$, iterating (204)–(205), using the stepsize scaling $\eta_m \asymp n^{\alpha}$, the horizon length $M_T \asymp n^{1-\alpha}$, and the fact that $\kappa_n \asymp n^{\alpha}$, yields

$$\begin{aligned}\sup_{m \le M_T} \mathbb{E}\|\Delta_m\|_{\mathrm{F}}^2 + \sup_{m \le M_T} \sum_{a=1}^q \mathbb{E}\|V_m^a\|_{\mathrm{F}}^2 \le \frac{C_{T,q}}{d} + C_{T,q} \sum_{r=0}^{M_T-1} \eta_r n^{-1} \Bigg[ &\sup_{\ell \le r} \frac{1}{n} \sum_{i=1}^n \mathbb{E}\|\delta_{i,\ell}^{(-j)}\|_2^2 \\ &+ \sup_{\ell \le r} \sum_{a=1}^q \frac{1}{n} \sum_{i=1}^n \mathbb{E}\|\nu_{i,\ell}^{a,(-j)}\|_2^2 \Bigg].\end{aligned} \tag{208}$$

Now Gaussianity of the design gives

$$\mathbb{E}\|\delta_{i,m}^{(-j)}\|_2^2 = \frac{1}{d}\mathbb{E}\|\Delta_m\|_{\mathrm{F}}^2, \qquad \mathbb{E}\|\nu_{i,m}^{a,(-j)}\|_2^2 = \frac{1}{d}\mathbb{E}\|V_m^a\|_{\mathrm{F}}^2.$$

Substituting these identities into (208) and using

$$\sum_{r=0}^{M_T-1} \eta_r n^{-1} \le C_T$$

shows that the quantity

$$\mathfrak{b}_m := \max_{\ell \le m} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{E}\|\delta_{i,\ell}^{(-j)}\|_2^2 + \sum_{a=1}^q \frac{1}{n} \sum_{i=1}^n \mathbb{E}\|\nu_{i,\ell}^{a,(-j)}\|_2^2 \right\}$$

satisfies

$$\mathfrak{b}_m \le \frac{C_{T,q}}{d} + C_{T,q} \sum_{r=0}^{m-1} \eta_r n^{-1}\, \mathfrak{b}_r.$$

Discrete Grönwall now yields

$$\sup_{m \le M_T} \mathfrak{b}_m \le \frac{C_{T,q}}{d},$$

which is exactly (202).

For (203), combine (202) with the deterministic remainder bounds (116)–(117) from Lemma D.2, the finite-horizon row envelopes from Proposition D.4, and Cauchy–Schwarz. Every term in $\mathcal{E}^{\vartheta,(-j)}$ and $\mathcal{E}^{u,a,(-j)}$ contains at least one factor $\delta_{i,m}^{(-j)}$ or $\nu_{i,m}^{a,(-j)}$, hence the entire expectation vanishes uniformly on $m \le M_T$ as $d \to \infty$. $\square$

The next proposition is the key bridge from the exact cavity recursion to the deterministic Volterra closure.

**Proposition F.11** (Approximate fixed-point equation for the empirical kernels)**.** *Under the assumptions of Lemma F.10, fix $T > 0$ and $q \geq 1$, and let $\lambda \geq \lambda_\star$ be chosen as in Theorem E.11 for a radius $R$ large enough to contain the empirical kernel tuple $\mathfrak{X}_T^{(q,d)}$ with probability tending to one. Then there exists a random remainder*

$$\mathfrak{R}_T^{(q,d)} = (\mathbf{R}_Q^{(d)}, \mathbf{R}_S^{(d)}, \mathbf{R}_M^{(d)}, \mathbf{R}_Z^{(d)}) \in \mathfrak{X}_T^{(q)}$$

*such that*

$$\mathfrak{X}_T^{(q,d)} = \mathcal{T}_T^{(q)}\big(\mathfrak{X}_T^{(q,d)}\big) + \mathfrak{R}_T^{(q,d)} \tag{209}$$

*and*

$$\mathbb{E}\Big[\|\mathfrak{R}_T^{(q,d)}\|_{\lambda,T}\Big] \longrightarrow 0 \qquad \text{as } n, d \to \infty. \tag{210}$$

*Proof.* We prove the statement for $\mathbf{M}^{(d)}, \mathbf{S}^{(d)}, \mathbf{Q}^{(d)}$; the $\mathbf{Z}^{(d)}$-component is then automatic because $\mathbf{Z}^{(d)} = \mathbf{Q}^{(d)}$ and $\mathscr{G}_T$ is the identity lift.

**Step 1: discrete kernel identities.** Take the exact row recursions (82) and (87), multiply them by $\Theta_{\star,j}$, $\bar{\Theta}_{j,m_s}$, or $\bar{U}^b_{j,m_s}$, and then average over $j \in [d]$. Using the continuous-time interpolants of Definition F.9, one obtains the discrete identities

$$M_t^{a,(d)} = M_0^{a,(d)} - \sum_{r<m_t} \eta_r \Big(\mathsf{A}_r^{(d)} M_r^{a,(d)} + \mathsf{B}_r^{a,(d)} C_r^{\vartheta,\star,(d)}\Big) + R_{M,t}^{(d)}, \tag{211}$$

$$S_{t,s}^{a,(d)} = S_{0,s}^{a,(d)} - \sum_{r<m_t} \eta_r \Big(\mathsf{A}_r^{(d)} S_{r,s}^{a,(d)} + \mathsf{B}_r^{a,(d)} C_{r,s}^{\vartheta,\vartheta,(d)}\Big) + \sum_{r<m_t \wedge m_s} \eta_r^2 \mathsf{J}_r^{a,(d)} + R_{S,t,s}^{(d)}, \tag{212}$$

$$\begin{aligned} Q_{t,s}^{ab,(d)} = Q_0^{ab,(d)} &- \sum_{r<m_t} \eta_r \Big(\mathsf{A}_r^{(d)} Q_{r,s}^{ab,(d)} + \mathsf{B}_r^{a,(d)} (S_{s,r}^{b,(d)})^\top\Big) \\ &- \sum_{r<m_s} \eta_r \Big(Q_{t,r}^{ab,(d)} (\mathsf{A}_r^{(d)})^\top + S_{t,r}^{a,(d)} (\mathsf{B}_r^{b,(d)})^\top\Big) + \sum_{r<m_t \wedge m_s} \eta_r^2 \mathsf{H}_r^{ab,(d)} + R_{Q,t,s}^{ab,(d)}, \end{aligned} \tag{213}$$

where:

$$\mathsf{A}_r^{(d)} := \Gamma + \frac{n}{d}\, \mathfrak{A}_{j,r}^{(-j)}\big|_{\text{averaged over } j}, \tag{214}$$

$$\mathsf{B}_r^{a,(d)} := \frac{n}{d}\, \mathfrak{B}_{j,r}^{a,(-j)}\big|_{\text{averaged over } j}, \tag{215}$$

$$\mathsf{J}_r^{a,(d)} := \frac{n}{d} \frac{1}{n} \sum_{i=1}^{n} H_{i,r}^{a,(-j)} \otimes F_{i,r}^{(-j)}\big|_{\text{averaged over } j}, \tag{216}$$

$$\mathsf{H}_r^{ab,(d)} := \frac{n}{d} \frac{1}{n} \sum_{i=1}^{n} H_{i,r}^{a,(-j)} \otimes H_{i,r}^{b,(-j)}\big|_{\text{averaged over } j}, \tag{217}$$

and the remainders $R_M^{(d)}, R_S^{(d)}, R_Q^{(d)}$ collect:

1. the centered batch martingales $\widetilde{\mathcal{F}}, \widetilde{\mathcal{A}}, \widetilde{\mathcal{H}}, \widetilde{\mathcal{B}}$;
2. the cavity backreaction terms $\mathcal{E}^{\vartheta,(-j)}$ and $\mathcal{E}^{u,a,(-j)}$;
3. the coordinate-square replacement errors $x_{ij}^2 - d^{-1}$;
4. the time-discretization errors caused by passing from $\eta_r$-sums to $\bar{\eta}(rn^{\alpha-1})\,dr$ integrals.

**Step 2: control of the four remainder classes.** We show that each class is negligible in the weighted norm.

*(i) Centered batch martingales.* By Corollary D.10, the accumulated centered batch martingales are uniformly bounded in $L^2$, and their operator-type components are $o_{L^2}(1)$ on compact horizons. After the row-average normalization $d^{-1}\sum_j$, these terms contribute $o_{L^1}(1)$ to the weighted kernel norm.

*(ii) Cavity backreaction terms.* Lemma F.10 gives

$$\sup_{m\le M_T}\max_j \mathbb{E}\Big[\|\mathcal{E}^{\vartheta,(-j)}_{j,m}\|_2+\sum_{a=1}^{q}\|\mathcal{E}^{u,a,(-j)}_{j,m}\|_2\Big]\to 0.$$

Summing these terms against the stepsizes $\eta_r$ over $r\le M_T$ and using $\sum_{r<M_T}\eta_r n^{-1}\le C_T$ shows that their contribution to $R^{(d)}_M, R^{(d)}_S, R^{(d)}_Q$ is $o_{L^1}(1)$.

*(iii) Coordinate-square replacement.* Because the cavity fields are independent of the removed coordinate $x_{ij}$, Gaussianity gives

$$\mathbb{E}[x_{ij}^2 \mid x_i^{(-j)}, \mathscr{H}_r]=\frac{1}{d}.$$

Hence the difference between the empirical quantities $\mathsf{A}^{(d)}_r, \mathsf{B}^{a,(d)}_r, \mathsf{J}^{a,(d)}_r, \mathsf{H}^{ab,(d)}_r$ and their $d^{-1}$-replaced counterparts is a centered triangular array with variance $O(d^{-1})$. Averaging over $j$ therefore yields another $o_{L^2}(1)$ contribution.

*(iv) Riemann-sum discretization.* By Assumption A.2,

$$\sup_{t\le T}\Big|\frac{\eta_{\lfloor tn^{1-\alpha}\rfloor}}{n^\alpha}-\bar\eta(t)\Big|\to 0.$$

Since all coefficient arrays remain bounded on the relevant event, the discrete sums in (211)–(213) converge uniformly to the corresponding Volterra integrals, and the discretization error is $o(1)$.

Collecting these four bounds produces a remainder $\mathfrak{R}^{(q,d)}_T$ satisfying (210).

**Step 3: identification of the limit coefficients.** It remains to identify the deterministic limit of the discrete coefficients. Because the iterate-level field process $(\xi_t, w_\star)$ is Gaussian under the present Gaussian-design hypothesis, and because the tangent projected fields have empirical second moments exactly given by $(\mathbf{Q}^{(d)}, \mathbf{S}^{(d)}, \mathbf{M}^{(d)})$, the coefficient arrays $\mathsf{A}^{(d)}_r, \mathsf{B}^{a,(d)}_r, \mathsf{J}^{a,(d)}_r, \mathsf{H}^{ab,(d)}_r$ converge, uniformly on compact horizons, to the concrete coefficients $A_t, B^a_t(\mathfrak{X}^{(q,d)}_T), J^a_t(\mathfrak{X}^{(q,d)}_T)$, and $H^{ab}_t(\mathfrak{X}^{(q,d)}_T)$ of Definition F.6. This is an immediate consequence of the imported iterate-level DMFT from Proposition A.9, the Gaussian interpolation estimate of Lemma F.7, and the fact that all nonlinearities have bounded derivatives up to order three.

Substituting these deterministic coefficient limits into (211)–(213) yields the approximate fixed-point relation (209). □

### F.2. Closure of the Overlap Dynamics

We can now close the argument.

**Theorem F.12** (Finite-horizon replicated tangent DMFT: kernel form)**.** *Assume Assumptions A.2, A.3, A.4, A.5, and F.1. Fix $T>0$ and $q\ge 1$. Then there exists a unique physical replicated kernel tuple*

$$\mathfrak{X}^{(q)}_{T,\star}=(\mathbf{Q}_\star,\mathbf{S}_\star,\mathbf{M}_\star,\mathbf{Z}_\star)\in\mathfrak{X}^{(q)}_{+,T}$$

*such that*

$$\mathcal{T}^{(q)}_T(\mathfrak{X}^{(q)}_{T,\star})=\mathfrak{X}^{(q)}_{T,\star}. \tag{218}$$

*Moreover, for every $\lambda\ge\lambda_\star$ chosen as in Theorem E.11,*

$$\mathbb{E}\Big[\|\mathfrak{X}^{(q,d)}_T-\mathfrak{X}^{(q)}_{T,\star}\|_{\lambda,T}\Big]\longrightarrow 0 \qquad \text{as } n,d\to\infty. \tag{219}$$

*In particular, each component converges uniformly on compact time intervals:*

$$\max_{a,b\in[q]} \sup_{t,s\in[0,T]} \|Q^{ab,(d)}_{t,s} - Q^{ab}_{\star,t,s}\|_{\mathrm{F}} \xrightarrow[L^1]{} 0, \tag{220}$$

$$\max_{a\in[q]} \sup_{t,s\in[0,T]} \|S^{a,(d)}_{t,s} - S^{a}_{\star,t,s}\|_{\mathrm{F}} \xrightarrow[L^1]{} 0, \tag{221}$$

$$\max_{a\in[q]} \sup_{t\in[0,T]} \|M^{a,(d)}_{t} - M^{a}_{\star,t}\|_{\mathrm{F}} \xrightarrow[L^1]{} 0. \tag{222}$$

*Proof.* By Proposition F.8, the concrete coefficient family is admissible on the closed complete metric space $\mathfrak{B}^{(q),+}_{R,\lambda,T}$. Hence Theorem E.11 applies and yields a unique fixed point $\mathfrak{X}^{(q)}_{T,\star}$ of $\mathcal{T}^{(q)}_T$ in that space.

Now take the empirical tuple $\mathfrak{X}^{(q,d)}_T$. By Proposition F.11,

$$\mathfrak{X}^{(q,d)}_T = \mathcal{T}^{(q)}_T\big(\mathfrak{X}^{(q,d)}_T\big) + \mathfrak{R}^{(q,d)}_T.$$

Subtracting the fixed-point identity (218) and applying the contraction estimate from Theorem E.11,

$$\|\mathfrak{X}^{(q,d)}_T - \mathfrak{X}^{(q)}_{T,\star}\|_{\lambda,T} \le \rho\|\mathfrak{X}^{(q,d)}_T - \mathfrak{X}^{(q)}_{T,\star}\|_{\lambda,T} + \|\mathfrak{R}^{(q,d)}_T\|_{\lambda,T},$$

where $\rho \in (0,1)$ is the contraction factor. Rearranging,

$$\|\mathfrak{X}^{(q,d)}_T - \mathfrak{X}^{(q)}_{T,\star}\|_{\lambda,T} \le \frac{1}{1-\rho}\|\mathfrak{R}^{(q,d)}_T\|_{\lambda,T}.$$

Taking expectations and using (210) proves (219). The componentwise convergence statements follow immediately from the definition of the norm $\|\cdot\|_{\lambda,T}$. □

**Corollary F.13** (Projected-field version of the replicated tangent DMFT)**.** *Under the assumptions of Theorem F.12, fix a finite time grid*

$$0 \le t_1 \le \cdots \le t_m \le T.$$

*For each $d$, define the projected tangent and iterate fields*

$$Z^{a,(d)}_{i,\ell} := x_i^\top \bar U^a_{m_{t_\ell}}, \qquad \Xi^{(d)}_{i,\ell} := x_i^\top \bar\Theta_{m_{t_\ell}}, \qquad W^{\star,(d)}_i := x_i^\top \Theta_\star,$$

*where $m_{t_\ell} = \lfloor t_\ell n^{1-\alpha}\rfloor$. Let*

$$\big(z^a_{t_\ell,\star},\, \xi_{t_\ell,\star},\, w_\star,\, \varepsilon\big)_{a\in[q],\,\ell\in[m]}$$

*be the jointly Gaussian field lift induced by the limiting kernel tuple $\mathfrak{X}^{(q)}_{T,\star}$. Then for every*

$$\varphi \in \mathrm{PL}_2\big((\mathbb{R}^k)^{mq} \times (\mathbb{R}^k)^m \times \mathbb{R}^{k_\star} \times \mathbb{R}\big),$$

$$\frac{1}{n}\sum_{i=1}^n \varphi\Big((Z^{a,(d)}_{i,\ell})_{a\in[q],\,\ell\in[m]}, (\Xi^{(d)}_{i,\ell})_{\ell\in[m]}, W^{\star,(d)}_i, \varepsilon_i\Big) \xrightarrow[L^1]{} \mathbb{E}\Big[\varphi\Big((z^a_{t_\ell,\star})_{a\in[q],\,\ell\in[m]}, (\xi_{t_\ell,\star})_{\ell\in[m]}, w_\star, \varepsilon\Big)\Big]. \tag{223}$$

*Proof.* Under the Gaussian design, conditional on the row arrays, the field vectors

$$\Big((Z^{a,(d)}_{i,\ell})_{a,\ell}, (\Xi^{(d)}_{i,\ell})_\ell, W^{\star,(d)}_i\Big)$$

are centered Gaussian with covariance determined exactly by the empirical kernel tuple $\mathfrak{X}^{(q,d)}_T$. By Theorem F.12, this covariance converges uniformly to that of the limiting Gaussian lift associated with $\mathfrak{X}^{(q)}_{T,\star}$. Since $\varphi$ is second-order pseudo-Lipschitz and the Gaussian moments are uniformly bounded on compact horizons, the convergence of the Gaussian covariances implies convergence of the corresponding expectations. Averaging over $i$ and using exchangeability of the samples yields (223). □

*Remark* F.14 (What Appendix F proves and what it does not). Appendix F proves the finite-horizon replicated tangent DMFT at the *kernel* level and, by Corollary F.13, at the level of projected field observables. This is already sufficient for the two subsequent steps of the paper: Appendix G, which studies characteristic $q$-volume growth, and Appendix H, which studies the large-time stochastic edge. A full row-level path-space propagation-of-chaos statement can be obtained by standard martingale-problem arguments once the kernel closure is known, but it is not needed for the present paper's main conclusions and is therefore omitted.

*Proof of Theorem 3.1.* Replace the placeholder statement of Theorem 3.1 in the main text by the kernel-level convergence statement of Theorem F.12. The proof is exactly the one given above. □

## G. Proof of the Characteristic Volume-Growth Law

This appendix turns the kernel-level replicated tangent DMFT from Appendix F into a deterministic law for the $q$-dimensional tangent-volume observable introduced in Definition 2.4. The key point is that the characteristic $q$-volume depends only on the scalar $q \times q$ Gram matrix formed from the replica–replica overlaps at equal time, and those overlaps are already part of the limiting kernel tuple of Theorem F.12.

The proof therefore has two steps.

1. We extract from the matrix-valued replica kernel $\mathbf{Q}^{(d)} = (Q^{ab,(d)})_{a,b\in[q]}$ the scalar Gram matrix governing the tangent volume, and prove its uniform convergence on compact horizons.
2. On any time interval on which the limiting scalar Gram matrix stays uniformly positive definite, the logarithmic determinant is a Lipschitz function of the Gram matrix. This yields the deterministic limit of the characteristic $q$-volume growth.

### G.1. Convergence of the Tangent Gram Matrix

Recall from Definition F.9 that, for $t, s \in [0, T]$,

$$Q^{ab,(d)}_{t,s} = \frac{1}{d}\sum_{j=1}^{d} \bar{U}^a_{j,m_t} \otimes \bar{U}^b_{j,m_s} \in \mathbb{R}^{k\times k}, \qquad m_t := \lfloor tn^{1-\alpha} \rfloor,$$

and that Theorem F.12 yields a deterministic limit $Q^{ab}_{\star,t,s}$ for every $a, b \in [q]$.

**Definition G.1** (Scalar tangent Gram matrices)**.** For $t \in [0, T]$, define the empirical scalar tangent Gram matrix

$$\mathbf{G}^{(d,q)}_t := \Big[\mathrm{Tr}\big(Q^{ab,(d)}_{t,t}\big)\Big]_{a,b=1}^{q} = \Big[\frac{1}{d}\,\langle \bar{U}^a_{m_t}, \bar{U}^b_{m_t}\rangle_{\mathrm{F}}\Big]_{a,b=1}^{q} \in \mathbb{R}^{q\times q}, \tag{224}$$

and the limiting scalar tangent Gram matrix

$$\mathbf{G}^{(q)}_{\star,t} := \Big[\mathrm{Tr}\big(Q^{ab}_{\star,t,t}\big)\Big]_{a,b=1}^{q} \in \mathbb{R}^{q\times q}. \tag{225}$$

*Remark* G.2 (Compatibility with Definition 2.4). Definition 2.4 used the ambient-dimension normalization

$$\mathbf{Q}^{(d,q)}_m = \Big[\frac{1}{p_d}\,\langle \bar{U}^a_m, \bar{U}^b_m\rangle_{\mathrm{F}}\Big]_{a,b=1}^{q}, \qquad p_d = dk.$$

By Corollary B.5 in Appendix B,

$$\mathbf{G}^{(d,q)}_t = k\,\mathbf{Q}^{(d,q)}_{m_t},$$

and hence the determinant ratios that define finite-time $q$-volume growth are identical in the two normalizations.

**Lemma G.3** (Positivity structure of the limiting Gram matrix)**.** *For every $t \in [0, T]$, the matrix $\mathbf{G}^{(q)}_{\star,t}$ is symmetric positive semidefinite. Moreover,*

$$\mathbf{G}^{(q)}_{\star,0} = \Big[\mathrm{Tr}(Q^{ab}_0)\Big]_{a,b=1}^{q} = \mathbf{Q}^{(q)}_0, \tag{226}$$

*where $\mathbf{Q}^{(q)}_0$ is the strictly positive definite initial Gram matrix from Assumption A.5.*

*Proof.* By Theorem F.12, the limiting kernel tuple $\mathfrak{X}_{T,\star}^{(q)}$ is physical in the sense of Definition F.4. Therefore, for each fixed $t \in [0, T]$, there exists a jointly Gaussian lift

$$(z_t^1, \ldots, z_t^q, \xi_t, w_\star) \in (\mathbb{R}^k)^q \times \mathbb{R}^k \times \mathbb{R}^{k_\star}$$

such that

$$Q_{\star,t,t}^{ab} = \mathbb{E}[z_t^a \otimes z_t^b] \qquad \text{for all } a, b \in [q].$$

Now let $c = (c_1, \ldots, c_q)^\top \in \mathbb{R}^q$. Then

$$\begin{aligned} c^\top \mathbf{G}_{\star,t}^{(q)} c &= \sum_{a,b=1}^{q} c_a c_b \ \mathrm{Tr}(Q_{\star,t,t}^{ab}) = \sum_{a,b=1}^{q} c_a c_b \ \mathrm{Tr}\Big(\mathbb{E}[z_t^a \otimes z_t^b]\Big) \\ &= \mathbb{E}\left[\sum_{a,b=1}^{q} c_a c_b \ \left\langle z_t^a,\ z_t^b \right\rangle\right] = \mathbb{E}\left[\left\|\sum_{a=1}^{q} c_a z_t^a\right\|_2^2\right] \geq 0. \end{aligned}$$

Hence $\mathbf{G}_{\star,t}^{(q)}$ is symmetric positive semidefinite.

At $t = 0$, the boundary data in Definition E.3 yield

$$Q_0^{ab} = \mathbb{E}[u_0^a \otimes u_0^b].$$

Therefore

$$\mathrm{Tr}(Q_0^{ab}) = \mathbb{E}[\langle u_0^a,\ u_0^b \rangle] = \int \left\langle u^a,\ u^b \right\rangle \ \nu_{0,\star,q}(du^1, \ldots, du^q, d\vartheta_0, d\vartheta_\star),$$

which is exactly the $(a, b)$-entry of $\mathbf{Q}_0^{(q)}$. This proves (226). The strict positive definiteness follows from Assumption A.5. □

**Proposition G.4** (Uniform convergence of the scalar tangent Gram matrix)**.** *Under the assumptions of Theorem F.12, for every finite horizon $T > 0$,*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left\|\mathbf{G}_t^{(d,q)} - \mathbf{G}_{\star,t}^{(q)}\right\|_{\mathrm{F}}\right] \longrightarrow 0 \qquad \textit{as } n, d \to \infty. \tag{227}$$

*In particular,*

$$\sup_{t\in[0,T]} \left\|\mathbf{G}_t^{(d,q)} - \mathbf{G}_{\star,t}^{(q)}\right\|_{\mathrm{F}} \xrightarrow[L^1]{} 0 \qquad \textit{and hence} \qquad \xrightarrow[\mathbb{P}]{} 0. \tag{228}$$

*Proof.* By Definition G.1,

$$\big(\mathbf{G}_t^{(d,q)} - \mathbf{G}_{\star,t}^{(q)}\big)_{ab} = \mathrm{Tr}\big(Q_{t,t}^{ab,(d)} - Q_{\star,t,t}^{ab}\big).$$

For every $k \times k$ matrix $A$,

$$|\,\mathrm{Tr}(A)| \leq \sqrt{k}\, \|A\|_{\mathrm{F}}.$$

Therefore,

$$\sup_{t\in[0,T]} \left|\big(\mathbf{G}_t^{(d,q)} - \mathbf{G}_{\star,t}^{(q)}\big)_{ab}\right| \leq \sqrt{k} \sup_{t\in[0,T]} \|Q_{t,t}^{ab,(d)} - Q_{\star,t,t}^{ab}\|_{\mathrm{F}}.$$

Taking the Frobenius norm over the $q \times q$ block index,

$$\sup_{t\in[0,T]} \|\mathbf{G}_t^{(d,q)} - \mathbf{G}_{\star,t}^{(q)}\|_{\mathrm{F}} \leq q\sqrt{k} \max_{a,b\in[q]} \sup_{t\in[0,T]} \|Q_{t,t}^{ab,(d)} - Q_{\star,t,t}^{ab}\|_{\mathrm{F}}.$$

Now Theorem F.12 gives

$$\max_{a,b\in[q]} \sup_{t,s\in[0,T]} \|Q_{t,s}^{ab,(d)} - Q_{\star,t,s}^{ab}\|_{\mathrm{F}} \xrightarrow[L^1]{} 0.$$

Restricting to the diagonal $s = t$ proves (227). □

### G.2. Deterministic Limit of $\Lambda^{(d)}_{q,M}$

To pass from Gram-matrix convergence to a volume-growth law, we need uniform positive definiteness on the time interval of interest.

**Definition G.5** ($q$-volume-admissible interval)**.** Let $0 < \tau \le T < \infty$. The interval $[\tau, T]$ is called *$q$-volume admissible* for the limiting replicated tangent theory if

$$\gamma^{(q)}_{\tau,T} := \frac{1}{2}\min\left\{\lambda_{\min}\big(\mathbf{G}^{(q)}_{\star,0}\big), \inf_{t\in[\tau,T]} \lambda_{\min}\big(\mathbf{G}^{(q)}_{\star,t}\big)\right\} > 0. \tag{229}$$

*Remark* G.6. Because $\mathbf{G}^{(q)}_{\star,0} = \mathbf{Q}^{(q)}_0 \succ 0$ by Lemma G.3, the only nontrivial part of Definition G.5 is the strict positivity of the limiting tangent Gram matrix on $[\tau, T]$. Appendix H will work precisely with intervals on which this positivity persists.

**Lemma G.7** (Lipschitz continuity of $\log\det$ on a uniformly positive cone)**.** *Let $q \ge 1$ and $\gamma > 0$. If $A, B \in \mathbb{R}^{q\times q}$ are symmetric and satisfy*

$$A \succeq \gamma I_q, \qquad B \succeq \gamma I_q,$$

*then*

$$\big|\log\det A - \log\det B\big| \le \frac{\sqrt{q}}{\gamma}\,\|A - B\|_{\mathrm{F}}. \tag{230}$$

*Proof.* Consider the line segment

$$A_\theta := (1-\theta)A + \theta B, \qquad \theta \in [0, 1].$$

Since the cone of symmetric positive semidefinite matrices is convex, $A_\theta \succeq \gamma I_q$ for every $\theta \in [0, 1]$. Hence

$$\|A_\theta^{-1}\|_{\mathrm{op}} \le \gamma^{-1}.$$

By the fundamental theorem of calculus and the Fréchet derivative $D(\log\det)(M)[H] = \mathrm{Tr}(M^{-1}H)$,

$$\log\det B - \log\det A = \int_0^1 \mathrm{Tr}\big(A_\theta^{-1}(B - A)\big)\, d\theta.$$

Therefore,

$$\begin{aligned}\big|\log\det B - \log\det A\big| &\le \int_0^1 \|A_\theta^{-1}\|_{\mathrm{op}}\, \|B - A\|_*\, d\theta\\ &\le \frac{1}{\gamma}\|B - A\|_* \le \frac{\sqrt{q}}{\gamma}\|B - A\|_{\mathrm{F}},\end{aligned}$$

because the nuclear norm satisfies $\|M\|_* \le \sqrt{q}\,\|M\|_{\mathrm{F}}$ for $q \times q$ matrices. □

**Definition G.8** (Epoch-normalized characteristic $q$-volume growth)**.** Let $0 < t \le T$. Whenever $\mathbf{G}^{(d,q)}_t$ and $\mathbf{G}^{(d,q)}_0$ are positive definite, define

$$\Lambda^{(d)}_{q,t} := \frac{1}{2t}\log\det\Big(\mathbf{G}^{(d,q)}_t\big(\mathbf{G}^{(d,q)}_0\big)^{-1}\Big). \tag{231}$$

Likewise, whenever $\mathbf{G}^{(q)}_{\star,t}$ and $\mathbf{G}^{(q)}_{\star,0}$ are positive definite, define

$$\Lambda_{q,t} := \frac{1}{2t}\log\det\Big(\mathbf{G}^{(q)}_{\star,t}\big(\mathbf{G}^{(q)}_{\star,0}\big)^{-1}\Big). \tag{232}$$

**Theorem G.9** (Deterministic law for characteristic $q$-volume growth)**.** *Assume the hypotheses of Theorem F.12, and let $[\tau, T] \subset (0, \infty)$ be a $q$-volume-admissible interval in the sense of Definition G.5. Then the following hold.*

1. *There exists an event $\mathcal{A}^{(d)}_{\tau,T}$ such that*

$$\mathbb{P}\big(\mathcal{A}^{(d)}_{\tau,T}\big) \to 1 \qquad \textit{as } n, d \to \infty, \tag{233}$$

*and on $\mathcal{A}^{(d)}_{\tau,T}$,*

$$\lambda_{\min}\big(\mathbf{G}_0^{(d,q)}\big) \geq \gamma^{(q)}_{\tau,T}, \qquad \inf_{t\in[\tau,T]} \lambda_{\min}\big(\mathbf{G}_t^{(d,q)}\big) \geq \gamma^{(q)}_{\tau,T}. \tag{234}$$

*In particular, $\Lambda^{(d)}_{q,t}$ is well defined for every $t \in [\tau, T]$ on $\mathcal{A}^{(d)}_{\tau,T}$.*

2. *On $\mathcal{A}^{(d)}_{\tau,T}$,*

$$\sup_{t\in[\tau,T]} \big|\Lambda^{(d)}_{q,t} - \Lambda_{q,t}\big| \leq \frac{\sqrt{q}}{2\tau\,\gamma^{(q)}_{\tau,T}} \left( \sup_{t\in[\tau,T]} \|\mathbf{G}_t^{(d,q)} - \mathbf{G}^{(q)}_{\star,t}\|_{\mathrm{F}} + \|\mathbf{G}_0^{(d,q)} - \mathbf{G}^{(q)}_{\star,0}\|_{\mathrm{F}} \right). \tag{235}$$

3. *Consequently,*

$$\mathbb{E}\left[ \mathbf{1}_{\mathcal{A}^{(d)}_{\tau,T}} \sup_{t\in[\tau,T]} \big|\Lambda^{(d)}_{q,t} - \Lambda_{q,t}\big| \right] \longrightarrow 0 \qquad \textit{as } n, d \to \infty, \tag{236}$$

*and hence*

$$\sup_{t\in[\tau,T]} \big|\Lambda^{(d)}_{q,t} - \Lambda_{q,t}\big| \underset{\mathbb{P}}{\to} 0 \qquad \textit{on the positive-definite event.} \tag{237}$$

*Proof.* We prove the three claims in order.

**Step 1: positivity of the empirical Gram matrix.** By Proposition G.4,

$$\sup_{t\in[0,T]} \|\mathbf{G}_t^{(d,q)} - \mathbf{G}^{(q)}_{\star,t}\|_{\mathrm{F}} \underset{\mathbb{P}}{\to} 0.$$

Define

$$\mathcal{A}^{(d)}_{\tau,T} := \left\{ \|\mathbf{G}_0^{(d,q)} - \mathbf{G}^{(q)}_{\star,0}\|_{\mathrm{op}} \leq \gamma^{(q)}_{\tau,T} \right\} \cap \left\{ \sup_{t\in[\tau,T]} \|\mathbf{G}_t^{(d,q)} - \mathbf{G}^{(q)}_{\star,t}\|_{\mathrm{op}} \leq \gamma^{(q)}_{\tau,T} \right\}. \tag{238}$$

Since $\|\cdot\|_{\mathrm{op}} \leq \|\cdot\|_{\mathrm{F}}$, Proposition G.4 implies $\mathbb{P}(\mathcal{A}^{(d)}_{\tau,T}) \to 1$.

Now let $\omega \in \mathcal{A}^{(d)}_{\tau,T}$. By Weyl's inequality,

$$\lambda_{\min}\big(\mathbf{G}_0^{(d,q)}\big) \geq \lambda_{\min}\big(\mathbf{G}^{(q)}_{\star,0}\big) - \|\mathbf{G}_0^{(d,q)} - \mathbf{G}^{(q)}_{\star,0}\|_{\mathrm{op}} \geq 2\gamma^{(q)}_{\tau,T} - \gamma^{(q)}_{\tau,T} = \gamma^{(q)}_{\tau,T}.$$

Likewise, for every $t \in [\tau, T]$,

$$\lambda_{\min}\big(\mathbf{G}_t^{(d,q)}\big) \geq \lambda_{\min}\big(\mathbf{G}^{(q)}_{\star,t}\big) - \|\mathbf{G}_t^{(d,q)} - \mathbf{G}^{(q)}_{\star,t}\|_{\mathrm{op}} \geq 2\gamma^{(q)}_{\tau,T} - \gamma^{(q)}_{\tau,T} = \gamma^{(q)}_{\tau,T}.$$

This proves (234).

**Step 2: uniform $\log\det$ control.** Fix $\omega \in \mathcal{A}^{(d)}_{\tau,T}$. For every $t \in [\tau, T]$,

$$\begin{aligned}
\big|\Lambda^{(d)}_{q,t} - \Lambda_{q,t}\big| &= \frac{1}{2t}\Big|\log\det \mathbf{G}_t^{(d,q)} - \log\det \mathbf{G}^{(q)}_{\star,t} - \log\det \mathbf{G}_0^{(d,q)} + \log\det \mathbf{G}^{(q)}_{\star,0}\Big| \\
&\leq \frac{1}{2\tau}\Big( \big|\log\det \mathbf{G}_t^{(d,q)} - \log\det \mathbf{G}^{(q)}_{\star,t}\big| + \big|\log\det \mathbf{G}_0^{(d,q)} - \log\det \mathbf{G}^{(q)}_{\star,0}\big| \Big).
\end{aligned}$$

Since all four matrices in the last display are bounded below by $\gamma^{(q)}_{\tau,T} I_q$ on $\mathcal{A}^{(d)}_{\tau,T}$, we may apply Lemma G.7 twice to obtain

$$\big|\Lambda^{(d)}_{q,t} - \Lambda_{q,t}\big| \leq \frac{\sqrt{q}}{2\tau\,\gamma^{(q)}_{\tau,T}} \Big( \|\mathbf{G}_t^{(d,q)} - \mathbf{G}^{(q)}_{\star,t}\|_{\mathrm{F}} + \|\mathbf{G}_0^{(d,q)} - \mathbf{G}^{(q)}_{\star,0}\|_{\mathrm{F}} \Big).$$

Taking the supremum over $t \in [\tau, T]$ proves (235).

**Step 3: convergence in $L^1$ and in probability.** Take expectations in (235) and use Proposition G.4. This gives (236). The convergence in probability (237) follows immediately by Markov's inequality. □

**Corollary G.10** (Compatibility with the discrete-time definition)**.** *Fix $m \in \mathbb{N}$ and set*

$$t_m := mn^{\alpha-1}.$$

*Assume that the Gram matrices at times 0 and $t_m$ are positive definite. Then the epoch-normalized characteristic $q$-volume growth of Definition G.8 satisfies the exact identity*

$$\Lambda^{(d)}_{q,t_m} = n^{1-\alpha} \cdot \frac{1}{2m} \log\det\Big(\mathbf{Q}^{(d,q)}_m \big(\mathbf{Q}^{(d,q)}_0\big)^{-1}\Big), \tag{239}$$

*where $\mathbf{Q}^{(d,q)}_m$ is the step-$m$ Gram matrix from Definition 2.4.*

*Proof.* By Definition G.1 and the definition of $t_m$,

$$\mathbf{G}^{(d,q)}_{t_m} = \Big[\frac{1}{d}\,\langle \bar{U}^a_m, \bar{U}^b_m\rangle_{\mathrm{F}}\Big]^q_{a,b=1}.$$

By Remark G.2 and Corollary B.5,

$$\det\Big(\mathbf{G}^{(d,q)}_{t_m}(\mathbf{G}^{(d,q)}_0)^{-1}\Big) = \det\Big(\mathbf{Q}^{(d,q)}_m(\mathbf{Q}^{(d,q)}_0)^{-1}\Big).$$

Therefore,

$$\Lambda^{(d)}_{q,t_m} = \frac{1}{2t_m}\log\det\Big(\mathbf{G}^{(d,q)}_{t_m}(\mathbf{G}^{(d,q)}_0)^{-1}\Big) = \frac{n^{1-\alpha}}{2m}\log\det\Big(\mathbf{Q}^{(d,q)}_m(\mathbf{Q}^{(d,q)}_0)^{-1}\Big),$$

which is exactly (239). □

**Corollary G.11** (Exterior-power and QR representations of the limiting law)**.** *Fix $m \in \mathbb{N}$ and let $t_m = mn^{\alpha-1}$. On the positive-definite event of Theorem G.9, the empirical characteristic $q$-volume growth admits the two equivalent representations*

$$\Lambda^{(d)}_{q,t_m} = \frac{1}{t_m}\log\frac{\big\|(\bigwedge^q \Phi_m)\,\omega_q(\mathbf{U}_0)\big\|}{\|\omega_q(\mathbf{U}_0)\|}, \tag{240}$$

$$\Lambda^{(d)}_{q,t_m} = \frac{1}{t_m}\sum_{r=0}^{m-1}\log\det R^{(d)}_r, \tag{241}$$

*where:*

1. *$\Phi_m$ is the stochastic stability cocycle from Definition 2.2,*
2. *$\mathbf{U}_0 = (U^1_0, \dots, U^q_0)$ is the initial tangent $q$-tuple,*
3. *$\omega_q(\mathbf{U}_0)$ is the exterior-power representative from Definition B.7,*
4. *$R^{(d)}_r$ are the local QR factors from Proposition B.12.*

*Consequently, Theorem G.9 gives a deterministic limit for both the exterior-power growth and the accumulated QR growth of the tangent cocycle.*

*Proof.* By Corollary G.10, the left-hand side is the epoch-rescaled version of the step-normalized determinant growth from Definition 2.4. The two representations then follow immediately from Corollary B.9 and Proposition B.12 in Appendix B, using again the normalization invariance of determinant ratios from Corollary B.5. □

*Proof of Theorem 3.3.* Replace the placeholder statement of Theorem 3.3 in the main text by the content of Theorem G.9. The proof is exactly the one given above. □

# H. Proof of the Stationary Stochastic-Edge Theorem

This appendix upgrades the finite-horizon characteristic $q$-volume law of Appendix G to a long-time stationary theory. The key idea is simple: once the limiting tangent-volume Gram matrix becomes asymptotically positive definite and sufficiently regular in time, its logarithmic determinant has an instantaneous growth rate, and the finite-time characteristic $q$-volume growth is the Cesàro average of that instantaneous rate. The stochastic edge is then the zero-crossing of the top ($q = 1$) stationary exponent.

There are two distinct tasks in this appendix.

1. We first formulate a stationary-regime hypothesis directly at the level of the limiting Gram matrix and prove existence of the long-time exponent.
2. We then consider a family of such stationary regimes indexed by a control parameter (for example, a learning-rate scale or a batch-size scale), prove continuity of the top exponent under a uniform tail hypothesis, and deduce existence and uniqueness of the stochastic edge under a monotone crossing assumption.

## H.1. Stationary Effective Regime

Throughout this subsection, fix a replica number $q \geq 1$, and let

$$t \longmapsto \mathbf{G}^{(q)}_{\star,t} \in \mathbb{R}^{q\times q}$$

denote the deterministic limiting scalar tangent Gram matrix from Definition G.1. Recall that $\mathbf{G}^{(q)}_{\star,0} = \mathbf{Q}^{(q)}_0 \succ 0$ by Lemma G.3.

**Definition H.1** (Asymptotically stationary $q$-volume regime)**.** We say that the limiting replicated tangent theory is in an *asymptotically stationary $q$-volume regime* if there exists $t^{(q)}_{\mathrm{st}} \geq 0$ such that the following hold.

1. The path $t \mapsto \mathbf{G}^{(q)}_{\star,t}$ is $C^1$ and takes values in $\mathrm{Sym}_{++}(q)$ on $[t^{(q)}_{\mathrm{st}}, \infty)$.
2. The *instantaneous characteristic $q$-volume rate*

$$\mathfrak{g}_q(t) := \frac{1}{2}\,\mathrm{Tr}\Big(\big(\mathbf{G}^{(q)}_{\star,t}\big)^{-1}\dot{\mathbf{G}}^{(q)}_{\star,t}\Big), \qquad t \geq t^{(q)}_{\mathrm{st}}, \tag{242}$$

   admits a finite limit

$$\lambda_q := \lim_{t\to\infty} \mathfrak{g}_q(t) \in \mathbb{R}. \tag{243}$$

*Remark* H.2. By Jacobi's formula, $\mathfrak{g}_q(t)$ is exactly the instantaneous growth rate of the $q$-dimensional tangent volume. When $q = 1$,

$$\mathbf{G}^{(1)}_{\star,t} = \big[g^{(1)}_{\star,t}\big]$$

is scalar, and (242) reduces to

$$\mathfrak{g}_1(t) = \frac{1}{2}\frac{d}{dt}\log g^{(1)}_{\star,t}.$$

Thus $\lambda_1$ is the asymptotic log-growth rate of the limiting tangent norm itself.

**Lemma H.3** (Log-determinant representation of the finite-time exponent)**.** *Assume Definition H.1. Then, for every $t \geq t^{(q)}_{\mathrm{st}}$,*

$$\Lambda_{q,t} = \frac{1}{2t}\log\det\Big(\mathbf{G}^{(q)}_{\star,t^{(q)}_{\mathrm{st}}}\big(\mathbf{G}^{(q)}_{\star,0}\big)^{-1}\Big) + \frac{1}{t}\int_{t^{(q)}_{\mathrm{st}}}^{t} \mathfrak{g}_q(r)\,dr, \tag{244}$$

*where $\Lambda_{q,t}$ is the limiting characteristic $q$-volume growth from Definition G.8.*

*Proof.* For $t \geq t^{(q)}_{\mathrm{st}}$, the map

$$t \mapsto \log\det \mathbf{G}^{(q)}_{\star,t}$$

is $C^1$ because $\mathbf{G}^{(q)}_{\star,t} \in \mathrm{Sym}_{++}(q)$ and $t \mapsto \mathbf{G}^{(q)}_{\star,t}$ is $C^1$. Jacobi's formula yields

$$\frac{d}{dt} \log\det \mathbf{G}^{(q)}_{\star,t} = \mathrm{Tr}\Big((\mathbf{G}^{(q)}_{\star,t})^{-1}\dot{\mathbf{G}}^{(q)}_{\star,t}\Big) = 2\mathfrak{g}_q(t).$$

Integrating from $t^{(q)}_{\mathrm{st}}$ to $t$ gives

$$\log\det \mathbf{G}^{(q)}_{\star,t} = \log\det \mathbf{G}^{(q)}_{\star,t^{(q)}_{\mathrm{st}}} + 2\int_{t^{(q)}_{\mathrm{st}}}^{t} \mathfrak{g}_q(r)\, dr.$$

Subtract $\log\det \mathbf{G}^{(q)}_{\star,0}$, divide by $2t$, and use Definition G.8. □

**Theorem H.4** (Existence of the stationary characteristic $q$-volume exponent)**.** *Assume Definition H.1. Then the limiting characteristic $q$-volume growth satisfies*

$$\Lambda_{q,t} \longrightarrow \lambda_q \qquad \text{as } t \to \infty, \tag{245}$$

*where $\lambda_q$ is the limit from (243). More quantitatively, for every $t \geq t^{(q)}_{\mathrm{st}}$,*

$$\big|\Lambda_{q,t} - \lambda_q\big| \leq \frac{1}{2t}\left|\log\det\Big(\mathbf{G}^{(q)}_{\star,t^{(q)}_{\mathrm{st}}}\big(\mathbf{G}^{(q)}_{\star,0}\big)^{-1}\Big)\right| + \frac{1}{t}\int_{t^{(q)}_{\mathrm{st}}}^{t} \big|\mathfrak{g}_q(r) - \lambda_q\big|\, dr. \tag{246}$$

*Proof.* Subtract $\lambda_q$ from both sides of (244):

$$\Lambda_{q,t} - \lambda_q = \frac{1}{2t}\log\det\Big(\mathbf{G}^{(q)}_{\star,t^{(q)}_{\mathrm{st}}}\big(\mathbf{G}^{(q)}_{\star,0}\big)^{-1}\Big) + \frac{1}{t}\int_{t^{(q)}_{\mathrm{st}}}^{t} \big(\mathfrak{g}_q(r) - \lambda_q\big)\, dr.$$

Taking absolute values yields (246). The first term on the right-hand side tends to zero because the numerator is constant and $t \to \infty$. The second term is the Cesàro average of a function that converges to zero, hence also tends to zero. Therefore (245) holds. □

**Corollary H.5** (Uniform tail control of the limiting finite-time exponent)**.** *Assume Definition H.1. Then for every $\varepsilon > 0$ there exists $T_\varepsilon \geq t^{(q)}_{\mathrm{st}}$ such that*

$$\sup_{t \geq T_\varepsilon} \big|\Lambda_{q,t} - \lambda_q\big| \leq \varepsilon. \tag{247}$$

*Proof.* Fix $\varepsilon > 0$. Since $\mathfrak{g}_q(t) \to \lambda_q$, there exists $T_1 \geq t^{(q)}_{\mathrm{st}}$ such that

$$\sup_{r \geq T_1} |\mathfrak{g}_q(r) - \lambda_q| \leq \frac{\varepsilon}{3}.$$

Then for $t \geq T_1$,

$$\begin{aligned}\frac{1}{t}\int_{t^{(q)}_{\mathrm{st}}}^{t} |\mathfrak{g}_q(r) - \lambda_q|\, dr &\leq \frac{1}{t}\int_{t^{(q)}_{\mathrm{st}}}^{T_1} |\mathfrak{g}_q(r) - \lambda_q|\, dr + \frac{1}{t}\int_{T_1}^{t} |\mathfrak{g}_q(r) - \lambda_q|\, dr \\ &\leq \frac{C_\varepsilon}{t} + \frac{\varepsilon}{3},\end{aligned}$$

where

$$C_\varepsilon := \int_{t^{(q)}_{\mathrm{st}}}^{T_1} |\mathfrak{g}_q(r) - \lambda_q|\, dr < \infty.$$

Choose $T_2 \geq T_1$ such that

$$\frac{C_\varepsilon}{T_2} \leq \frac{\varepsilon}{3}, \qquad \frac{1}{2T_2}\left|\log\det\Big(\mathbf{G}^{(q)}_{\star,t^{(q)}_{\mathrm{st}}}(\mathbf{G}^{(q)}_{\star,0})^{-1}\Big)\right| \leq \frac{\varepsilon}{3}.$$

Then (246) implies (247) with $T_\varepsilon := T_2$. □

**Corollary H.6** (Two-scale empirical approximation on long windows)**.** *Assume Definition H.1. Suppose, in addition, that for every $T \geq t_{\mathrm{st}}^{(q)}$, the interval $[T, 2T]$ is q-volume-admissible in the sense of Definition G.5. Then for every $\varepsilon > 0$ there exists $T_\varepsilon \geq t_{\mathrm{st}}^{(q)}$ such that*

$$\limsup_{n,d\to\infty} \mathbb{P}\left( \sup_{t\in[T_\varepsilon, 2T_\varepsilon]} |\Lambda_{q,t}^{(d)} - \lambda_q| > \varepsilon \right) = 0. \tag{248}$$

*Proof.* Fix $\varepsilon > 0$. By Corollary H.5, choose $T_\varepsilon$ large enough that

$$\sup_{t\geq T_\varepsilon} |\Lambda_{q,t} - \lambda_q| \leq \frac{\varepsilon}{2}.$$

Then, on the window $[T_\varepsilon, 2T_\varepsilon]$,

$$|\Lambda_{q,t}^{(d)} - \lambda_q| \leq |\Lambda_{q,t}^{(d)} - \Lambda_{q,t}| + \frac{\varepsilon}{2}.$$

Hence

$$\mathbb{P}\left( \sup_{t\in[T_\varepsilon, 2T_\varepsilon]} |\Lambda_{q,t}^{(d)} - \lambda_q| > \varepsilon \right) \leq \mathbb{P}\left( \sup_{t\in[T_\varepsilon, 2T_\varepsilon]} |\Lambda_{q,t}^{(d)} - \Lambda_{q,t}| > \frac{\varepsilon}{2} \right).$$

The right-hand side tends to zero by Theorem G.9, because the window $[T_\varepsilon, 2T_\varepsilon]$ is $q$-volume-admissible by assumption. □

### H.2. Definition and Characterization of the Stochastic Edge

The stochastic edge is a statement about the *top* stationary characteristic exponent, i.e. the $q = 1$ case, along a family of stationary regimes indexed by a control parameter.

**Definition H.7** (Parameterized stationary family)**.** Let $\mathcal{P} \subset \mathbb{R}$ be a nonempty compact interval. A *parameterized stationary tangent family* on $\mathcal{P}$ consists of a collection

$$\left\{ \mathbf{G}_{\star,t}^{(1)}(p) : t \geq 0,\ p \in \mathcal{P} \right\}$$

such that, for every $p \in \mathcal{P}$, the path $t \mapsto \mathbf{G}_{\star,t}^{(1)}(p)$ satisfies Definition H.1 with $q = 1$. The corresponding top stationary exponent is denoted by

$$\lambda_1(p) := \lim_{t\to\infty} \frac{1}{2} \operatorname{Tr}\Big( \big(\mathbf{G}_{\star,t}^{(1)}(p)\big)^{-1} \dot{\mathbf{G}}_{\star,t}^{(1)}(p) \Big). \tag{249}$$

**Definition H.8** (Top stochastic edge)**.** For a parameterized stationary tangent family on $\mathcal{P}$, the *top stochastic edge set* is

$$\mathcal{E}_{\mathrm{stoch}} := \{ p \in \mathcal{P} : \lambda_1(p) = 0 \}. \tag{250}$$

Any $p_c \in \mathcal{E}_{\mathrm{stoch}}$ is called a *stochastic-edge point*.

**Proposition H.9** (Continuity of the top stationary exponent)**.** *Let $\{\mathbf{G}_{\star,t}^{(1)}(p)\}_{t\geq 0,\, p\in\mathcal{P}}$ be a parameterized stationary tangent family. Assume the following.*

1. *For every fixed $t \geq 0$, the map $p \mapsto \mathbf{G}_{\star,t}^{(1)}(p)$ is continuous on $\mathcal{P}$.*

2. *The corresponding instantaneous rates*

$$\mathfrak{g}_1(t;p) := \frac{1}{2} \operatorname{Tr}\Big( \big(\mathbf{G}_{\star,t}^{(1)}(p)\big)^{-1} \partial_t \mathbf{G}_{\star,t}^{(1)}(p) \Big)$$

   *converge to $\lambda_1(p)$ uniformly in $p$: for every $\varepsilon > 0$, there exists $T_\varepsilon$ such that*

$$\sup_{p\in\mathcal{P}} \sup_{t\geq T_\varepsilon} |\mathfrak{g}_1(t;p) - \lambda_1(p)| \leq \varepsilon. \tag{251}$$

*Then $p \mapsto \lambda_1(p)$ is continuous on $\mathcal{P}$.*

*Proof.* Fix $p \in \mathcal{P}$ and a sequence $p_n \to p$. Let $\varepsilon > 0$. By (251), there exists $T_\varepsilon$ such that

$$|\lambda_1(p_n) - \mathfrak{g}_1(T_\varepsilon; p_n)| \le \varepsilon, \qquad |\lambda_1(p) - \mathfrak{g}_1(T_\varepsilon; p)| \le \varepsilon$$

for all $n$. Hence

$$\begin{aligned} |\lambda_1(p_n) - \lambda_1(p)| &\le |\lambda_1(p_n) - \mathfrak{g}_1(T_\varepsilon; p_n)| + |\mathfrak{g}_1(T_\varepsilon; p_n) - \mathfrak{g}_1(T_\varepsilon; p)| + |\mathfrak{g}_1(T_\varepsilon; p) - \lambda_1(p)| \\ &\le 2\varepsilon + |\mathfrak{g}_1(T_\varepsilon; p_n) - \mathfrak{g}_1(T_\varepsilon; p)|. \end{aligned}$$

By continuity of $p \mapsto \mathbf{G}^{(1)}_{\star, T_\varepsilon}(p)$ and of its time derivative at fixed $T_\varepsilon$, the last term tends to zero as $n \to \infty$. Thus

$$\limsup_{n\to\infty} |\lambda_1(p_n) - \lambda_1(p)| \le 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, $\lambda_1(p_n) \to \lambda_1(p)$. □

**Theorem H.10** (Existence and uniqueness of the stochastic edge)**.** *Let $\mathcal{P} = [p_-, p_+] \subset \mathbb{R}$, and let $\{\mathbf{G}^{(1)}_{\star,t}(p)\}_{t\ge 0,\, p\in\mathcal{P}}$ be a parameterized stationary tangent family. Assume:*

1. *$p \mapsto \lambda_1(p)$ is continuous on $\mathcal{P}$;*
2. *$p \mapsto \lambda_1(p)$ is strictly monotone on $\mathcal{P}$;*
3. *$\lambda_1(p_-)\lambda_1(p_+) < 0$.*

*Then there exists a unique $p_c \in \mathcal{P}$ such that*

$$\lambda_1(p_c) = 0. \tag{252}$$

*Equivalently,*

$$\mathcal{E}_{\mathrm{stoch}} = \{p_c\}. \tag{253}$$

*If, moreover, $\lambda_1$ is strictly decreasing, then*

$$\lambda_1(p) > 0 \text{ for } p < p_c, \qquad \lambda_1(p) < 0 \text{ for } p > p_c. \tag{254}$$

*The same conclusion with the inequalities reversed holds if $\lambda_1$ is strictly increasing.*

*Proof.* By continuity of $\lambda_1$ and the sign condition at the endpoints, the intermediate value theorem yields at least one $p_c \in [p_-, p_+]$ such that $\lambda_1(p_c) = 0$. Strict monotonicity implies that such a point is unique, because a strictly monotone function on an interval can vanish at most once. This proves (252) and (253). The sign structure (254) follows immediately from strict monotonicity. □

**Proposition H.11** (Frozen-minimum reduction to the characteristic Lyapunov exponent)**.** *Assume $q = 1$. Suppose there exists $t_{\mathrm{fr}} \ge 0$ and a post-transient linear cocycle $(\Psi_t)_{t\ge 0}$ on $\mathbb{R}^k$ such that*

$$\mathbf{G}^{(1)}_{\star, t_{\mathrm{fr}}+t} = \left[\|\Psi_t u_{\mathrm{fr}}\|_2^2\right] \qquad \text{for all } t \ge 0, \tag{255}$$

*for some nonzero $u_{\mathrm{fr}} \in \mathbb{R}^k$, and that the limit*

$$\lambda_{\mathrm{fr}} := \lim_{t\to\infty} \frac{1}{t} \log \|\Psi_t u_{\mathrm{fr}}\|_2 \tag{256}$$

*exists. Then the top stationary exponent of Definition H.1 satisfies*

$$\lambda_1 = \lambda_{\mathrm{fr}}. \tag{257}$$

*In particular, if $\Psi_t$ is the frozen-minimum stochastic-gradient cocycle analyzed in Chemnitz & Engel (2025), then $\lambda_1$ coincides with their characteristic Lyapunov exponent after matching time normalizations.*

*Proof.* Since $q = 1$, the limiting Gram matrix is scalar. By (255),

$$\Lambda_{1,t_{\mathrm{fr}}+t} = \frac{1}{2(t_{\mathrm{fr}}+t)} \log \frac{\|\Psi_t u_{\mathrm{fr}}\|_2^2}{\mathbf{G}^{(1)}_{\star,0}}.$$

Therefore

$$\Lambda_{1,t_{\mathrm{fr}}+t} = \frac{t}{t_{\mathrm{fr}}+t} \cdot \frac{1}{t} \log \|\Psi_t u_{\mathrm{fr}}\|_2 - \frac{1}{2(t_{\mathrm{fr}}+t)} \log \mathbf{G}^{(1)}_{\star,0}.$$

Letting $t \to \infty$, the second term vanishes, the prefactor $t/(t_{\mathrm{fr}}+t) \to 1$, and the first factor converges to $\lambda_{\mathrm{fr}}$ by (256). Hence

$$\lim_{t\to\infty} \Lambda_{1,t} = \lambda_{\mathrm{fr}}.$$

By Theorem H.4 with $q = 1$, the left-hand side is $\lambda_1$. This proves (257). $\square$

*Proof of Theorem 3.4.* Replace the placeholder statement of Theorem 3.4 in the main text by the conjunction of Theorem H.4 and Theorem H.10. The existence of the stationary characteristic exponent is exactly (245), and the existence/uniqueness of the stochastic edge is exactly (253). The frozen-minimum consistency claim is given by Proposition H.11. $\square$

# I. Brownian Comparison and Jump-Cumulant Separation

This appendix compares the replicated tangent effective theory for true mini-batch SGD with the Brownian reference theory induced by the high-dimensional SGF /SME limit. The iterate-level Brownian reference process was already recorded in Proposition A.11; our task here is to lift that comparison from the iterate sector to the replicated tangent sector and then to the characteristic $q$-volume law.

The logic is as follows.

1. We first construct the Brownian replicated tangent fixed point by repeating the Volterra closure of Appendices E and F with the imported Brownian iterate kernels in place of the Poissonian SGD iterate kernels.
2. We then subtract the two fixed-point equations and derive a linear Volterra system for the difference

   $$\Delta\mathfrak{X}^{(q)}_T := \mathfrak{X}^{(q),\mathrm{SGD}}_{T,\star} - \mathfrak{X}^{(q),\mathrm{SGF}}_{T,\star}.$$

   The forcing in that linear system is exactly the kernel-level mismatch between the Poissonian and Brownian tangent environments.
3. Finally, we push that kernel-level difference through the log-determinant representation of Appendix G. This yields the exact finite-horizon jump-cumulant correction to the characteristic $q$-volume law.

Throughout this appendix, we work under the standing assumptions of Appendix F, including Assumption F.1, and we fix a horizon

$$0 < T < T^{\star}_{\mathrm{SGF}},$$

where $T^{\star}_{\mathrm{SGF}}$ is the Brownian well-posedness horizon from Proposition A.11. We also fix the same initial tangent law $\nu_{0,\star,q}$ and the same boundary data as in Definition E.3.

## I.1. Replicated Tangent SGF/SME System

We begin by defining the Brownian analogue of the replicated tangent effective system.

**Definition I.1** (Brownian iterate kernels)**.** Let

$$(\vartheta^{\mathrm{SGF}}_t, \xi^{\mathrm{SGF}}_t, \vartheta_\star, w_\star, \varepsilon)_{t\in[0,T]}$$

denote the imported Brownian iterate effective process from Proposition A.11. Define

$$C^{\vartheta,\vartheta,\mathrm{SGF}}_{t,s} := \mathbb{E}[\vartheta^{\mathrm{SGF}}_t \otimes \vartheta^{\mathrm{SGF}}_s] \in \mathbb{R}^{k\times k}, \qquad C^{\vartheta,\star,\mathrm{SGF}}_{t} := \mathbb{E}[\vartheta^{\mathrm{SGF}}_t \otimes \vartheta_\star] \in \mathbb{R}^{k\times k_\star}. \tag{258}$$

**Definition I.2** (Brownian physical kernel tuples)**.** A tuple

$$\mathfrak{X}^{\mathrm{SGF}} = (\mathbf{Q}^{\mathrm{SGF}}, \mathbf{S}^{\mathrm{SGF}}, \mathbf{M}^{\mathrm{SGF}}, \mathbf{Z}^{\mathrm{SGF}}) \in \mathfrak{X}_T^{(q)}$$

is called *Brownian-physical* if:

1. $\mathbf{Z}^{\mathrm{SGF}} = \mathbf{Q}^{\mathrm{SGF}}$,
2. for every $t \in [0,T]$, the block matrix

$$\Sigma_t^{\mathrm{SGF}}(\mathfrak{X}^{\mathrm{SGF}}) := \begin{pmatrix} \big[Q_{t,t}^{ab,\mathrm{SGF}}\big]_{a,b\in[q]} & \big[S_{t,t}^{a,\mathrm{SGF}}\big]_{a\in[q]} & \big[M_t^{a,\mathrm{SGF}}\big]_{a\in[q]} \\ \big[(S_{t,t}^{a,\mathrm{SGF}})^\top\big]_{a\in[q]} & C_{t,t}^{\vartheta,\vartheta,\mathrm{SGF}} & C_t^{\vartheta,\star,\mathrm{SGF}} \\ \big[(M_t^{a,\mathrm{SGF}})^\top\big]_{a\in[q]} & (C_t^{\vartheta,\star,\mathrm{SGF}})^\top & C^{\star,\star} \end{pmatrix} \succeq 0 \tag{259}$$

holds.

The corresponding closed subset of $\mathfrak{X}_T^{(q)}$ is denoted by $\mathfrak{X}_{+,T}^{(q),\mathrm{SGF}}$.

**Definition I.3** (Brownian replicated tangent coefficient family)**.** For $\mathfrak{X}^{\mathrm{SGF}} \in \mathfrak{X}_{+,T}^{(q),\mathrm{SGF}}$, let

$$(z_t^{1,\mathrm{SGF}}, \dots, z_t^{q,\mathrm{SGF}}, \xi_t^{\mathrm{SGF}}, w_\star)$$

be the jointly Gaussian lift with covariance $\Sigma_t^{\mathrm{SGF}}(\mathfrak{X}^{\mathrm{SGF}})$. Define:

$$A_t^{\mathrm{SGF}} := \Gamma + \phi\,\mathbb{E}\big[D_1 f(\xi_t^{\mathrm{SGF}}, w_\star, \varepsilon)\big]\,, \tag{260}$$

$$B_t^{a,\mathrm{SGF}}(\mathfrak{X}^{\mathrm{SGF}})[h] := \phi\,\mathbb{E}\Big[D_1^2 f(\xi_t^{\mathrm{SGF}}, w_\star, \varepsilon)[h, z_t^{a,\mathrm{SGF}}]\Big]\,, \qquad h \in \mathbb{R}^k, \tag{261}$$

$$J_t^{a,\mathrm{SGF}}(\mathfrak{X}^{\mathrm{SGF}}) := \phi\,\mathbb{E}\Big[\big(D_1 f(\xi_t^{\mathrm{SGF}}, w_\star, \varepsilon)[z_t^{a,\mathrm{SGF}}]\big) \otimes f(\xi_t^{\mathrm{SGF}}, w_\star, \varepsilon)\Big]\,, \tag{262}$$

$$H_t^{ab,\mathrm{SGF}}(\mathfrak{X}^{\mathrm{SGF}}) := \phi\,\mathbb{E}\Big[\big(D_1 f(\xi_t^{\mathrm{SGF}}, w_\star, \varepsilon)[z_t^{a,\mathrm{SGF}}]\big) \otimes \big(D_1 f(\xi_t^{\mathrm{SGF}}, w_\star, \varepsilon)[z_t^{b,\mathrm{SGF}}]\big)\Big]\,. \tag{263}$$

The lifting operator is again

$$\mathscr{G}_T^{\mathrm{SGF}}(\mathbf{Q}, \mathbf{S}, \mathbf{M}) := \mathbf{Q}. \tag{264}$$

**Theorem I.4** (Finite-horizon Brownian replicated tangent fixed point)**.** *Under the standing assumptions of this appendix, there exists a unique Brownian-physical replicated kernel tuple*

$$\mathfrak{X}_{T,\star}^{(q),\mathrm{SGF}} = \big(\mathbf{Q}_\star^{\mathrm{SGF}}, \mathbf{S}_\star^{\mathrm{SGF}}, \mathbf{M}_\star^{\mathrm{SGF}}, \mathbf{Z}_\star^{\mathrm{SGF}}\big) \in \mathfrak{X}_{+,T}^{(q),\mathrm{SGF}}$$

*solving the Brownian analogue of the Volterra fixed-point equations* (172)–(175)*, obtained by replacing*

$$(C^{\vartheta,\vartheta}, C^{\vartheta,\star}, A, B, J, H, \mathscr{G}_T)$$

*with their Brownian counterparts*

$$(C^{\vartheta,\vartheta,\mathrm{SGF}}, C^{\vartheta,\star,\mathrm{SGF}}, A^{\mathrm{SGF}}, B^{\mathrm{SGF}}, J^{\mathrm{SGF}}, H^{\mathrm{SGF}}, \mathscr{G}_T^{\mathrm{SGF}}).$$

*Proof.* The proof is identical to that of Theorem F.12. Indeed, the Gaussian interpolation estimate of Lemma F.7 and the bounded-derivative assumptions on $f$ imply that the Brownian coefficient family of Definition I.3 is admissible on bounded weighted balls, exactly as in Proposition F.8. Since the Volterra map has the same causal block-triangular form, Theorem E.11 applies verbatim. □

**Definition I.5** (Brownian scalar tangent Gram matrix and $q$-volume law)**.** For $t \in [0,T]$, define

$$\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} := \Big[\mathrm{Tr}(Q_{\star,t,t}^{ab,\mathrm{SGF}})\Big]_{a,b=1}^{q} \in \mathbb{R}^{q\times q}, \tag{265}$$

and, whenever $\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}}$ is positive definite, define

$$\Lambda_{q,t}^{\mathrm{SGF}} := \frac{1}{2t}\log\det\Big(\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}}\big(\mathbf{G}_{\star,0}^{(q)}\big)^{-1}\Big), \qquad t > 0. \tag{266}$$

*Remark* I.6. The initial Gram matrix is the same for the Poissonian and Brownian replicated systems, because both are driven by the same initial tangent law $\nu_{0,\star,q}$. Hence

$$\mathbf{G}_{\star,0}^{(q),\mathrm{SGD}} = \mathbf{G}_{\star,0}^{(q),\mathrm{SGF}} = \mathbf{G}_{\star,0}^{(q)} = \mathbf{Q}_0^{(q)}.$$

### I.2. Poisson–Brownian Cumulant Comparison

We now compare the Poissonian and Brownian replicated fixed points.

**Definition I.7** (Poisson–Brownian kernel differences)**.** Let

$$\mathfrak{X}_{T,\star}^{(q),\mathrm{SGD}} = (\mathbf{Q}_\star^{\mathrm{SGD}}, \mathbf{S}_\star^{\mathrm{SGD}}, \mathbf{M}_\star^{\mathrm{SGD}}, \mathbf{Z}_\star^{\mathrm{SGD}})$$

be the Poissonian replicated fixed point from Theorem F.12, and let

$$\mathfrak{X}_{T,\star}^{(q),\mathrm{SGF}} = (\mathbf{Q}_\star^{\mathrm{SGF}}, \mathbf{S}_\star^{\mathrm{SGF}}, \mathbf{M}_\star^{\mathrm{SGF}}, \mathbf{Z}_\star^{\mathrm{SGF}})$$

be the Brownian replicated fixed point from Theorem I.4. Define the differences

$$\Delta Q_{t,s}^{ab} := Q_{\star,t,s}^{ab,\mathrm{SGD}} - Q_{\star,t,s}^{ab,\mathrm{SGF}}, \tag{267}$$

$$\Delta S_{t,s}^{a} := S_{\star,t,s}^{a,\mathrm{SGD}} - S_{\star,t,s}^{a,\mathrm{SGF}}, \tag{268}$$

$$\Delta M_{t}^{a} := M_{\star,t}^{a,\mathrm{SGD}} - M_{\star,t}^{a,\mathrm{SGF}}, \tag{269}$$

$$\Delta \mathbf{G}_{q,t} := \mathbf{G}_{\star,t}^{(q),\mathrm{SGD}} - \mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} = \Big[\mathrm{Tr}(\Delta Q_{t,t}^{ab})\Big]_{a,b=1}^{q}. \tag{270}$$

We also define the iterate-kernel differences

$$\Delta C_{t,s}^{\vartheta,\vartheta} := C_{t,s}^{\vartheta,\vartheta,\mathrm{SGD}} - C_{t,s}^{\vartheta,\vartheta,\mathrm{SGF}}, \tag{271}$$

$$\Delta C_{t}^{\vartheta,\star} := C_{t}^{\vartheta,\star,\mathrm{SGD}} - C_{t}^{\vartheta,\star,\mathrm{SGF}}, \tag{272}$$

and the coefficient differences

$$\Delta A_t := A_t^{\mathrm{SGD}} - A_t^{\mathrm{SGF}}, \tag{273}$$

$$\Delta B_t^a := B_t^{a,\mathrm{SGD}} - B_t^{a,\mathrm{SGF}}, \tag{274}$$

$$\Delta J_t^a := J_t^{a,\mathrm{SGD}} - J_t^{a,\mathrm{SGF}}, \tag{275}$$

$$\Delta H_t^{ab} := H_t^{ab,\mathrm{SGD}} - H_t^{ab,\mathrm{SGF}}. \tag{276}$$

**Proposition I.8** (Linear Volterra system for the Poisson–Brownian difference)**.** *For every $a, b \in [q]$ and $t, s \in [0, T]$, the difference kernels satisfy*

$$\Delta M_t^a = -\int_0^t \bar{\eta}(r)\Big(A_r^{\mathrm{SGD}} \Delta M_r^a + \mathcal{C}_{M,r}^a\Big)\, dr, \tag{277}$$

$$\Delta S_{t,s}^a = -\int_0^t \bar{\eta}(r)\Big(A_r^{\mathrm{SGD}} \Delta S_{r,s}^a + \mathcal{C}_{S,r,s}^a\Big)\, dr + \int_0^{t\wedge s} \bar{\eta}(r)^2 \Delta J_r^a \, dr, \tag{278}$$

$$\begin{aligned} \Delta Q_{t,s}^{ab} = &-\int_0^t \bar{\eta}(r)\Big(A_r^{\mathrm{SGD}} \Delta Q_{r,s}^{ab} + \mathcal{C}_{Q,r,s}^{ab,\mathrm{L}}\Big)\, dr \\ &- \int_0^s \bar{\eta}(r)\Big(\Delta Q_{t,r}^{ab} (A_r^{\mathrm{SGD}})^\top + \mathcal{C}_{Q,t,r}^{ab,\mathrm{R}}\Big)\, dr + \int_0^{t\wedge s} \bar{\eta}(r)^2 \Delta H_r^{ab}\, dr, \end{aligned} \tag{279}$$

*where the forcing terms are*

$$\mathcal{C}_{M,r}^a := \Delta A_r\, M_r^{a,\mathrm{SGF}} + B_r^{a,\mathrm{SGD}}\, \Delta C_r^{\vartheta,\star} + \Delta B_r^a\, C_r^{\vartheta,\star,\mathrm{SGF}}, \tag{280}$$

$$\mathcal{C}_{S,r,s}^a := \Delta A_r\, S_{r,s}^{a,\mathrm{SGF}} + B_r^{a,\mathrm{SGD}}\, \Delta C_{r,s}^{\vartheta,\vartheta} + \Delta B_r^a\, C_{r,s}^{\vartheta,\vartheta,\mathrm{SGF}}, \tag{281}$$

$$\mathcal{C}_{Q,r,s}^{ab,\mathrm{L}} := \Delta A_r\, Q_{r,s}^{ab,\mathrm{SGF}} + B_r^{a,\mathrm{SGD}}\, (\Delta S_{s,r}^b)^\top + \Delta B_r^a\, (S_{s,r}^{b,\mathrm{SGF}})^\top, \tag{282}$$

$$\mathcal{C}_{Q,t,r}^{ab,\mathrm{R}} := Q_{t,r}^{ab,\mathrm{SGF}} (\Delta A_r)^\top + \Delta S_{t,r}^a (B_r^{b,\mathrm{SGD}})^\top + S_{t,r}^{a,\mathrm{SGF}} (\Delta B_r^b)^\top. \tag{283}$$

*Proof.* Subtract the Brownian fixed-point equations from the Poissonian fixed-point equations. For $M^a$, both systems start from the same initial datum $M_0^a$, so the initial terms cancel. One obtains

$$\begin{aligned}\Delta M_t^a &= -\int_0^t \bar{\eta}(r)\Big(A_r^{\mathrm{SGD}} M_r^{a,\mathrm{SGD}} + B_r^{a,\mathrm{SGD}} C_r^{\vartheta,\star,\mathrm{SGD}} - A_r^{\mathrm{SGF}} M_r^{a,\mathrm{SGF}} - B_r^{a,\mathrm{SGF}} C_r^{\vartheta,\star,\mathrm{SGF}}\Big)\,dr \\ &= -\int_0^t \bar{\eta}(r)\Big(A_r^{\mathrm{SGD}} \Delta M_r^a + \Delta A_r M_r^{a,\mathrm{SGF}} + B_r^{a,\mathrm{SGD}} \Delta C_r^{\vartheta,\star} + \Delta B_r^a C_r^{\vartheta,\star,\mathrm{SGF}}\Big)\,dr,\end{aligned}$$

which is (277).

The proof of (278) is identical, with the additional noise term $\int_0^{t\wedge s} \bar{\eta}(r)^2 \Delta J_r^a\,dr$ surviving after subtraction.

For (279), subtract the two $Q$-equations and add/subtract the mixed terms

$$B_r^{a,\mathrm{SGD}} (S_{s,r}^{b,\mathrm{SGF}})^\top, \qquad S_{t,r}^{a,\mathrm{SGF}} (B_r^{b,\mathrm{SGD}})^\top,$$

to separate the difference into a part involving $\Delta Q, \Delta S$ and a part involving the coefficient mismatch $\Delta A, \Delta B, \Delta H$. This gives exactly (279) with the forcing terms (282) and (283). $\square$

**Corollary I.9** (Lipschitz Brownian comparison)**.** *Fix $\lambda \geq \lambda_\star$ and a radius $R$ for which both fixed points belong to the invariant ball of Theorem E.11. Then there exists a constant $L_{q,T}^{\mathrm{cmp}} < \infty$ such that*

$$\|\mathfrak{X}_{T,\star}^{(q),\mathrm{SGD}} - \mathfrak{X}_{T,\star}^{(q),\mathrm{SGF}}\|_{\lambda,T} \leq L_{q,T}^{\mathrm{cmp}} \left( \sup_{t\in[0,T]} \|\Delta C_t^{\vartheta,\star}\|_{\mathrm{F}} + \sup_{t,s\in[0,T]} \|\Delta C_{t,s}^{\vartheta,\vartheta}\|_{\mathrm{F}} \right). \tag{284}$$

*Consequently,*

$$\sup_{t\in[0,T]} \|\Delta \mathbf{G}_{q,t}\|_{\mathrm{F}} \leq q\sqrt{k}\, L_{q,T}^{\mathrm{cmp}} \left( \sup_{t\in[0,T]} \|\Delta C_t^{\vartheta,\star}\|_{\mathrm{F}} + \sup_{t,s\in[0,T]} \|\Delta C_{t,s}^{\vartheta,\vartheta}\|_{\mathrm{F}} \right). \tag{285}$$

*Proof.* This is exactly the Lipschitz dependence statement of Corollary E.12 applied to the two coefficient families generated by the Poissonian and Brownian iterate inputs. Under the standing assumptions and the bounded-derivative hypotheses on $f$, the Gaussian interpolation estimate of Lemma F.7 implies that the coefficient differences $\Delta A, \Delta B, \Delta J, \Delta H$ are bounded linearly by the iterate-kernel differences $\Delta C^{\vartheta,\vartheta}$ and $\Delta C^{\vartheta,\star}$. Hence the fixed-point difference obeys (284). The bound (285) then follows from the definition of $\Delta \mathbf{G}_{q,t}$ and the inequality $|\operatorname{Tr}(A)| \leq \sqrt{k}\|A\|_{\mathrm{F}}$. $\square$

We now isolate the finite-horizon correction to the characteristic $q$-volume law.

**Definition I.10** (Finite-horizon jump-cumulant correction)**.** Assume that, for some $0 < \tau \leq T < T_{\mathrm{SGF}}^\star$, both $\mathbf{G}_{\star,t}^{(q),\mathrm{SGD}}$ and $\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}}$ are positive definite on $[\tau, T]$. For $t \in [\tau, T]$, define the *finite-horizon jump-cumulant correction*

$$\mathcal{J}_{q,t} := \Lambda_{q,t}^{\mathrm{SGD}} - \Lambda_{q,t}^{\mathrm{SGF}} = \frac{1}{2t} \log\det\Big(\mathbf{G}_{\star,t}^{(q),\mathrm{SGD}} \big(\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}}\big)^{-1}\Big). \tag{286}$$

**Lemma I.11** (Exact integral representation of the jump-cumulant correction)**.** *Assume Definition I.10. Then for every $t \in [\tau, T]$,*

$$\mathcal{J}_{q,t} = \frac{1}{2t} \int_0^1 \operatorname{Tr}\Big( \big(\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta \Delta \mathbf{G}_{q,t}\big)^{-1} \Delta \mathbf{G}_{q,t} \Big)\, d\theta. \tag{287}$$

*Proof.* Set

$$A_\theta := \mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta \Delta \mathbf{G}_{q,t}, \qquad \theta \in [0,1].$$

Because $\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}}$ and $\mathbf{G}_{\star,t}^{(q),\mathrm{SGD}}$ are positive definite, every convex combination $A_\theta$ is positive definite. Hence

$$\frac{d}{d\theta} \log\det A_\theta = \operatorname{Tr}(A_\theta^{-1} \Delta \mathbf{G}_{q,t})$$

by Jacobi's formula. Integrating from $\theta = 0$ to $\theta = 1$ gives

$$\log\det \mathbf{G}_{\star,t}^{(q),\mathrm{SGD}} - \log\det \mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} = \int_0^1 \mathrm{Tr}(A_\theta^{-1}\Delta\mathbf{G}_{q,t})\,d\theta.$$

Divide by $2t$ and use Definition I.10. □

**Theorem I.12** (Jump–diffusion separation at the multiplicative level)**.** *Fix $0 < \tau \le T < T^\star_{\mathrm{SGF}}$. Assume that the interval $[\tau, T]$ is q-volume admissible for both the Poissonian and Brownian replicated tangent systems, and define the common positivity margin*

$$\gamma_{\tau,T}^{(q),\mathrm{com}} := \frac{1}{2}\min\left\{\inf_{t\in[\tau,T]}\lambda_{\min}\big(\mathbf{G}_{\star,t}^{(q),\mathrm{SGD}}\big),\ \inf_{t\in[\tau,T]}\lambda_{\min}\big(\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}}\big)\right\} > 0. \tag{288}$$

*Then, for every $t \in [\tau, T]$, the finite-horizon jump-cumulant correction satisfies*

$$|\mathcal{J}_{q,t}| \le \frac{\sqrt{q}}{2\tau\,\gamma_{\tau,T}^{(q),\mathrm{com}}}\|\Delta\mathbf{G}_{q,t}\|_{\mathrm{F}}, \tag{289}$$

*and hence*

$$\sup_{t\in[\tau,T]}|\mathcal{J}_{q,t}| \le \frac{\sqrt{q}}{2\tau\,\gamma_{\tau,T}^{(q),\mathrm{com}}}\sup_{t\in[\tau,T]}\|\Delta\mathbf{G}_{q,t}\|_{\mathrm{F}} \tag{290}$$

$$\le \frac{q\sqrt{kq}\,L_{q,T}^{\mathrm{cmp}}}{2\tau\,\gamma_{\tau,T}^{(q),\mathrm{com}}}\left(\sup_{t\in[0,T]}\|\Delta C_t^{\vartheta,\star}\|_{\mathrm{F}} + \sup_{t,s\in[0,T]}\|\Delta C_{t,s}^{\vartheta,\vartheta}\|_{\mathrm{F}}\right). \tag{291}$$

*In particular, the difference between the Poissonian and Brownian characteristic q-volume laws is controlled entirely by the mismatch of the two iterate effective environments.*

*Proof.* Fix $t \in [\tau, T]$. By Lemma I.11,

$$\mathcal{J}_{q,t} = \frac{1}{2t}\int_0^1 \mathrm{Tr}\Big((\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta\Delta\mathbf{G}_{q,t})^{-1}\Delta\mathbf{G}_{q,t}\Big)\,d\theta.$$

For every $\theta \in [0, 1]$, the matrix

$$\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta\Delta\mathbf{G}_{q,t} = (1-\theta)\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta\mathbf{G}_{\star,t}^{(q),\mathrm{SGD}}$$

is positive definite and satisfies

$$\lambda_{\min}\big(\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta\Delta\mathbf{G}_{q,t}\big) \ge 2\gamma_{\tau,T}^{(q),\mathrm{com}}.$$

Hence

$$\Big\|(\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta\Delta\mathbf{G}_{q,t})^{-1}\Big\|_{\mathrm{op}} \le \frac{1}{2\gamma_{\tau,T}^{(q),\mathrm{com}}}.$$

Using $|\mathrm{Tr}(AB)| \le \|A\|_{\mathrm{op}}\|B\|_*$ and $\|B\|_* \le \sqrt{q}\,\|B\|_{\mathrm{F}}$, we obtain

$$\begin{aligned}|\mathcal{J}_{q,t}| &\le \frac{1}{2t}\int_0^1 \Big\|(\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta\Delta\mathbf{G}_{q,t})^{-1}\Big\|_{\mathrm{op}}\|\Delta\mathbf{G}_{q,t}\|_*\,d\theta \\ &\le \frac{\sqrt{q}}{2t}\cdot\frac{1}{2\gamma_{\tau,T}^{(q),\mathrm{com}}}\|\Delta\mathbf{G}_{q,t}\|_{\mathrm{F}} \le \frac{\sqrt{q}}{2\tau\,\gamma_{\tau,T}^{(q),\mathrm{com}}}\|\Delta\mathbf{G}_{q,t}\|_{\mathrm{F}},\end{aligned}$$

which proves (289). Taking the supremum over $t \in [\tau, T]$ gives (290). Finally, (291) follows from Corollary I.9. □

**Corollary I.13** (Vanishing of the jump correction in the linear regime)**.** *Assume, in addition, that the learner-side nonlinearity is linear in its first argument, i.e.*

$$D_1^2 f(\xi, w_\star, \varepsilon) \equiv 0 \qquad \textit{for all } (\xi, w_\star, \varepsilon), \tag{292}$$

*and that the imported iterate effective kernels coincide:*

$$C_{t,s}^{\vartheta,\vartheta,\mathrm{SGD}} = C_{t,s}^{\vartheta,\vartheta,\mathrm{SGF}}, \qquad C_t^{\vartheta,\star,\mathrm{SGD}} = C_t^{\vartheta,\star,\mathrm{SGF}} \qquad \textit{for all } t,s \in [0,T]. \tag{293}$$

*Then*

$$\mathcal{J}_{q,t} \equiv 0 \qquad \textit{for all } t \in (0,T]. \tag{294}$$

*In particular, in the linear models where the iterate dynamics of* SGD *and their Brownian reference limit coincide, the multiplicative jump correction is exactly zero, in agreement with the linear deterministic-equivalence picture (Atanasov et al., 2025; Fan & Wang, 2026).*

*Proof.* If $D_1^2 f \equiv 0$, then $B_t^{a,\mathrm{SGD}} = B_t^{a,\mathrm{SGF}} \equiv 0$ for every $a \in [q]$. Moreover, $D_1 f$ is constant in $\xi$, so $A_t^{\mathrm{SGD}} = A_t^{\mathrm{SGF}}$, $J_t^{a,\mathrm{SGD}} = J_t^{a,\mathrm{SGF}}$, and $H_t^{ab,\mathrm{SGD}} = H_t^{ab,\mathrm{SGF}}$ whenever the iterate kernels coincide. Thus

$$\Delta A_t = \Delta B_t^a = \Delta J_t^a = \Delta H_t^{ab} = 0, \qquad \Delta C_{t,s}^{\vartheta,\vartheta} = 0, \qquad \Delta C_t^{\vartheta,\star} = 0.$$

The linear Volterra system of Proposition I.8 then has zero forcing and zero initial data, so uniqueness implies

$$\Delta M_t^a = \Delta S_{t,s}^a = \Delta Q_{t,s}^{ab} = 0 \qquad \text{for all } a,b \in [q],\ t,s \in [0,T].$$

Hence $\Delta \mathbf{G}_{q,t} = 0$ for all $t$, and Definition I.10 gives $\mathcal{J}_{q,t} = 0$. □

**Proposition I.14** (Sign criterion for nontrivial jump separation)**.** *Fix $t \in [\tau, T]$ under the assumptions of Theorem I.12. Then:*

1. *If $\Delta \mathbf{G}_{q,t} \succeq 0$ and $\Delta \mathbf{G}_{q,t} \neq 0$, then $\mathcal{J}_{q,t} > 0$.*
2. *If $\Delta \mathbf{G}_{q,t} \preceq 0$ and $\Delta \mathbf{G}_{q,t} \neq 0$, then $\mathcal{J}_{q,t} < 0$.*

*Proof.* Assume $\Delta \mathbf{G}_{q,t} \succeq 0$ and $\Delta \mathbf{G}_{q,t} \neq 0$. For every $\theta \in [0,1]$, the matrix

$$\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta \Delta \mathbf{G}_{q,t}$$

is positive definite, so its inverse is positive definite as well. Therefore

$$\mathrm{Tr}\Big( (\mathbf{G}_{\star,t}^{(q),\mathrm{SGF}} + \theta \Delta \mathbf{G}_{q,t})^{-1} \Delta \mathbf{G}_{q,t} \Big) \geq 0$$

for every $\theta$, and it is strictly positive for at least one $\theta$ because $\Delta \mathbf{G}_{q,t} \neq 0$. The integral representation (287) then gives $\mathcal{J}_{q,t} > 0$. The case $\Delta \mathbf{G}_{q,t} \preceq 0$ is identical after replacing $\Delta \mathbf{G}_{q,t}$ by $-\Delta \mathbf{G}_{q,t}$. □

*Proof of Theorem 4.1.* Replace the placeholder statement of Theorem 4.1 in the main text by Theorem I.12, together with the linear-regime reduction of Corollary I.13. The comparison of the two replicated tangent theories is furnished by Theorem I.4 and Proposition I.8, while the exact correction formula is (287). □

# J. Rank-One Logistic Specialization

This appendix gives a fully explicit specialization of the general replicated tangent theory to the rank-one logistic classification model of Assumption A.7. The goal is not to re-prove the entire high-dimensional DMFT from first principles, but to use the general kernel theory of Appendices F–H to identify an explicit low-dimensional sector decomposition and an explicit stochastic-edge equation.

The key simplification is that, once the iterate sector is assumed to be signal-aligned, the tangent dynamics split into two scalar sectors:

1. a *signal* sector parallel to the teacher direction, and
2. an *orthogonal bulk* sector transverse to that direction.

Each sector has an explicit scalar growth law. For $q = 1$, the top stochastic edge is therefore the first positive root of a one-dimensional deterministic equation.

### J.1. Reduced Iterate Dynamics

We first record the scalar logistic objects used throughout the appendix.

**Definition J.1** (Scalar logistic derivatives)**.** Let

$$\ell(u; y) := \log(1 + e^{-yu}), \qquad u \in \mathbb{R},\ y \in \{-1, 1\}.$$

Define

$$\ell_1(u; y) := \partial_u \ell(u; y) = -\frac{y}{1 + e^{yu}}, \tag{295}$$

$$\ell_2(u; y) := \partial_u^2 \ell(u; y) = \frac{e^{yu}}{(1 + e^{yu})^2}, \tag{296}$$

$$\ell_3(u; y) := \partial_u^3 \ell(u; y) = y\, \frac{e^{yu}(1 - e^{yu})}{(1 + e^{yu})^3}. \tag{297}$$

*Remark* J.2. The derivatives in Definition J.1 satisfy

$$|\ell_1(u; y)| \le 1, \qquad 0 < \ell_2(u; y) \le \frac{1}{4}, \qquad |\ell_3(u; y)| \le \frac{1}{6\sqrt{3}},$$

uniformly in $u \in \mathbb{R}$ and $y \in \{-1, 1\}$. In particular, the bounded-derivative assumptions of Appendix A hold automatically in the present specialization.

**Definition J.3** (Teacher-label channel and scalar response functions)**.** Let $W \sim \mathcal{N}(0, 1)$ and $\varepsilon \sim \mathrm{Unif}[0, 1]$ be independent, and define the rank-one logistic teacher label

$$Y_\beta(W, \varepsilon) := 2\mathbf{1}\left\{\varepsilon \le \frac{1}{1 + e^{-\beta W}}\right\} - 1 \in \{-1, 1\}. \tag{298}$$

For $m \in \mathbb{R}$, define

$$\Psi_\beta(m) := \mathbb{E}\big[W\, \ell_1\big(mW; Y_\beta(W, \varepsilon)\big)\big]\,, \tag{299}$$

$$\chi_\parallel(m) := \mathbb{E}\big[W^2\, \ell_2\big(mW; Y_\beta(W, \varepsilon)\big)\big]\,, \tag{300}$$

$$\chi_\perp(m) := \mathbb{E}\big[\ell_2\big(mW; Y_\beta(W, \varepsilon)\big)\big]\,, \tag{301}$$

$$\tau_\beta(m) := \mathbb{E}\big[W\, \ell_3\big(mW; Y_\beta(W, \varepsilon)\big)\big]\,, \tag{302}$$

$$\upsilon_\parallel(m) := \mathbb{E}\Big[W^2\, \ell_2\big(mW; Y_\beta(W, \varepsilon)\big)^2\Big]\,, \tag{303}$$

$$\upsilon_\perp(m) := \mathbb{E}\Big[\ell_2\big(mW; Y_\beta(W, \varepsilon)\big)^2\Big]\,. \tag{304}$$

**Lemma J.4** (Basic properties of the scalar response functions)**.** *The functions in Definition J.3 satisfy:*

1. $\Psi_\beta \in C^1(\mathbb{R})$ *and*
$$\Psi_\beta'(m) = \chi_\parallel(m) \qquad \textit{for all } m \in \mathbb{R}. \tag{305}$$
2. $\chi_\parallel(m) > 0$, $\chi_\perp(m) > 0$, $\upsilon_\parallel(m) > 0$, *and* $\upsilon_\perp(m) > 0$ *for every* $m \in \mathbb{R}$.
3. $\Psi_\beta$ *is globally Lipschitz and*
$$|\Psi_\beta(m)| \le \mathbb{E}|W| = \sqrt{\frac{2}{\pi}} \qquad \textit{for all } m \in \mathbb{R}.$$

*Proof.* Because $|\ell_1(u;y)| \le 1$ and $\ell_2(u;y) \le 1/4$, all expressions are integrable. Differentiating under the expectation in (299) is justified by dominated convergence and gives

$$\Psi'_\beta(m) = \mathbb{E}\big[W^2 \ell_2(mW; Y_\beta(W,\varepsilon))\big] = \chi_\|(m),$$

which proves (305). Since $\ell_2(mW; Y_\beta(W,\varepsilon)) > 0$ almost surely and $W^2 > 0$ with positive probability, both $\chi_\|$ and $\chi_\perp$ are strictly positive. The same argument applies to $\upsilon_\|$ and $\upsilon_\perp$. Finally,

$$|\Psi_\beta(m)| \le \mathbb{E}[|W|\, |\ell_1(mW; Y_\beta(W,\varepsilon))|] \le \mathbb{E}|W|,$$

and the Lipschitz property follows from the bound $|\Psi'_\beta(m)| = \chi_\|(m) \le \mathbb{E}[W^2]/4 = 1/4$. □

The explicit theorem in this appendix is derived under the following structural restriction on the iterate effective theory.

**Assumption J.5** (Signal-aligned iterate regime)**.** In addition to Assumption A.7, assume that the limiting iterate effective theory from Proposition A.9 is signal-aligned, i.e. there exists a continuous scalar path $m : [0,\infty) \to \mathbb{R}$ such that

$$C^{\vartheta,\vartheta}_{t,s} = m_t m_s, \qquad C^{\vartheta,\star}_t = m_t \qquad \text{for all } t, s \ge 0. \tag{306}$$

Equivalently, the projected iterate field satisfies

$$\xi_t = m_t W \qquad \text{in law for every } t \ge 0. \tag{307}$$

*Remark* J.6. Assumption J.5 is a model restriction on the limiting iterate regime, not a claim about all finite-$d$ trajectories. It is exactly the regime in which the tangent sector closes to two scalar channels and the stochastic edge becomes explicit.

**Proposition J.7** (Reduced scalar iterate dynamics)**.** *Under Assumption J.5, the alignment path $m_t$ satisfies the scalar Volterra equation*

$$m_t = m_0 - \int_0^t \bar\eta(r)\Big[\lambda m_r + \phi \Psi_\beta(m_r)\Big]\, dr. \tag{308}$$

*Consequently, if $\bar\eta(t) \equiv \eta > 0$ is constant, then $m_t$ solves the scalar ODE*

$$\dot m_t = -\eta\Big[\lambda m_t + \phi\Psi_\beta(m_t)\Big]. \tag{309}$$

*Moreover, the map*

$$F_{\mathrm{it}}(m) := \lambda m + \phi \Psi_\beta(m) \tag{310}$$

*is $C^1$, strictly increasing, and has a unique zero $m_\infty \in \mathbb{R}$.*

*Proof.* Under Assumption J.5, the imported iterate field is $\xi_t = m_t W$, so the rank-one specialization of the iterate fixed-point equation from Proposition A.9 reduces to (308). The ODE (309) is its constant-$\eta$ form.

Now

$$F'_{\mathrm{it}}(m) = \lambda + \phi\Psi'_\beta(m) = \lambda + \phi\chi_\|(m) > 0$$

by Lemma J.4. Thus $F_{\mathrm{it}}$ is strictly increasing. Since $|\Psi_\beta(m)| \le \sqrt{2/\pi}$,

$$F_{\mathrm{it}}(m) = \lambda m + O(1) \qquad \text{as } |m| \to \infty.$$

Hence

$$\lim_{m\to-\infty} F_{\mathrm{it}}(m) = -\infty, \qquad \lim_{m\to+\infty} F_{\mathrm{it}}(m) = +\infty.$$

By continuity and strict monotonicity, $F_{\mathrm{it}}$ has a unique zero $m_\infty$. □

### J.2. Signal–Bulk Tangent Decomposition

We now identify the explicit scalar tangent channels. The first is the signal channel parallel to the teacher direction; the second is the orthogonal bulk channel.

**Definition J.8** (Initial signal and bulk weights)**.** For each replica $a \in [q]$, define its limiting initial signal coefficient by

$$\mu_0^a := M_0^a \in \mathbb{R}, \tag{311}$$

and define the initial signal Gram and bulk Gram by

$$\Sigma_0^{ab} := \mu_0^a \mu_0^b, \tag{312}$$

$$R_0^{ab} := Q_0^{ab} - \Sigma_0^{ab}, \tag{313}$$

where $Q_0^{ab}$ is the $(a, b)$-entry of the initial tangent Gram matrix $\mathbf{Q}_0^{(q)}$. Since $C^{\star,\star} = 1$, the matrix $(R_0^{ab})_{a,b\in[q]}$ is symmetric positive semidefinite.

**Definition J.9** (Sector growth coefficients)**.** For $t \geq 0$, define

$$\kappa_\parallel(t) := \lambda + \phi\chi_\perp(m_t) + \phi m_t \tau_\beta(m_t), \tag{314}$$

$$\kappa_\perp(t) := \lambda + \phi\chi_\perp(m_t), \tag{315}$$

$$\Gamma_\parallel(t) := -2\bar{\eta}(t)\kappa_\parallel(t) + \bar{\eta}(t)^2\phi\upsilon_\parallel(m_t), \tag{316}$$

$$\Gamma_\perp(t) := -2\bar{\eta}(t)\kappa_\perp(t) + \bar{\eta}(t)^2\phi\upsilon_\perp(m_t). \tag{317}$$

**Proposition J.10** (Signal–bulk splitting in the rank-one logistic model)**.** *Assume Assumption J.5. Then the one-time limiting replica Gram entries admit the decomposition*

$$Q_{\star,t,t}^{ab} = \Sigma_t^{ab} + R_t^{ab}, \qquad a, b \in [q],\ t \geq 0, \tag{318}$$

*where $\Sigma_t = (\Sigma_t^{ab})_{a,b\in[q]}$ and $R_t = (R_t^{ab})_{a,b\in[q]}$ are the unique symmetric matrix-valued solutions of the scalar linear systems*

$$\dot{\Sigma}_t^{ab} = \Gamma_\parallel(t)\,\Sigma_t^{ab}, \qquad \Sigma_0^{ab} \text{ given by (312)}, \tag{319}$$

$$\dot{R}_t^{ab} = \Gamma_\perp(t)\,R_t^{ab}, \qquad R_0^{ab} \text{ given by (313)}. \tag{320}$$

*Equivalently,*

$$\Sigma_t^{ab} = \Sigma_0^{ab} \exp\Big(\int_0^t \Gamma_\parallel(r)\,dr\Big), \tag{321}$$

$$R_t^{ab} = R_0^{ab} \exp\Big(\int_0^t \Gamma_\perp(r)\,dr\Big). \tag{322}$$

*Proof.* Because $k = k_\star = 1$, all kernels in Appendices F–H are scalar. Under Assumption J.5, the projected iterate field is

$$\xi_t = m_t W,$$

where $W \sim \mathcal{N}(0, 1)$. For each replica $a$, the Gaussian lift of the tangent projected field may therefore be decomposed as

$$z_t^a = \mu_t^a W + \zeta_t^a, \tag{323}$$

where

$$\mathbb{E}[\zeta_t^a W] = 0, \qquad \mathbb{E}[\zeta_t^a \zeta_t^b] = R_t^{ab}, \qquad \mathbb{E}[z_t^a z_t^b] = Q_{\star,t,t}^{ab},$$

and $\mu_t^a$ is the signal-direction scalar. Since the Gaussian lift is fully determined by second moments,

$$Q_{\star,t,t}^{ab} = \Sigma_t^{ab} + R_t^{ab}, \qquad \Sigma_t^{ab} := \mathbb{E}[\mu_t^a \mu_t^b].$$

We now specialize the scalar coefficient formulas of Appendix F.

**Signal sector.** The coefficient $A_t$ from (192) becomes

$$A_t = \lambda + \phi \mathbb{E}[\ell_2(m_t W; Y_\beta(W, \varepsilon))] = \kappa_\perp(t).$$

The mixed coefficient $B_t^a[h]$ from (193) becomes

$$B_t^a[h] = \phi h \, \mathbb{E}[\ell_3(m_t W; Y_\beta(W, \varepsilon)) \, z_t^a].$$

Using the decomposition (323) and the orthogonality $\mathbb{E}[\zeta_t^a W] = 0$, we obtain

$$\mathbb{E}[\ell_3(m_t W; Y_\beta) z_t^a] = \mu_t^a \mathbb{E}[W \, \ell_3(m_t W; Y_\beta)] = \mu_t^a \tau_\beta(m_t).$$

Since the iterate signal coefficient is $m_t$, the corresponding signal second-moment contribution is

$$-2\bar{\eta}(t)\Big(\lambda + \phi \chi_\perp(m_t) + \phi m_t \tau_\beta(m_t)\Big)\Sigma_t^{ab}.$$

The $H$-term from (195) contributes

$$\phi \bar{\eta}(t)^2 \mathbb{E}\big[\ell_2(m_t W; Y_\beta)^2 z_t^a z_t^b\big] .$$

Again using (323), the cross terms vanish by orthogonality, and the signal part is

$$\phi \bar{\eta}(t)^2 \upsilon_\parallel(m_t) \Sigma_t^{ab}.$$

Collecting the signal-only contributions yields (319).

**Orthogonal bulk sector.** In the orthogonal complement of the teacher direction, the scalar projected field is independent of $W$ and centered. Hence the mixed $B$-term vanishes identically on the bulk sector, while the drift is still $A_t = \kappa_\perp(t)$. The $H$-term contributes

$$\phi \bar{\eta}(t)^2 \mathbb{E}\big[\ell_2(m_t W; Y_\beta)^2\big] \, R_t^{ab} = \phi \bar{\eta}(t)^2 \upsilon_\perp(m_t) R_t^{ab}.$$

Therefore $R_t^{ab}$ solves (320).

Since both systems are scalar linear ODEs, their explicit solutions are (321)–(322). This proves the decomposition (318). $\square$

**Corollary J.11** (Strict contraction criterion for the orthogonal bulk)**.** *Under the assumptions of Proposition J.10, suppose there exist $t_0 \geq 0$ and $c_\perp > 0$ such that*

$$\Gamma_\perp(t) \leq -c_\perp \qquad \textit{for all } t \geq t_0. \tag{324}$$

*Then the orthogonal bulk decays exponentially:*

$$\|R_t\|_{\mathrm{F}} \leq e^{-c_\perp (t - t_0)} \|R_{t_0}\|_{\mathrm{F}} \qquad \textit{for all } t \geq t_0. \tag{325}$$

*Proof.* From (322),

$$R_t^{ab} = R_{t_0}^{ab} \exp\Big(\int_{t_0}^{t} \Gamma_\perp(r) \, dr\Big).$$

If $\Gamma_\perp(r) \leq -c_\perp$ on $[t_0, \infty)$, then

$$|R_t^{ab}| \leq |R_{t_0}^{ab}| \, e^{-c_\perp (t - t_0)}.$$

Taking the Frobenius norm over the $q \times q$ replica matrix yields (325). $\square$

### J.3. Scalar Edge Equation and Uniqueness of the Critical Root

We now specialize further to the constant-stepsize stationary regime and derive the explicit stochastic-edge equation.

**Assumption J.12** (Constant-stepsize stationary rank-one regime)**.** In addition to Assumption J.5, assume:

1. $\bar{\eta}(t) \equiv \eta > 0$ is constant;
2. the iterate alignment path $m_t$ from Proposition J.7 converges to the unique stationary root $m_\infty$ of $F_{\mathrm{it}}(m) = 0$.

**Definition J.13** (Asymptotic sector exponents)**.** Under Assumption J.12, define the signal and bulk stationary exponents by

$$\lambda_{\|}(\eta) := -\eta\Big[\lambda + \phi\chi_{\perp}(m_{\infty}) + \phi m_{\infty}\tau_{\beta}(m_{\infty})\Big] + \frac{\eta^2}{2}\phi v_{\|}(m_{\infty}), \tag{326}$$

$$\lambda_{\perp}(\eta) := -\eta\Big[\lambda + \phi\chi_{\perp}(m_{\infty})\Big] + \frac{\eta^2}{2}\phi v_{\perp}(m_{\infty}). \tag{327}$$

The associated positive critical roots are

$$\eta_{\|,c} := \frac{2\big[\lambda + \phi\chi_{\perp}(m_{\infty}) + \phi m_{\infty}\tau_{\beta}(m_{\infty})\big]}{\phi v_{\|}(m_{\infty})}, \tag{328}$$

$$\eta_{\perp,c} := \frac{2\big[\lambda + \phi\chi_{\perp}(m_{\infty})\big]}{\phi v_{\perp}(m_{\infty})}. \tag{329}$$

*Remark* J.14. The denominators in (328)–(329) are strictly positive by Lemma J.4. The numerator in (329) is strictly positive because $\lambda > 0$ and $\chi_{\perp} > 0$. For the signal sector, positivity of the numerator is the mild stability condition

$$\lambda + \phi\chi_{\perp}(m_{\infty}) + \phi m_{\infty}\tau_{\beta}(m_{\infty}) > 0.$$

Whenever this holds, both sector roots are well defined and strictly positive.

**Lemma J.15** (Asymptotic sector growth laws)**.** *Under Assumption J.12, for every $a, b \in [q]$,*

$$\frac{1}{2t}\log\frac{\Sigma_t^{ab}}{\Sigma_0^{ab}} \longrightarrow \lambda_{\|}(\eta) \qquad \textit{whenever } \Sigma_0^{ab} > 0, \tag{330}$$

$$\frac{1}{2t}\log\frac{R_t^{ab}}{R_0^{ab}} \longrightarrow \lambda_{\perp}(\eta) \qquad \textit{whenever } R_0^{ab} > 0. \tag{331}$$

*Proof.* By (321),

$$\frac{1}{2t}\log\frac{\Sigma_t^{ab}}{\Sigma_0^{ab}} = \frac{1}{2t}\int_0^t \Gamma_{\|}(r)\,dr.$$

Because $m_t \to m_{\infty}$ and $\bar\eta(t) \equiv \eta$, we have

$$\Gamma_{\|}(t) \longrightarrow -2\eta\Big[\lambda + \phi\chi_{\perp}(m_{\infty}) + \phi m_{\infty}\tau_{\beta}(m_{\infty})\Big] + \eta^2\phi v_{\|}(m_{\infty}) = 2\lambda_{\|}(\eta).$$

The Cesàro theorem therefore yields (330).

The proof of (331) is identical, using (322) and the convergence

$$\Gamma_{\perp}(t) \longrightarrow -2\eta\big[\lambda + \phi\chi_{\perp}(m_{\infty})\big] + \eta^2\phi v_{\perp}(m_{\infty}) = 2\lambda_{\perp}(\eta).$$

$\square$

**Theorem J.16** (Explicit stochastic-edge equation in the rank-one logistic model)**.** *Assume Assumption J.12, and consider the single-replica case $q = 1$. Let $\Sigma_t$ and $R_t$ denote the signal and bulk scalar Gram components from Proposition J.10, with initial data $\Sigma_0 \geq 0$, $R_0 \geq 0$, not both zero. Then:*

1. *The limiting scalar tangent Gram satisfies*
$$G_{\star,t}^{(1)} = \Sigma_t + R_t. \tag{332}$$

2. *The top stationary exponent is*
$$\lambda_1(\eta) = \max\big\{\lambda_{\|}(\eta)\mathbf{1}_{\{\Sigma_0>0\}}, \lambda_{\perp}(\eta)\mathbf{1}_{\{R_0>0\}}\big\}, \tag{333}$$
*where the absent sector is ignored if its initial weight is zero.*

*3. If both sector roots in Definition J.13 are positive, then the unique positive stochastic-edge point is*

$$\eta_c = \min\Big\{\eta_{\|,c}\mathbf{1}_{\{\Sigma_0>0\}} + \infty\mathbf{1}_{\{\Sigma_0=0\}}, \eta_{\perp,c}\mathbf{1}_{\{R_0>0\}} + \infty\mathbf{1}_{\{R_0=0\}}\Big\}. \tag{334}$$

*Equivalently, $\eta_c$ is the unique positive root of the scalar edge equation*

$$\max\big\{\lambda_{\|}(\eta)\mathbf{1}_{\{\Sigma_0>0\}}, \lambda_{\perp}(\eta)\mathbf{1}_{\{R_0>0\}}\big\} = 0. \tag{335}$$

*Proof.* The decomposition (332) is the $q=1$ specialization of (318).

Now assume first that $\Sigma_0 > 0$ and $R_0 > 0$. By Lemma J.15,

$$\Sigma_t = \Sigma_0\, e^{2\lambda_{\|}(\eta)t+o(t)}, \qquad R_t = R_0\, e^{2\lambda_{\perp}(\eta)t+o(t)}.$$

Hence

$$G^{(1)}_{\star,t} = \Sigma_0\, e^{2\lambda_{\|}(\eta)t+o(t)} + R_0\, e^{2\lambda_{\perp}(\eta)t+o(t)}.$$

Taking $\frac{1}{2t}\log(\cdot)$ and using the elementary fact that the logarithmic growth rate of a sum of two nonnegative exponentials is the maximum of their growth rates, we obtain

$$\frac{1}{2t}\log\frac{G^{(1)}_{\star,t}}{G^{(1)}_{\star,0}} \longrightarrow \max\{\lambda_{\|}(\eta), \lambda_{\perp}(\eta)\}.$$

By Appendix H, the left-hand side is the top stationary exponent $\lambda_1(\eta)$. This proves (333) when both sectors are present. If one sector has zero initial weight, the same proof applies after deleting the absent term.

For the edge equation, observe that for any sector $\sharp \in \{\|, \perp\}$,

$$\lambda_\sharp(\eta) = \eta\Big(-\kappa_\sharp^\infty + \frac{\eta}{2}\phi\upsilon_\sharp^\infty\Big),$$

where $\kappa_\sharp^\infty > 0$ and $\upsilon_\sharp^\infty > 0$. Therefore $\lambda_\sharp(\eta)$ has exactly two roots on $[0,\infty)$: the trivial root $\eta = 0$ and the unique positive root $\eta_{\sharp,c}$. It is negative on $(0, \eta_{\sharp,c})$ and positive on $(\eta_{\sharp,c}, \infty)$.

Hence the maximum of the active sector exponents is negative for all positive $\eta$ strictly below the smallest active sector root, zero at that smallest root, and positive for all larger $\eta$. This proves the uniqueness of the positive edge point and identifies it as (334), which is equivalent to (335). ☐

**Corollary J.17** (Strict bulk-contraction regime)**.** *Assume Assumption J.12. If*

$$\eta < \eta_{\perp,c}, \tag{336}$$

*then the orthogonal bulk sector is strictly contracting:*

$$\lambda_\perp(\eta) < 0.$$

*In this regime, whenever $\Sigma_0 > 0$, the top exponent is carried by the signal sector:*

$$\lambda_1(\eta) = \lambda_{\|}(\eta) \qquad \textit{if } \Sigma_0 > 0 \textit{ and } R_0 = 0,$$

*and more generally by the larger of $\lambda_{\|}(\eta)$ and the still negative bulk exponent if both sectors are present.*

*Proof.* The inequality $\eta < \eta_{\perp,c}$ is exactly equivalent to $\lambda_\perp(\eta) < 0$ by the definition of $\eta_{\perp,c}$. The remaining statements follow immediately from (333). ☐

*Proof of Proposition 4.2.* Replace the placeholder statement of Proposition 4.2 in the main text by the content of Proposition J.10 together with Corollary J.11. The proof is exactly the one given above. ☐

*Proof of Theorem 4.3.* Replace the placeholder statement of Theorem 4.3 in the main text by the content of Theorem J.16. The explicit critical surface is the scalar equation (335), and the unique positive critical root is (334). ☐

## K. Exact Reductions and Consistency Checks

This appendix verifies that the general replicated tangent theory reduces exactly to the appropriate earlier theories in three benchmark regimes:

1. the zero-replica reduction, which recovers the imported iterate-level $\mathrm{DMFT}$ of Proposition A.9;
2. the affine-linear learner-side regime, where the replicated tangent closure collapses to a linear deterministic-equivalence system and the Poissonian/Brownian discrepancy vanishes exactly;
3. the frozen-minimum regime, where the top stationary exponent reduces to the characteristic Lyapunov exponent of the local stochastic-gradient cocycle.

These reductions are not cosmetic. They are the main internal consistency checks for the theory developed in Appendices F–J.

### K.1. Recovery of the Imported Iterate-Level Effective Theory

The first reduction is the formal $q = 0$ limit, in which no tangent replicas are present.

**Definition K.1** (Zero-replica convention)**.** Throughout this subsection we adopt the following convention for $q = 0$.

1. Any family indexed by $a \in [q]$ or $(a, b) \in [q]^2$ is empty.
2. The replicated state-kernel space $\mathfrak{X}_T^{(0)}$ is the singleton $\{\varnothing\}$.
3. The tangent-sector kernels
$$\mathbf{Q}, \quad \mathbf{S}, \quad \mathbf{M}, \quad \mathbf{Z}$$
are absent, as are the tangent coefficient maps
$$B, \quad J, \quad H, \quad \mathscr{G}_T.$$
4. All statements involving tangent replicas are interpreted as vacuous.

**Proposition K.2** (Formal $q = 0$ reduction of the augmented cavity system)**.** *Under the zero-replica convention of Definition K.1, the augmented cavity expansion of Appendix C reduces exactly to the iterate-only cavity expansion. More precisely:*

1. *The tangent recursion (66), the tangent forcing $\mathcal{H}^{a,(-j)}$, the bilinear operator $\mathcal{B}^{a,(-j)}$, the tangent remainders $\mathcal{E}^{u,a,(-j)}$, and all $u$-indexed backreaction variables are absent.*
2. *The only surviving row-level cavity equation is Proposition C.8, namely the exact iterate recursion*
$$\bar{\Theta}_{j,m+1} = \bar{\Theta}_{j,m} - \eta_m\Big(\mathcal{F}_{j,m}^{(-j)} + \mathcal{A}_{j,m}^{(-j)}[\bar{\Theta}_{j,m}] + \mathcal{E}_{j,m}^{\vartheta,(-j)} + \frac{1}{n}g(\bar{\Theta}_{j,m})\Big).$$

*Proof.* This is immediate by inspection of Definitions C.2 and C.15, Propositions C.8 and C.9, and the zero-replica convention Definition K.1. With no tangent replicas present, every $u$-indexed object disappears identically and the only nonvacuous cavity equation is the iterate recursion. $\square$

**Proposition K.3** (Formal $q = 0$ reduction of the deterministic fixed-point theory)**.** *Under the zero-replica convention of Definition K.1, the replicated Volterra theory of Appendices E and F collapses to the imported iterate-level effective theory of Proposition A.9. In particular:*

1. *the replicated Volterra map $\mathcal{T}_T^{(0)}$ acts on the singleton $\mathfrak{X}_T^{(0)} = \{\varnothing\}$;*
2. *the tangent fixed-point condition becomes vacuous;*

3. *the only remaining nontrivial effective objects are the iterate kernels*

$$C^{\vartheta,\vartheta} \qquad \text{and} \qquad C^{\vartheta,\star},$$

*together with the imported iterate effective process* $(\vartheta_t, \xi_t, \vartheta_\star, w_\star, \varepsilon)$ *from Proposition A.9.*

*Hence the* $q = 0$ *specialization of the present theory is exactly the imported iterate-level* DMFT.

*Proof.* In Appendix E, the unknown kernel tuple is

$$(\mathbf{Q}, \mathbf{S}, \mathbf{M}, \mathbf{Z}).$$

Under Definition K.1, this tuple is empty, so the state space $\mathfrak{X}_T^{(0)}$ is a singleton and the Volterra map is trivial. In Appendix F, the empirical tangent kernels $\mathbf{Q}^{(d)}, \mathbf{S}^{(d)}, \mathbf{M}^{(d)}$ are likewise absent. Thus the only remaining content is the imported iterate-level kernel theory from Proposition A.9. No additional fixed-point constraint survives. □

**Corollary K.4** (Recovery of the imported iterate-level DMFT)**.** *Under the formal* $q = 0$ *convention, the present theory reduces exactly to the iterate-level multi-pass* SGD DMFT *of Fan & Wang (2026).*

*Proof.* Combine Propositions K.2 and K.3. □

### K.2. Linear-Model Reduction

We now show that the nonlinear replicated tangent theory collapses to an exact linear deterministic-equivalence system when the learner-side nonlinearity is affine.

**Assumption K.5** (Affine-linear learner-side regime)**.** In addition to Assumptions A.2, A.3, A.4, and F.1, assume that there exist deterministic matrices

$$A_{\text{lin}} \in \mathbb{R}^{k\times k}, \qquad B_{\text{lin}} \in \mathbb{R}^{k\times k_\star}, \qquad \Gamma \in \mathbb{R}^{k\times k}$$

such that

$$f(\xi, w_\star, \varepsilon) = A_{\text{lin}}(\xi - B_{\text{lin}} w_\star), \qquad g(u) = \Gamma u, \tag{337}$$

for all $(\xi, w_\star, \varepsilon) \in \mathbb{R}^k \times \mathbb{R}^{k_\star} \times \mathbb{R}$ and all $u \in \mathbb{R}^k$.

*Remark* K.6. Under Assumption K.5,

$$D_1 f(\xi, w_\star, \varepsilon) \equiv A_{\text{lin}}, \qquad D_1^2 f(\xi, w_\star, \varepsilon) \equiv 0, \qquad D_1^3 f(\xi, w_\star, \varepsilon) \equiv 0,$$

and

$$Dg(u) \equiv \Gamma.$$

Thus every genuinely nonlinear term in the general theory disappears.

**Proposition K.7** (Linear reduction of the replicated coefficient family)**.** *Under Assumption K.5, the concrete replicated coefficient family of Definition F.6 simplifies to*

$$A_t \equiv A_{\text{eff}} := \Gamma + \phi A_{\text{lin}}, \tag{338}$$

$$B_t^a(\mathfrak{X}) \equiv 0, \tag{339}$$

$$J_t^a(\mathfrak{X}) = \phi\, A_{\text{lin}}\big(S_{t,t}^a - M_t^a B_{\text{lin}}^\top\big) A_{\text{lin}}^\top, \tag{340}$$

$$H_t^{ab}(\mathfrak{X}) = \phi\, A_{\text{lin}} Q_{t,t}^{ab} A_{\text{lin}}^\top. \tag{341}$$

*In particular, the replicated tangent effective system becomes a closed linear Volterra/Lyapunov system.*

*Proof.* The identity (338) follows from (192):

$$A_t = \Gamma + \phi\mathbb{E}[D_1 f(\xi_t, w_\star, \varepsilon)] = \Gamma + \phi A_{\text{lin}}.$$

Since $D_1^2 f \equiv 0$, (193) gives $B_t^a(\mathfrak{X}) \equiv 0$.

For $J_t^a$, use (194) and (337):

$$\begin{aligned} J_t^a(\mathfrak{X}) &= \phi\,\mathbb{E}[(A_{\text{lin}} z_t^a) \otimes A_{\text{lin}}(\xi_t - B_{\text{lin}} w_\star)] \\ &= \phi\, A_{\text{lin}} \mathbb{E}[z_t^a \otimes (\xi_t - B_{\text{lin}} w_\star)]\, A_{\text{lin}}^\top \\ &= \phi\, A_{\text{lin}} \big(S_{t,t}^a - M_t^a B_{\text{lin}}^\top\big) A_{\text{lin}}^\top, \end{aligned}$$

which is (340).

Likewise,

$$H_t^{ab}(\mathfrak{X}) = \phi\,\mathbb{E}\big[(A_{\text{lin}} z_t^a) \otimes (A_{\text{lin}} z_t^b)\big] = \phi\, A_{\text{lin}} Q_{t,t}^{ab} A_{\text{lin}}^\top,$$

which is (341). □

**Theorem K.8** (Exact linear deterministic-equivalence system)**.** *Under Assumption K.5, the replicated tangent fixed-point equations reduce exactly to the linear system*

$$M_t^a = M_0^a - \int_0^t \bar\eta(r)\, A_{\text{eff}} M_r^a\, dr, \tag{342}$$

$$S_{t,s}^a = S_{0,s}^a - \int_0^t \bar\eta(r)\, A_{\text{eff}} S_{r,s}^a\, dr + \int_0^{t\wedge s} \bar\eta(r)^2\, \phi A_{\text{lin}} \big(S_{r,r}^a - M_r^a B_{\text{lin}}^\top\big) A_{\text{lin}}^\top\, dr, \tag{343}$$

$$\begin{aligned} Q_{t,s}^{ab} = Q_0^{ab} &- \int_0^t \bar\eta(r)\, A_{\text{eff}} Q_{r,s}^{ab}\, dr - \int_0^s \bar\eta(r)\, Q_{t,r}^{ab} A_{\text{eff}}^\top\, dr \\ &+ \int_0^{t\wedge s} \bar\eta(r)^2\, \phi A_{\text{lin}} Q_{r,r}^{ab} A_{\text{lin}}^\top\, dr. \end{aligned} \tag{344}$$

*This is a closed linear Volterra system whose solution is unique on every finite horizon. In particular, the tangent sector is entirely determined by the pairwise covariance kernels, exactly as in the linear deterministic-equivalence regime of high-dimensional linear* SGD *(Atanasov et al., 2025).*

*Proof.* Substitute the coefficient simplifications (338)–(341) into the general Volterra equations (172)–(174). Since $B \equiv 0$, all nonlinear mixed-drift terms disappear, and the remaining system is exactly (342)–(344). Uniqueness follows either from the general contraction theorem (Theorem E.11) or directly from linear Volterra theory. □

**Corollary K.9** (Exact coincidence of Poissonian and Brownian theories in the linear regime)**.** *Under Assumption K.5, if the imported iterate kernels of the Poissonian and Brownian reference systems coincide, i.e.*

$$C_{t,s}^{\vartheta,\vartheta,\text{SGD}} = C_{t,s}^{\vartheta,\vartheta,\text{SGF}}, \qquad C_t^{\vartheta,\star,\text{SGD}} = C_t^{\vartheta,\star,\text{SGF}} \qquad \textit{for all } t, s,$$

*then*

$$\mathfrak{X}_{T,\star}^{(q),\text{SGD}} = \mathfrak{X}_{T,\star}^{(q),\text{SGF}}, \qquad \Lambda_{q,t}^{\text{SGD}} = \Lambda_{q,t}^{\text{SGF}} \quad \textit{for all } t \in (0,T]. \tag{345}$$

*Equivalently, the jump-cumulant correction vanishes identically:*

$$\mathcal{J}_{q,t} \equiv 0.$$

*Proof.* Under Assumption K.5, we have $B \equiv 0$ and the coefficient families for the Poissonian and Brownian systems are the same whenever the iterate kernels coincide. Thus the two fixed-point equations are identical and have the same boundary data. By uniqueness of solutions to the linear Volterra system (342)–(344), the two kernel tuples coincide. The equality of the characteristic $q$-volume laws then follows from Appendix G, and the vanishing of the jump correction is precisely Corollary I.13 reinterpreted in the present notation. □

### K.3. Frozen-Minimum Reduction

We finally connect the present theory to the local dynamical-stability picture near a fixed minimum.

**Definition K.10** (Frozen-minimum discrete cocycle)**.** Fix $q = 1$. Let $\theta_{\rm fr} \in \mathbb{R}^{d\times k}$ be a local minimizer of the empirical risk, and let

$$H_{B_m}(\theta_{\rm fr}) := \nabla^2 \widehat{L}_{B_m}(\theta_{\rm fr})$$

be the batch Hessian evaluated at $\theta_{\rm fr}$. Define the local discrete tangent cocycle

$$\Phi_m^{\rm fr} := \prod_{r=0}^{m-1} \Big( I - \eta_r H_{B_r}(\theta_{\rm fr}) \Big), \qquad m \geq 1, \tag{346}$$

and let $u_0 \neq 0$ be a fixed tangent initialization.

**Proposition K.11** (Exact discrete-time reduction of the top characteristic exponent)**.** *Assume $q = 1$, and suppose that after some finite transient $m_{\rm fr}$, the iterate is frozen:*

$$\bar{\Theta}_m \equiv \theta_{\rm fr} \qquad \textit{for all } m \geq m_{\rm fr}.$$

*Then for every $m \geq 1$,*

$$\Lambda_{1,t_m}^{(d)} = \frac{1}{t_m} \log \frac{\|\Phi_m^{\rm fr} u_0\|_2}{\|u_0\|_2}, \qquad t_m := m n^{\alpha-1}. \tag{347}$$

*Consequently, if the per-step characteristic Lyapunov exponent*

$$\lambda_{\rm fr}^{\rm step} := \lim_{m\to\infty} \frac{1}{m} \log \frac{\|\Phi_m^{\rm fr} u_0\|_2}{\|u_0\|_2} \tag{348}$$

*exists, then the epoch-time stationary exponent satisfies*

$$\lambda_1 = n^{1-\alpha} \lambda_{\rm fr}^{\rm step}, \tag{349}$$

*after matching the two time normalizations.*

*Proof.* With $q = 1$, the tangent Gram is scalar and Definition 2.4 reduces to

$$\mathbf{Q}_m^{(d,1)} = \Big[ \frac{1}{dk} \|\bar{U}_m\|_{\rm F}^2 \Big].$$

By Corollary G.10,

$$\Lambda_{1,t_m}^{(d)} = \frac{n^{1-\alpha}}{2m} \log \frac{\|\bar{U}_m\|_{\rm F}^2}{\|U_0\|_{\rm F}^2} = \frac{n^{1-\alpha}}{m} \log \frac{\|\bar{U}_m\|_{\rm F}}{\|U_0\|_{\rm F}}.$$

In the frozen regime,

$$\bar{U}_m = \Phi_m^{\rm fr} u_0,$$

so (347) follows immediately because $t_m = m n^{\alpha-1}$. If (348) exists, then dividing (347) by 1 and passing to the limit yields (349). □

**Corollary K.12** (Consistency with local Lyapunov theory near minima)**.** *In the frozen-minimum regime of Proposition K.11, the top stationary exponent of the present theory coincides with the characteristic Lyapunov exponent of the local stochastic-gradient cocycle after matching the discrete-step and epoch-time normalizations. In particular, this is consistent with the local dynamical-stability analysis of Chemnitz & Engel (2025).*

*Proof.* This is exactly Proposition K.11 together with Proposition H.11 of Appendix H. The former gives the exact discrete-time identity, while the latter identifies the limiting exponent with the stationary $q = 1$ exponent $\lambda_1$. □

*Proof of Corollaries 3.2 and 4.4.* Corollary 3.2 is exactly Corollary K.4.

For Corollary 4.4, combine:

1. the iterate-level recovery of Corollary K.4;
2. the linear deterministic-equivalence reduction of Theorem K.8 and Corollary K.9;
3. the frozen-minimum reduction of Corollary K.12.

These three statements are precisely the exact consistency checks advertised in the main text. □