Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sharp Cyclic Forgetting in Linear Regression

Jiangze Zhu

quifox@163.com

Qingdao University of Science and Technology

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11817786/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Sharp Cyclic Forgetting in Linear Regression

Jiangze Zhu (ORCID: 0009-0004-4839-4886)[1*]

[1*]Qingdao University of Science and Technology, Qingdao, China.

Corresponding author(s). E-mail(s): quifox@163.com;

Abstract

Sequentially fitting jointly realizable linear-regression tasks in a fixed cycle can reintroduce residual error on tasks that were fitted exactly. We characterize the largest such endpoint error under orthogonal-projection updates. For $\boldsymbol{T}$ tasks and $\boldsymbol{m}$ complete cycles, when $\boldsymbol{q}$ is the minimum dimension of a task solution subspace, the worst-case average squared residual is $\boldsymbol{\Theta(\min\{1, T\min(T,q)/m\})}$, uniformly over finite-dimensional task families. This identifies a transition from linear to quadratic dependence on the task count as $\boldsymbol{q}$ increases. We also determine the exact leading constant in the unrestricted iterated asymptotic regime: $\boldsymbol{\lim_{T\to\infty}\lim_{m\to\infty} mM_T(m)/T^2 = 1/(2\pi^4)}$. The corresponding constant for the maximum task residual is $\boldsymbol{1/(8\pi^2)}$. The proof connects the cycle product and endpoint observations through closed paths of subspaces. A sharp dissipative semigroup bound provides the upper constant, while a finite real projection construction matches it. The argument distinguishes endpoint residuals from the norm decrease in the following cycle, a distinction essential for the lower bound. The result gives a worst-case characterization of exact cyclic fitting without assumptions on how task subspaces are distributed.

Keywords: Continual learning, cyclic projections, linear regression, worst-case forgetting, sharp asymptotic bounds

1 Introduction

Fitting a linear regression task can increase the error on an earlier task even when all tasks share a solution. Under exact minimum-change fitting, each task projects the current parameter onto its affine solution space. A repeated pass through the tasks thus gives a product of orthogonal projections. Its learning performance at the end of a pass is measured by evaluating every task at that same parameter.

Evron et al. (2022) proved an $\Omega(T/m)$ worst-case lower bound after m cycles through T tasks in the regime $m \geq T$. Their upper bound includes the term $O(Tq/m)$, where q is the minimum task nullity. Swartworth et al. (2023) established the dimension-independent upper bound $O(T^2/m)$ and asked whether the factor between that bound and the known lower bound could be removed. Kong (2026, Chapter 4, Problem 1) asks both whether the rate is order-sharp and what the exact limiting constant is. The first theorem closes the finite-budget order gap, and the second determines the exact iterated-limit constant in the unrestricted case. Together they settle both parts of this cyclic question.

Evron et al. (2022, Appendix C.2.1) already explored whether a contraction with substantial late-cycle energy loss could be turned into a hard projection instance. General quantitative factorizations of singular strict contractions can require many factors (Oikhberg 1999, Theorem 1). For a lower bound on forgetting, the same T factors must be used both to generate the orbit and to define the residual measured on it.

Our construction matches the nullity-dependent and dimension-independent upper bounds within a single family of examples. For every $T \geq 2$, $m \geq 1$, and $q \geq 1$, the worst-case residual is of order $\min\{1, T\min(T,q)/m\}$. The inverse-linear tail in m therefore has a coefficient of order Tq when $q \leq T$ and T^2 when $q \geq T$. The result is uniform over finite task families with the stated parameters. The family can be chosen for the horizon m and stays fixed throughout its m cycles.

Two features of the hard instances account for this dependence. First, a rotation spreads the initial error over Fourier modes. Its correlation with the contracting direction stays small in squared sum before a prescribed return. Second, the product step is realized as the overlap between adjacent subspaces on a closed path. Adjacent subspaces consume little squared norm, while evaluating the endpoint against the entire path produces a larger residual. We quantify this difference for the same product and observation. A trace bound shows that an orbit confined to one plane has a coefficient at most linear in T in the $1/m$ rate.

The exact-constant proof uses a sharper version of this same path geometry: every finite projection cycle has an exact closed-path factorization on its terminal invariant space. The resulting Dirichlet observation and skew-sector estimates lead directly to a sharp semigroup bound; a finite real absorber construction matches its leading constant. No compactness argument over varying ambient dimensions is needed.

Related work.

Ding et al. (2024) analyze the effects of task similarity, task order, and step size in sequential linear regression. Their analysis concerns a different task-order and learning contract from the fixed-cyclic endpoint minimax criterion considered here.

Beyond the cyclic minimax setting, Lin et al. (2023) analyze expected forgetting and generalization across sequences of overparameterized linear tasks, relating them to task similarity and ordering. Doan et al. (2021) study catastrophic forgetting through neural tangent kernel task overlap and examine projected-gradient methods for reducing it. These analyses address different learning dynamics and evaluation criteria:

neither their expected-loss nor kernel-overlap conclusions substitute for a uniform worst-case endpoint bound for exact cyclic projections.

Reich and Zalas (2023) bound the root mean square of the distances to the subspaces and prove sharpness of the time exponent in a Hilbert-space setting with a nonclosed sum of orthogonal complements. Our question concerns uniform dependence on T and nullity among finite-dimensional instances. The numerical-range analysis of Swartworth et al. (2023) controls $(I-A)A^m$ for the cycle product A. A lower bound on forgetting also needs control of the observation defined by the individual projections. Their real numerical-range boundary construction lies in a four-dimensional invariant block (Swartworth et al. 2023, Supplement, Proposition 18). The trace bound in Section 4 limits the residual on this block to $O(T/m)$. Harmonic families of projections occur in comparisons of deterministic and randomized products (Recht and Ré 2012). Evron et al. (2026) study expected forgetting under random task orderings. Tsipory et al. (2025) analyze geometry-dependent task orderings. Our analysis instead fixes the cyclic schedule and evaluates the loss at the common endpoint of each cycle.

2 Model and main result

Task i has a data matrix $X_i \in \mathbb{R}^{s_i \times D}$ with $\|X_i\| \le 1$, and labels $b_i = X_i w_\star$ for a shared solution $w_\star$. All vector norms are Euclidean and all matrix norms without a subscript are the induced operator norm. Assume $\|w_0 - w_\star\| \le 1$. Each update selects the closest parameter to the current one among those fitting the task exactly. If P_i projects onto $\ker X_i$, the update on the error $w - w_\star$ is multiplication by P_i. Put

$$A = P_T \cdots P_1, \qquad Q_i = I - P_i, \qquad x = w_0 - w_\star.$$

After m passes in the order $1, \dots, T$, the error is $A^m x$. The reported loss is the average of the squared residuals $\|X_i A^m x\|^2$ over tasks, each measured at the cycle endpoint. Each task achieves zero loss at its own most recent fit, so this endpoint loss equals the forgetting relative to that fit.

Since $X_i^\top X_i \preceq Q_i$, the residual is bounded by the squared projection distance. Conversely, for any P_i with nonzero Q_i, a matrix of orthonormal rows spanning $\operatorname{ran}(Q_i)$ gives $X_i^\top X_i = Q_i$. If $Q_i = 0$, one zero row suffices. Thus the worst-case forgetting is precisely

$$\mathcal{F}(T, m, q) = \sup_{\substack{D \ge 1,\ P_i = P_i^\top = P_i^2 \\ \min_i \operatorname{rank}(P_i) = q,\ \|x\| \le 1}} F_m(x), \qquad F_m(x) = \frac{1}{T} \sum_{i=1}^{T} \|Q_i A^m x\|^2 . \tag{1}$$

The ambient dimension D is free. The corresponding number of individual task fits is $k = mT$.

Theorem 1 (Sharp cyclic forgetting)**.** *There are absolute constants $c, C > 0$ such that for all integers $T \geq 2$, $m \geq 1$, and $q \geq 1$,*

$$c \min\left\{1, \frac{T \min(T,q)}{m}\right\} \leq \mathcal{F}(T,m,q) \leq C \min\left\{1, \frac{T \min(T,q)}{m}\right\}. \tag{2}$$

For the lower bound there are real instances with every task nullity equal to q, ambient dimension at most $3q$, and $\bigcap_i \operatorname{ran}(P_i) = \{0\}$. For these instances the next cycle loses squared norm

$$\Delta_m(x) = \|A^m x\|^2 - \left\|A^{m+1} x\right\|^2$$

of order $F_m(x)/T$, with absolute comparison constants.

Taking the supremum over q yields $\Theta(\min\{1, T^2/m\})$, equivalently $\Theta(\min\{1, T^3/k\})$ at a complete cycle. For $T = 1$ or $q = 0$, forgetting is zero after one cycle.

3 Exact unrestricted asymptotic constants

Let $M_T(m) = \sup_{q \geq 1} \mathcal{F}(T,m,q)$, and let $N_T(m)$ be the same unrestricted supremum with the average replaced by the maximum over tasks. The task family may depend on m, but remains fixed throughout all its cycles. Ambient dimensions are finite and may vary.

Theorem 2 (The exact constants in Kong's cyclic problem)**.** *For every fixed $T \geq 2$, both limits*

$$c_T^{\mathrm{av}} = \lim_{m \to \infty} m M_T(m), \qquad c_T^{\mathrm{mx}} = \lim_{m \to \infty} m N_T(m)$$

exist. Their large-task limits are

$$\lim_{T \to \infty} \frac{c_T^{\mathrm{av}}}{T^2} = \frac{1}{2\pi^4}, \qquad \lim_{T \to \infty} \frac{c_T^{\mathrm{mx}}}{T^2} = \frac{1}{8\pi^2}. \tag{3}$$

In particular,

$$2\pi^4 \lim_{T \to \infty} \lim_{m \to \infty} \frac{m}{T^2} \sup_{q \geq 1} \mathcal{F}(T,m,q) = 1.$$

The constants are unchanged if the common fixed space is required to be zero. These are iterated limits with m taken first.

The main upper bound comes directly from a *finite-cycle* identity, rather than a limiting approximation of the subspaces. Let $W = \operatorname{ran} A$, $A_W = A|_W$, and let B_W be the compressed endpoint observation. Appendix A proves that the deviations $x - P_i \cdots P_1 x$ form an exact closed-path factorization of $I_W - A_W$. The discrete Dirichlet and skew-spectrum inequalities then give

$$B_W \preceq 2\kappa_T D_W, \qquad -\cot(\pi/T) D_W \preceq C_W \preceq \cot(\pi/T) D_W, \tag{4}$$

where D_W and C_W are the Hermitian and skew-Hermitian parts of $I_W - A_W$, respectively, and $\kappa_T = [4T\sin^2(\pi/(2T))]^{-1}$. For the observation of any single task, the same argument yields $B_{W,i} \preceq 2\gamma_T D_W$, where $\gamma_T = \lfloor T^2/4 \rfloor / T$. These are finite-product inequalities, with the displayed sharp Dirichlet and skew-path constants. Their simultaneous optimality for the restricted class of genuine projection products is not asserted here.

Theorem 3 (Sharp sectorial dissipation)**.** *For finite complex matrices $H = D + iC$ with $D > 0$ and $C = C^*$, put*

$$\Gamma(s) = \sup_{\left\|D^{-1/2}CD^{-1/2}\right\| \le s} \left\| D^{1/2} e^{-H} \right\|^2 .$$

The supremum includes arbitrary finite dimensions. Then

$$\Gamma(s) = \frac{s}{4\pi} + o(s), \qquad \Gamma(s) \le \frac{s}{4\pi} + O(s^{2/3}) \quad (s \to \infty).$$

The same asymptotic value is attained by real $D > 0$ and real skew-symmetric $H - D$.

The upper bound is established by a weighted two-ray resolvent identity and a vanishing scalar regularization in Appendix C. A finite real absorber model attains the matching coefficient, as proved in Appendix B.

Proof of Theorem 2 The existence of the fixed-T inner limits follows from the short exact-path graph variational theorem in Appendix A. For the sharp upper bound, fix any projection family and put $n = m - 1$. Since $A^m x = A_W^n(Ax)$ and $\|Ax\| \le \|x\|$, we may work entirely on W. The uniform Poissonization estimate of Lemma 7 states that $\left\| A_W^n - e^{-n(I_W - A_W)} \right\| = O_T(n^{-1})$. Apply (4) to $H = n(I_W - A_W)$, $D = nD_W$; the nullspace of D_W can be removed exactly. Indeed, $2D_W = I - A_W^* A_W + (I - A_W)^*(I - A_W)$; hence $D_W v = 0$ implies $A_W v = A_W^* v = v$. Equality of the norms through the original cycle forces $P_i v = v$ for every i, so $B_W v = 0$. Thus $\ker D_W$ is a reducing fixed subspace with zero observation; on its orthogonal complement $D_W > 0$. Since $B_W \preceq 2\kappa_T D_W$,

$$\sqrt{n} \left\| B_W^{1/2} e^{-n(I_W - A_W)} \right\| \le \sqrt{2\kappa_T \, \Gamma(\cot(\pi/T))}.$$

The Poissonization error becomes $O_T(n^{-1/2})$ after multiplying by $\sqrt{n} B_W^{1/2}$, uniformly in dimension and in the chosen projection family. Hence

$$c_T^{\mathrm{av}} \le 2\kappa_T \Gamma(\cot(\pi/T)), \qquad c_T^{\mathrm{mx}} \le 2\gamma_T \Gamma(\cot(\pi/T)).$$

Using $\kappa_T \sim T/\pi^2$, $\gamma_T \sim T/4$, and $\Gamma(\cot(\pi/T)) \sim T/(4\pi^2)$ proves the matching upper coefficients in (3). The graph realization and the finite real absorber of Appendix B provide the corresponding lower coefficients with zero common fixed space. This proves the assertions, including the stated order of limits. □

4 Upper bounds and effective nullity

Let $y = A^m x$. Along the next pass write $y_0 = y$ and $y_i = P_i y_{i-1}$. The sum of the squared steps is $\sum_i \|y_i - y_{i-1}\|^2 = \|y\|^2 - \|Ay\|^2$. The nearest-point property of P_i

and Cauchy–Schwarz give

$$\|Q_i y\|^2 \le \|y - y_i\|^2 \le i \sum_{\ell=1}^{i} \|y_\ell - y_{\ell-1}\|^2 .$$

At a cycle endpoint $y \in \operatorname{ran}(P_T)$, so the last term of the task average vanishes. Summing over $i < T$ gives the energy bound

$$F_m(x) \le \frac{T-1}{2}\, \Delta_m(x). \tag{5}$$

This is the one-cycle argument of Evron et al. (2022, Appendix C.2.1). Summing (5) over m consecutive endpoints gives $\min_{1\le j\le m} F_j(x) \le (T-1)/(2m)$, since the squared norm lost over the corresponding cycles is at most one.

To see which dimensions enter their trace bound (Evron et al. 2022, Lemma 23), let $M = \bigcap_i \operatorname{ran}(P_i)$ and $x_\perp = x - P_M x$, where P_M projects onto M. Each projector fixes M and preserves $M^\perp$. The component in M contributes zero to F_m and Δ_m. Let $V_x = \operatorname{span}\{Ax_\perp, A^2 x_\perp, \ldots\}$, an A-invariant subspace of dimension $d_x \le q - \dim M$, and write $L = A|_{V_x}$. For $j \ge 0$ the positive semidefinite matrices

$$B_j = (L^j)^\top L^j - (L^{j+1})^\top L^{j+1}$$

satisfy $B_{j+1} = L^\top B_j L$. Since L is a contraction, $\operatorname{tr}(B_j)$ is nonincreasing. Telescoping the first m traces gives $m \operatorname{tr}(B_{m-1}) \le d_x$. With $u = Ax_\perp$, we have $\Delta_m(x) = u^\top B_{m-1} u \le d_x/m$ (the zero-dimensional case has zero residual). In particular,

$$F_m(x) \le \frac{(T-1)d_x}{2m} \le \frac{(T-1)(q - \dim M)}{2m}. \tag{6}$$

Combining this with $F_m(x) \le 1$ and the $O(T^2/m)$ upper bound of Swartworth et al. (2023, Theorem 1) proves the upper bound in Theorem 1.

The active dimension bound has two immediate consequences. If all the nontrivial cycle iterates stay in a plane, the residual is $O(T/m)$. More generally, to attain residual of order $T\min(T,q)/m$ in the inverse-linear regime, the orbit requires d_x of order at least $\min(T,q)$. Normal dynamics on V_x also give at most $O(T/m)$ in any dimension. In that case, L commutes with $L^\top$, and the spectral theorem gives $\Delta_m(x) \le \max_{0\le s\le 1}(1-s)s^{m-1} \le 1/m$ using $\|u\| \le 1$. The hard instances must therefore combine enough active directions with a cycle product that is nonnormal on their orbit when the coefficient grows faster than T.

The common solution directions can be removed exactly. If one fixes $\dim M = h \le q$, the supremum in (1) becomes $\mathcal{F}(T, m, q-h)$. Restriction to $M^\perp$ proves the upper direction. For the reverse, take any instance of minimum nullity $q - h$. Remove its common intersection, then restore any lost nullity with the zero-intersection padding in Section 7. This preserves its residual and gives an instance with zero common intersection. Adjoining a common h-dimensional subspace realizes the original residual at the prescribed q and h.

5 A delayed return in a finite product

For the main construction take integers $n \geq 1$, $T \geq 8n$, and $m \geq nT/8$. Write $N = mT$ and $d = 2n+1$. On $\mathbb{R}^d$, let U fix the first coordinate and rotate the j-th two-dimensional plane by $2\pi j/(16N)$ for $1 \leq j \leq n$. Let v have first coordinate $1/\sqrt{2n+1}$ and coordinates $(\sqrt{2/(2n+1)}, 0)$ in each plane. Set

$$z = U^{-N}v, \qquad \beta = \frac{n}{32N}, \qquad \eta = e^{-1/(2N)}, \qquad G = \eta U(I - \beta vv^\top). \tag{7}$$

The vectors z, v are unit, and $\|G\| \leq \eta < 1$. Without the rank-one damping the rotation returns z to v after N steps. The next lemma shows that the return survives the damping.

Lemma 4 (Return at step N)**.** *The matrix and vectors in* (7) *satisfy* $v^\top G^N z > 2/5$.

Proof Define $w_j = \eta^{-j}U^{-j}G^j z$ and $H_j = U^{-j}vv^\top U^j$. Then $w_0 = z$ and $w_{j+1} = (I - \beta H_j)w_j$. Put

$$b = \beta \sum_{j=0}^{N-1} \left\|H_j z\right\|^2, \quad c = \beta \sum_{j=0}^{N-1} \left\|H_j w_j\right\|^2, \quad s = \beta \sum_{j=0}^{N-1} (H_j z)^\top H_j w_j.$$

The energy identity and the recurrence give $\|w_N\|^2 = 1 - (2-\beta)c$ and $z^\top w_N = 1 - s$. Thus $\|w_N - z\|^2 \geq 0$ implies $s \geq (1-\beta/2)c$. Together with $s^2 \leq bc$, this yields $0 \leq s \leq b/(1-\beta/2)$.

For $\phi(k) = v^\top U^k v$, we have

$$\phi(k) = \frac{1}{2n+1} \sum_{j=-n}^{n} e^{2\pi \mathrm{i} jk/(16N)}, \qquad b = \beta \sum_{k=1}^{N} |\phi(k)|^2.$$

All $2n+1$ Fourier modes are distinct modulo $16N$. Their orthogonality gives $\sum_{k=0}^{16N-1} |\phi(k)|^2 = 16N/(2n+1)$. The terms at $k = 1, \ldots, N$ and at $16N - k$ are disjoint and equal in pairs, so $b \leq n/[4(2n+1)] < 1/8$. Since $N \geq 8n$, we have $\beta \leq 1/256$ and $s < 1/6$. We conclude that $v^\top G^N z = \eta^N z^\top w_N > 5/(6\sqrt{e}) > 2/5$. □

The symmetric step defect $B = I - (G + G^\top)/2$ both ensures that the subspace construction is feasible and retains the contracting direction. In fact,

$$B = \tfrac{1}{2}\left(I - G^\top G + (I-G)^\top(I-G)\right) \succeq \tfrac{3\beta}{4} vv^\top, \qquad B \succeq (1-\eta)I \succeq \frac{1}{3N} I. \tag{8}$$

The first inequality follows from $I - G^\top G = (1-\eta^2)I + \eta^2(2\beta - \beta^2)vv^\top$ and $\eta^2 \geq 7/8$, $\beta \leq 1/256$. The second uses $\|G\| \leq \eta$ and $1 - e^{-1/(2N)} \geq 1/(3N)$.

6 An exact cycle of projections

Lemma 5 (A closed path of subspaces)**.** *Let $0 \prec E \preceq I/2$, $W^\top = -W$, and $S = E^{-1/2} W E^{-1/2}$ with $\|S\| \le 1$. The real maps*

$$J(\theta) = \begin{pmatrix} (I-E)^{1/2} \\ (\cos\theta\, I + \sin\theta\, S) E^{1/2} \\ -\sin\theta\, (I + S^2)^{1/2} E^{1/2} \end{pmatrix} \tag{9}$$

are isometries from $\mathbb{R}^d$ to $\mathbb{R}^{3d}$, and

$$J(\theta)^\top J(\varphi) = I - (1 - \cos(\varphi - \theta))E + \sin(\varphi - \theta) W. \tag{10}$$

Let $T \ge 3$, $\delta = 2\pi/T$, $J_i = J(i\delta)$, and $P_i = J_i J_i^\top$ for $1 \le i \le T$. The path closes with $J_T = J_0$. At a point $J_0 y$ the averaged squared residual is $y^\top R_0 y$, where

$$R_0 = \frac{1}{T} \sum_{i=1}^{T} J_0^\top (I - P_i) J_0 = 2E - \tfrac{3}{2} E^2 + \tfrac{1}{2} W^2, \qquad E \preceq R_0 \preceq 2E. \tag{11}$$

Proof The square root in (9) exists since S is skew symmetric and $\|S\| \le 1$. Multiplication of the three blocks gives (10), which for $\theta = \varphi$ is the isometry identity. For the residual use $J_i^\top J_0 = I - (1 - \cos(i\delta))E - \sin(i\delta) W$. Averaging its squared Gram matrix, the sine, cosine and sine–cosine terms have zero mean, and the squares of sine and cosine have mean $1/2$. The signs in (11) follow from $W^\top = -W$. For the lower bound, $E^2 \preceq \|E\|\, E$ and $-W^2 = E^{1/2} S^\top E S E^{1/2} \preceq \|E\|\, E$, so that $R_0 \succeq 2(1 - \|E\|) E \succeq E$. The same identities give the upper bound $R_0 \preceq 2E$. □

To realize (7), define $a = 1 - \cos\delta$, $b = \sin\delta$, and $K = (G - G^\top)/2$, and set

$$E = B/a, \qquad W = -K/b. \tag{12}$$

The angle bounds $16/T^2 \le a \le 2\pi^2/T^2$ and $a/b = \tan(\pi/T) \le 4/T$ hold for $T \ge 8$. Let $\theta_{\max} = \pi n/(8N)$ be the largest angle of U. Taking the symmetric and skew parts of (7) gives

$$\|B\| \le (1 - \eta) + \theta_{\max}^2/2 + \beta \le \frac{9n}{16N}, \qquad \|K\| \le \theta_{\max} + \beta \le \frac{17n}{32N}.$$

Here $1 - \eta \le 1/(2N)$, $\theta_{\max} \le n/(2N)$, and $n/N \le 1/8$. Thus $\|E\| \le 9nT/(256m) \le 9/32$, and (8) yields

$$\left\| E^{-1/2} W E^{-1/2} \right\| \le \frac{a}{b}\, 3N\, \|K\| \le \frac{51}{64}.$$

Lemma 5 now gives $J_i^\top J_{i-1} = I - aE - bW = G$. A projection of any $J_{i-1}y$ onto the next subspace is J_iGy. Hence for $x = J_0z$ the endpoint satisfies the exact identity

$$A^m x = J_0 G^N z. \tag{13}$$

Writing $y = G^N z$ and using (11), (8) and Lemma 4, we get

$$F_m(x) = y^\top R_0 y \geq \frac{3\beta}{4a}(v^\top y)^2 \geq \frac{3nT}{1600\pi^2 m}. \tag{14}$$

Also $F_m(x) \leq 2\|E\| \leq 9nT/(128m)$.

6.1 Residual and energy loss for the same cycle

For the construction above write $C = aE + bW$, so $G = I - C$. The single-step energy defect is

$$I - G^\top G = 2aE - C^\top C, \qquad C^\top C \preceq \|E\|\,(a^2 + b^2)E = 2a\,\|E\|\,E.$$

The inequality follows by writing $C = E^{1/2}(aI + bS)E^{1/2}$ and using skew symmetry of S. Thus the defect is bounded above and below by $2aE$ and aE, respectively. Moreover

$$\left\| E^{1/2} G E^{-1/2} - I \right\| \leq \|E\|\sqrt{2a} \leq \sin(\pi/T) \leq \pi/T.$$

For $0 \leq j \leq T$, the singular values of the j-th power of $E^{1/2}GE^{-1/2}$ lie between $(1 - \pi/T)^j$ and $(1 + \pi/T)^j$. For $T \geq 8$ these bounds lie between absolute positive constants. In particular $\left\|E^{1/2}G^j y\right\|^2$ is within absolute factors of $y^\top Ey$. By telescoping the step losses,

$$\|y\|^2 - \left\|G^T y\right\|^2 = \sum_{j=0}^{T-1} (G^j y)^\top (I - G^\top G) G^j y \asymp Ta\, y^\top Ey \asymp \frac{y^\top Ey}{T}.$$

Here and below $\asymp$ denotes comparability with absolute positive constants. The actual residual of the same cycle is $y^\top R_0 y$, which is comparable to $y^\top Ey$. Thus the instances in (14) simultaneously have $F_m(x) \asymp nT/m$ and $\Delta_m(x) \asymp n/m$. The Fourier modes supply the factor n in the late-cycle loss of squared norm. The geometry of the closed path supplies the factor T between that loss and the observed residual.

7 All parameter regimes

Suppose $T \geq 8$, $q \geq 3$, and $m \geq T/8$. Choose

$$n = \min\left\{\lfloor T/8 \rfloor,\ \lfloor (q-1)/2 \rfloor,\ \lfloor 8m/T \rfloor\right\}.$$

Then $n \geq 1$, $T \geq 8n$, $2n + 1 \leq q$, and $m \geq nT/8$. The choice also gives $n \geq \min(T, q, m/T)/16$, since $\lfloor s \rfloor \geq s/2$ for $s \geq 1$. The lower bound (14) therefore has the

scale in (2). In the long-horizon regime $m \geq T \min(T, q)$, the active orbit has dimension at most $2n + 1 = O(\min(T, q))$. By (6) and the attained residual, it has dimension of order $\min(T, q)$. The common intersection is zero. Any shared vector would be $J_0 y$ and fixed by A, so that $G^T y = y$. Strict contractivity of G then gives $y = 0$.

A planar family of the type used by Evron et al. (2022, Appendix C.2.2) handles the remaining regimes. For any $T \geq 2$ and $m \geq 1$, let

$$\alpha^2 = \tfrac{1}{16} \min\{1, T/m\}, \quad \theta_i = \alpha \sin(\pi i/T), \quad u_i = (\cos\theta_i, \sin\theta_i)^\top, \quad P_i = u_i u_i^\top.$$

Take $x = u_0 = u_T$ and $\rho = \prod_{i=1}^{T} \cos(\theta_i - \theta_{i-1})$. Then $A^m x = \rho^m x$. The elementary sums

$$\frac{1}{T} \sum_{i=1}^{T} \sin^2 \theta_i \geq \frac{2\alpha^2}{\pi^2}, \qquad \sum_{i=1}^{T} (\theta_i - \theta_{i-1})^2 = 2T\alpha^2 \sin^2\Big(\frac{\pi}{2T}\Big) \leq \frac{\pi^2 \alpha^2}{2T}$$

follow from sine orthogonality on the grid and $|\sin t| \geq 2|t|/\pi$ for $|t| \leq \pi/2$. Using $\log\cos t \geq -t^2$ for $|t| \leq \pi/4$, we get $\rho^{2m} \geq e^{-\pi^2 m \alpha^2/T} \geq e^{-1}$. Consequently $F_m(x) \geq (8e\pi^2)^{-1} \min\{1, T/m\}$. Also

$$\Delta_m(x) = \rho^{2m}(1 - \rho^2) \leq \rho^{2m} \sum_i (\theta_i - \theta_{i-1})^2 \leq \frac{\pi^4}{4T} F_m(x).$$

Together with (5), this gives $F_m(x) \asymp T\Delta_m(x)$ in the planar example as well. When $T \leq 7$, $q \leq 2$, or $m < T/8$, the target rate in (2) is at most, respectively, 7, 2, or 1 times $\min\{1, T/m\}$. At least two lines differ, so their common intersection is zero.

Finally let q_0 be the attained rank, either $2n + 1$ or 1. To impose rank q when $q_0 < q$, adjoin two mutually orthogonal spaces of dimension $q - q_0$. In the auxiliary space, project onto the first in task 1 and onto the second in task 2. Use the first for any remaining tasks and extend x by zero. The orbit, residual, and energy loss on the active block are unchanged, the common intersection remains zero, and the new ambient dimension is at most $3q$. This completes the proof of Theorem 1.

8 Regression realizations and scope

For each hard instance, choose X_i with orthonormal rows spanning $\mathrm{ran}(Q_i)$. With $w_0 = 0$, $w_\star = -x$, and $b_i = X_i w_\star$, the task labels are compatible and the initial error has unit norm. For the objective $\frac{1}{2} \|X_i w - b_i\|^2$, a single full-gradient step of size one maps the error to $P_i(w - w_\star)$. Every update in the construction is therefore an exact fit by a single gradient step under this choice of data and objective.

Figure 1 shows the trajectory of one fixed family constructed for a target cycle m. The squared alignment peaks near that cycle, accompanied by a local rise in the residual and the squared norm lost in the next cycle. Within each fixed finite-dimensional family, the restriction of A to $M^\perp$ has norm less than one whenever $M^\perp$ is nonzero. Indeed, a unit vector preserving its norm over a cycle must preserve it at every projection, and hence belongs to M. The residual thus eventually decays geometrically

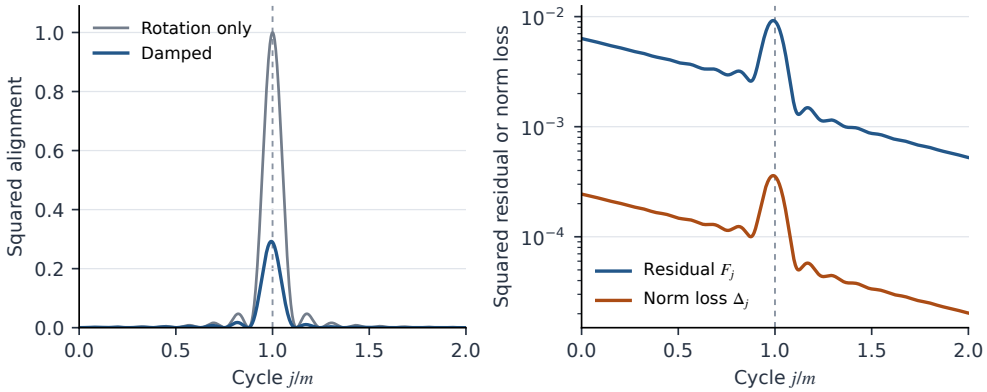


Fig. 1 A family chosen for $n = 64$, $T = 512$, and $m = 4096$. The left panel shows squared alignment with v after j cycles, for the damped step G and for the pure rotation U. The right panel shows the averaged task residual $F_j(x)$ and the next-cycle loss of squared norm $\Delta_j(x)$ on a logarithmic scale. Both panels use the same cycle index j, and the dashed line marks the prescribed horizon. The vector graphic was generated with Python/Matplotlib.

for fixed tasks. The rate in Theorem 1 concerns the supremum at each budget over families with the stated parameters.

The parameter q controls the dimensions of the spaces onto which the updates project. Fixing the row rank separately constrains the orthogonal complements as well. In particular, one row per task forces each nontrivial subspace to be a hyperplane, a distinct case in the questions of Kong (2026, Chapter 4, Problem 2). The one-row Kaczmarz setting instead restricts every task projection to a hyperplane, whereas our theorem permits general block projections. Theorem 1 allows general blocks and describes the worst case with free ambient dimension. The separate hyperplane-only extremal problem is not claimed to be solved by the exact unrestricted constant.

For unit-norm examples, another common loss is the mean squared error within each task. On the same orthonormal-row data, averaging over the rows divides the task residual by the number of rows. Appendix D shows that if the normalization constrains each row norm to be at most one, the worst case for that loss still equals $\mathcal{F}(T, m, q)$. The construction changes the covariance within each row space while keeping the projection trajectory fixed.

9 Discussion and conclusion

The matching bounds identify the cost of revisiting jointly realizable tasks when each update exactly fits the current task but the final error is evaluated simultaneously on all tasks. In particular, the quadratic task-count dependence of the dimension-independent upper bound is attained in the worst case once the minimum solution-space nullity is large enough. When that nullity is smaller, the sharp rate has the finer dependence Tq/m. Thus the known $O(T^2/m)$ guarantee cannot be improved uniformly over all admissible task families merely by exploiting joint realizability and exact fitting.

The minimax quantifiers are essential to this interpretation. For any *fixed* finite-dimensional family, the cycle product is strictly contractive away from its common fixed space, and the endpoint residual eventually decays geometrically. The difficult families in the finite-budget lower bound may depend on the prescribed horizon. The exact leading constants likewise concern the iterated limit in which m tends to infinity for fixed T, followed by T tending to infinity; they do not assert the same constants under arbitrary joint growth. The results characterize exact cyclic minimum-change fitting, not expected performance on randomly sampled tasks or arbitrary finite-step and nonlinear training procedures.

A natural next question is the sharp universal behavior when each task contains only one data row, forcing hyperplane projections rather than general block projections. Other changes to the learning contract, such as noncyclic orderings, inconsistent tasks, or inexact updates, require separate analysis. The present results specify a worst-case baseline against which such additional assumptions or algorithmic changes can be assessed.

A An exact path factorization and the graph limit

The main observation is that the closed-path model is already present *exactly* inside every finite product of orthogonal projections. Thus neither a limit of subspaces nor a resolvent compactness argument is needed to identify the asymptotic minimax value.

A.1 Exact path factorization

Let $P_1, \dots, P_T$ be orthogonal projections on a finite-dimensional real Hilbert space $\mathcal{H}$, let $A = P_T \cdots P_1$, and let $W = \operatorname{ran} A \subseteq \operatorname{ran} P_T$. On W let $A_W = A|_W$ and

$$B_W = \frac{1}{T} P_W \Big(\sum_{i=1}^{T} (I - P_i) \Big) P_W \Big|_W .$$

Let S be the lower shift on $\mathbb{R}^{T-1}$ and put $L_T = 2I - S - S^*$, $C_T = S^* - S$. Thus $\lambda_1(L_T) = 4\sin^2(\pi/(2T))$ and

$$\left\| L_T^{-1/2} C_T L_T^{-1/2} \right\| = \cot(\pi/T). \tag{15}$$

The same formulas hold after complexification.

Lemma 6 (Exact closed-path factorization)**.** *For $x \in W$, set $y_i(x) = P_i \cdots P_1 x$ and $U_i x = x - y_i(x)$ for $1 \le i < T$; put $U_0 = U_T = 0$. Let $U : W \to \mathcal{H}^{T-1}$ stack the maps U_i. Then*

$$I_W - A_W = U^*((I - S) \otimes I_{\mathcal{H}})U, \qquad B_W \preceq T^{-1} U^* U. \tag{16}$$

Consequently, writing $H = I_W - A_W$, $D = (H + H^*)/2$, *and* $C = (H - H^*)/(2i)$, *we have the* exact finite-cycle *estimates*

$$B_W \preceq 2\kappa_T D, \qquad -\cot(\pi/T)D \preceq C \preceq \cot(\pi/T)D, \qquad \kappa_T = \frac{1}{4T\sin^2(\pi/(2T))}. \tag{17}$$

Proof Fix $x, z \in W$. Write $d_i(x) = y_{i-1}(x) - y_i(x)$, where $y_0(x) = x$. Then $d_i(x) \perp \operatorname{ran} P_i$, and $z \perp d_T(x)$ because $z \in W \subseteq \operatorname{ran} P_T$. For $i < T$, $y_i(z) \in \operatorname{ran} P_i$, so

$$\langle z, d_i(x)\rangle = \langle z - y_i(z), d_i(x)\rangle = \langle U_i z, (U_i - U_{i-1})x\rangle.$$

Sum over i to obtain the first identity in (16) by polarization. Moreover, $(I - P_i)x = (I - P_i)U_i x$ for $i < T$, while $(I - P_T)x = 0$. Squaring and averaging proves the observation bound.

Now $D = \frac{1}{2}U^*(L_T \otimes I)U$ and $C = \frac{1}{2i}U^*(C_T \otimes I)U$. The discrete Dirichlet inequality $U^*U \preceq \lambda_1(L_T)^{-1}U^*(L_T \otimes I)U$ proves the first estimate of (17); (15) gives the second. □

For completeness, the eigenvalues behind (15) can be obtained from $(S^* - S)f = i\mu(2I - S - S^*)f$ with $f_0 = f_T = 0$: its characteristic roots are 1 and $(-1+i\mu)/(1+i\mu)$, and the boundary condition gives $\mu = \cot(k\pi/T)$ for $1 \le k < T$.

A.2 One uniform discrete-to-continuous estimate

Lemma 7 (Uniform Poissonization)**.** *For each fixed* T *there is* $K_T < \infty$ *such that all products* $A = P_T \cdots P_1$ *of* T *orthogonal projections satisfy*

$$\|(I - A)A^j\| \le \frac{K_T}{j+1}, \qquad \|A^n - e^{-n(I-A)}\| \le \frac{K_T}{n}, \quad n \ge 1. \tag{18}$$

The estimates are independent of the ambient dimension.

Proof Put $E = I - A$ and $D_E = (E + E^*)/2$. Along one cycle the orthogonal increments d_i satisfy $\sum_i \|d_i\|^2 = \|x\|^2 - \|Ax\|^2$ and $Ex = \sum_i d_i$. Cauchy–Schwarz and the identity $2D_E = I - A^*A + E^*E$ imply

$$D_E \preceq \tfrac{T+1}{2}(I - A^*A).$$

The same argument applied to the reversed product A^* gives $D_E \preceq \frac{T+1}{2}(I - AA^*)$. Since $\langle x, Ex\rangle = \sum_{j \le i}\langle d_j, d_i\rangle$, its skew part is bounded in quadratic form by $(T-1)D_E$. Thus

$$|\langle y, Ex\rangle| \le \sqrt{1 + (T-1)^2}\,\|D_E^{1/2}y\|\,\|D_E^{1/2}x\|.$$

The telescoping forward and adjoint energy sums are bounded by $(T+1)/2$ times the initial squared norm. Pairing

$$EA^j = \frac{1}{j+1}\sum_{\ell=0}^{j} A^{j-\ell}EA^{\ell}$$

with unit vectors and using Cauchy–Schwarz gives the first bound. It also implies $\|E^2A^j\| \le C_T(j+1)^{-2}$: write $E^2A^j = (EA^{\lfloor j/2\rfloor})(EA^{j-\lfloor j/2\rfloor})$ and use commutation.

Here is the uniform discrete remainder explicitly. For integers $k \geq n/2$, telescoping the difference $A^k - A^n$ once and then telescoping each first difference once more gives

$$A^k = A^n + (k-n)A^n(A-I) + R_{k,n}, \qquad \|R_{k,n}\| \leq C_T \frac{(k-n)^2}{n^2}.$$

For $k \geq n$, every summand in $R_{k,n}$ is a positive integer multiple of some $A^j(A-I)^2$ with $j \geq n$; for $n/2 \leq k < n$, the same is true with $k \leq j < n$. The preceding $O_T((j+1)^{-2})$ bound proves both cases, including $k = n$ (where the remainder is zero). For $k < n/2$ the simpler contraction bound $\|R_{k,n}\| \leq 2 + K_T|k-n|/n$ suffices.

Now let N_n be Poisson with mean n. Since $e^{-n(I-A)} = \mathbb{E}[A^{N_n}]$ and $\mathbb{E}(N_n - n) = 0$, averaging the expansion eliminates its first-order term. The second-order contribution is at most $C_T\mathbb{E}(N_n - n)^2/n^2 = C_T/n$. The event $N_n < n/2$ has exponentially small probability (by a standard Chernoff bound), and its remainder is uniformly bounded by $2 + K_T$. This proves the claimed C_T/n bound without dependence on dimension or on the projections. □

A.3 The full minimax variational identity

For finite real Hilbert spaces E, F let $X_0 = X_T = 0$ and $X_i : E \to F$ for $1 \leq i < T$. Define

$$H_X = \sum_{i=1}^{T} X_i^*(X_i - X_{i-1}), \qquad B_X = \frac{1}{T}\sum_{i=1}^{T-1} X_i^* X_i. \tag{19}$$

Let

$$c_{T,q}^{\mathrm{gr}} = \sup_{\dim E \leq q,\, \dim F < \infty,\, X} \|B_X^{1/2} e^{-H_X}\|^2,$$

with no restriction on $\dim E$ when $q = \infty$. In the statement below set $\mathcal{F}(T, m, \infty) := M_T(m)$, the unrestricted minimax loss. For fixed $\dim E = r$, one may take $\dim F \leq (T-1)r$ by factoring the joint Gram matrix of the X_i.

Theorem 8 (Graph variational principle)**.** *For every fixed $T \geq 2$ and $q \in \mathbb{N} \cup \{\infty\}$,*

$$\lim_{m\to\infty} m\mathcal{F}(T, m, q) = c_{T,q}^{\mathrm{gr}}.$$

The analogous maximum-task limit exists and equals $\sup_X \max_{i<T} \|X_i e^{-H_X}\|^2$ under the same dimension restriction. These assertions allow arbitrary finite ambient dimensions and families depending on m.

Proof For the upper bound fix any projection family and $m \geq 2$. Let $W = \operatorname{ran} A$ (so $\dim W \leq q$) and put $n = m - 1$. Apply Lemma 6 on W and set $X_i = \sqrt{n}\, U_i$ for $i < T$, with $X_0 = X_T = 0$. The exact factorization yields, *without passing to a limit*,

$$H_X = n(I_W - A_W), \qquad B_X \succeq nB_W. \tag{20}$$

Also $\|U_i\| \leq 2$ and therefore $\|B_X\| \leq 4n$. For $\|x\| \leq 1$, set $y = Ax \in W$, so $\|y\| \leq 1$ and $A^m x = A_W^n y$. Lemma 7, restricted to the invariant space W, now gives the uniform error

$$\|B_X^{1/2}(A_W^n - e^{-H_X})\| \leq \frac{2K_T}{\sqrt{n}}.$$

Consequently

$$\sqrt{mF_m(x)} \le \sqrt{\frac{n+1}{n}}\left(\|B_X^{1/2}e^{-H_X}\| + \frac{2K_T}{\sqrt{n}}\right) \le \sqrt{\frac{n+1}{n}}\left(\sqrt{c_{T,q}^{\mathrm{gr}}} + \frac{2K_T}{\sqrt{n}}\right).$$

Taking the supremum and then $m \to \infty$ proves the upper bound. Finiteness of $c_{T,q}^{\mathrm{gr}}$ follows from the sharp Dirichlet observation bound and the sectorial estimate of Theorem 3; the latter is independent of dimension.

Conversely, fix any graph model and unit vector $z \in E$. On $E \oplus F$, let $P_i(\epsilon)$ be the projection onto the graph $\{(u, \epsilon X_i u) : u \in E\}$. The transfer of E-coordinates between successive graphs is

$$(I + \epsilon^2 X_i^* X_i)^{-1}(I + \epsilon^2 X_i^* X_{i-1}).$$

Because $X_0 = X_T = 0$, the full cycle on $E \oplus \{0\}$ is $I - \epsilon^2 H_X + O(\epsilon^4)$. With $\epsilon = m^{-1/2}$, its mth power tends to e^{-H_X}, and m times the endpoint loss tends to $\|B_X^{1/2}e^{-H_X}z\|^2$. All these projections have rank $\dim E$; if $\dim E < q < \infty$, adjoin a common fixed space to make their rank exactly q. Taking the supremum proves the lower bound.

For maximum-task loss, replace B_W in the upper argument by each individual compressed $I - P_i$ and B_X by $X_i^* X_i$. The same error estimate applies because $\|X_i\| \le 2\sqrt{n}$. In the graph realization, every individual loss converges, so its maximum converges as well. □

The graph generator also has the familiar quadratic energy identity

$$(H_X)_s = \frac{1}{2}\sum_{i=1}^{T}(X_i - X_{i-1})^*(X_i - X_{i-1}), \qquad B_X \preceq 2\kappa_T (H_X)_s.$$

The skew part has relative norm at most $\cot(\pi/T)$, by (15). Thus both the original finite product and its limiting graph model are governed by the *same* Dirichlet path operator. This is the common structural reason for the sharp observation and sector constants used in the main theorem.

For $q = 1$, the graph generator is scalar and optimizing $2\kappa_T h e^{-2h}$ recovers $\lim_m m\mathcal{F}(T, m, 1) = \kappa_T/e$. For maximum-task loss, the discrete point-observation inequality gives the analogous constant $\lfloor T^2/4 \rfloor/(eT)$.

Zero common fixed space.

In the graph lower realization, the common fixed directions have E-coordinates in $\bigcap_i \ker X_i$. By embedding the codomain in a larger finite-dimensional space and adding an arbitrarily small injective component to one X_i, any fixed graph objective can be approximated with $\bigcap_i \ker X_i = \{0\}$. The matching absorber construction below is already injective, so in particular the exact unrestricted constants are unchanged under this restriction.

B A Cauchy-frequency absorber for the matching lower constants

We prove the matching lower bounds using the same closed graph geometry as Appendix A. The analytic step is a rank-one scattering limit, proved below by Volterra

equations, Fourier–Plancherel, and bounded approximation. A Cauchy spectral density makes its limiting transform rational, avoiding the logarithmic boundary-value calculation for a uniform spectral band.

Theorem 9. *For unrestricted real projection families,*

$$\liminf_{T\to\infty}\liminf_{m\to\infty}\frac{mM_T(m)}{T^2}\ge\frac{1}{2\pi^4},\qquad \liminf_{T\to\infty}\liminf_{m\to\infty}\frac{mN_T(m)}{T^2}\ge\frac{1}{8\pi^2}.$$

Both lower bounds admit finite real families with zero common fixed space.

B.1 Graph coordinates and an anchored compression

Let E,F be finite real Hilbert spaces, $X_0=X_T=0$, and $X_i:E\to F$. As in Theorem 8, set

$$H_X=\sum_{i=1}^{T}X_i^*(X_i-X_{i-1}),\qquad B_X=\frac{1}{T}\sum_{i=1}^{T-1}X_i^*X_i.$$

The graph realization already proved in that theorem implies

$$\liminf_{m\to\infty}mM_T(m)\ge\|B_X^{1/2}e^{-H_X}\|^2,\quad \liminf_{m\to\infty}mN_T(m)\ge\max_i\|X_ie^{-H_X}\|^2. \tag{21}$$

Put $J_T=L_T^{-1/2}(S^*-S)L_T^{-1/2}$, where $L_T=2I-S-S^*$ and S is the lower shift on $\mathbb{R}^{T-1}$. By (15), $\|J_T\|=\lambda_T=\cot(\pi/T)$. For $Y:E\to\mathbb{R}^d\otimes\mathbb{R}^{T-1}$, choosing $X=(I_d\otimes L_T^{-1/2})Y$ gives

$$H_X=\tfrac{1}{2}Y^*(I+I_d\otimes J_T)Y,\qquad B_X=\tfrac{1}{T}Y^*(I_d\otimes L_T^{-1})Y. \tag{22}$$

Write $e_T(j)=\sqrt{2/T}\sin(\pi j/T)$, $\kappa_T=[4T\sin^2(\pi/(2T))]^{-1}$, and $\gamma_j=j(T-j)/T$. The elementary sine-product identities imply

$$\|J_Te_T\|>T/16,\qquad \left\|J_TL_T^{-1/2}e_j/\sqrt{\gamma_j}\right\|>T/16\quad(j=\lfloor T/2\rfloor,\ T\ge4). \tag{23}$$

For reference, the first estimate follows by pairing J_Te_T with the second Dirichlet sine vector, giving $\|J_Te_T\|\ge2T/\pi^3>T/16$; the second follows by applying the same test vector to the Green-function column $(L_T^{-1})_{kj}=\min(k,j)-kj/T$, giving a lower bound $T/\pi^2>T/16$.

Lemma 10 (Anchored compression)**.** *Let $T\ge4$, $S_0\in\mathbb{R}^{d\times d}$ be skew symmetric, and $u\in\mathbb{R}^d$, $g\in\mathbb{R}^{T-1}$ be unit vectors. If, for some $0<\delta<1$,*

$$\|S_0\|\le(1-\delta)\lambda_T,\qquad \|S_0u\|^2\le\delta\|J_Tg\|^2,$$

then there is a real isometry $Q : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^{T-1}$ *with* $Q^*(I_d \otimes J_T)Q = S_0$ *and* $Qu = e_1 \otimes g$. *For any* $D > 0$, *taking* $Y = \sqrt{2}\, QD^{1/2}$ *in* (22) *realizes* $H_X = D + D^{1/2} S_0 D^{1/2}$. *Moreover*

$$g = e_T : \qquad B_X \succeq 2\kappa_T D^{1/2} uu^* D^{1/2},$$
$$g = L_T^{-1/2} e_j / \sqrt{\gamma_j} : \; X_j^* X_j \succeq 2\gamma_j D^{1/2} uu^* D^{1/2}.$$

Proof Use coordinates with $u = e_1$ and let $J = I_d \otimes J_T$, $e = e_1 \otimes g$, $t = \|Je\|$, $f = Je/t$. In block form write $S_0 = \left(\begin{smallmatrix} 0 & -a^* \\ a & S_1 \end{smallmatrix}\right)$, $c = a/t$ and $G = I - cc^*$. The assumptions yield $G \succeq (1-\delta)I$ and $\|G^{-1/2} S_1 G^{-1/2}\| \le \lambda_T$. Using the $d-1$ remaining fibers, take an isometry V with image orthogonal to the first fiber and $V^* J V = G^{-1/2} S_1 G^{-1/2}$. Here is an explicit compression, avoiding a dilation theorem. Orthogonally block-diagonalize the target skew matrix into $\left(\begin{smallmatrix} 0 & -\eta_j \\ \eta_j & 0 \end{smallmatrix}\right)$, $0 \le \eta_j \le \lambda_T$, and possibly one zero direction. For each two-dimensional target block, take two unused copies of J_T, and choose unit vectors a_j, b_j in a maximal-frequency plane of the first copy with $J_T a_j = \lambda_T b_j$. Take a unit vector c_j in the second copy, and map the target basis to

$$a_j, \qquad (\eta_j/\lambda_T) b_j + \sqrt{1 - (\eta_j/\lambda_T)^2}\, c_j.$$

These images are orthonormal and their skew compression is exactly the desired block. Different target blocks use disjoint fibers. There are $d-1$ fibers for a $(d-1)$-dimensional target, so $2\lfloor (d-1)/2 \rfloor \le d-1$ supplies all blocks; a possible remaining scalar uses the spare fiber. Then $Q = [\, e \quad fc^* + VG^{1/2} \,]$ is an isometry and $Q^* J Q = S_0$. Indeed, $e \perp f$, and the ranges of V and $\{e, f\}$ lie in orthogonal fibers, so the lower-right Gram block is $cc^* + G = I$. Moreover $e^* J f = -t$ and $f^* J f = 0$; the off-diagonal skew block is therefore $-tc^* = -a^*$, while the remaining skew block equals $G^{1/2}(V^* J V) G^{1/2} = S_1$. The two observation bounds follow respectively from $L_T^{-1} \succeq \lambda_1(L_T)^{-1} e_T e_T^*$ and from the jth row of $(I_d \otimes L_T^{-1/2})Y$. □

B.2 A rank-one transmission identity

What the limit actually uses.

This is a one-dimensional input–output calculation, not a theorem from abstract scattering theory. Its proof has four independently checkable steps: (i) Duhamel gives a scalar Volterra equation for the amplitude along v; (ii) its Laplace transform is $F/(1+bF)$; (iii) the dissipative energy estimate and Plancherel give convergence of the amplitudes in L^2; and (iv) the endpoint response is a bounded, strongly convergent Fourier map on a *fixed bounded* frequency interval. Only after this limit do we let the frequency cutoff grow. These qualifications are essential to the quantifiers.

Let μ be a symmetric probability measure supported on $[-R, R]$ with a bounded piecewise C^1 density, positive on the interior. This suffices for the truncated Cauchy laws below. On $\mathcal{H} = L^2(\mu)$ let $\Omega f(\xi) = \xi f(\xi)$, $v = 1$, and $P = vv^*$. Fix $b > 0$ and put

$$D_s = |\Omega| + sbP, \qquad H_s = D_s + is\Omega. \tag{24}$$

The following identity gives an explicit asymptotic output without assuming that D_s is bounded below on the auxiliary continuum.

Lemma 11 (Transmission identity). *Put* $F_\mu(z) = \int (z+i\xi)^{-1} d\mu(\xi)$ *for* $\Re z > 0$. *Then*

$$\lim_{s\to\infty} b\|e^{-H_s}v\|^2 = b\int_{-R}^{R} \frac{e^{-2|\xi|}}{|1+bF_\mu(-i\xi+0)|^2}\, d\mu(\xi). \tag{25}$$

The same limit holds for H_s^*.

Proof Use fast time $G_s = H_s/s = |\Omega|/s + bP + i\Omega$. Writing $g_s(t) = \langle v, e^{-tG_s}v\rangle$, Duhamel's formula and the Laplace transform give

$$\widehat{g}_s(z) = \frac{F_s(z)}{1+bF_s(z)}, \qquad F_s(z) = \int \frac{d\mu(\xi)}{z+|\xi|/s+i\xi}.$$

The contraction energy identity gives $2b\int_0^\infty |g_s(t)|^2 dt \le 1$. We record why the needed boundary convergence holds for this piecewise smooth density. Fix ω such that $\xi_0 = -\omega$ lies away from the density's finitely many breakpoints and from 0. Near ξ_0 let $h_s = |\xi_0|/s$. The elementary inequality $\big||\xi| - |\xi_0|\big| \le |\xi - \xi_0|$ gives

$$\left|\frac{1}{|\xi|/s + i(\xi-\xi_0)} - \frac{1}{h_s + i(\xi-\xi_0)}\right| \le \frac{1}{s\sqrt{h_s^2 + (\xi-\xi_0)^2}}.$$

The integral of the right-hand side against the bounded density is $O_\omega((1+\log s)/s)$. With the constant floor h_s, subtract the value of the density at ξ_0 in a small neighborhood; local Lipschitz continuity handles the difference, while the constant part integrates explicitly. Hence

$$F_s(i\omega) \longrightarrow F_\mu(i\omega+0) = \pi\rho(-\omega) - i\,\text{p.v.}\int \frac{\rho(\xi)}{\xi+\omega}\, d\xi.$$

For ω outside the support, ordinary dominated convergence applies. The excluded frequencies have measure zero. The same limit holds after applying $z \mapsto z/(1+bz)$, since $\Re F_s \ge 0$ and $b > 0$. As $\Re F_s \ge 0$, $|F_s/(1+bF_s)| \le 1/b$; for $|\omega| > 2R$ this is also at most $(|\omega|-R)^{-1}$. Dominated convergence and Plancherel therefore give $g_s \to g_\infty$ in $L^2(0,\infty)$, where $\widehat{g}_\infty = F_\mu/(1+bF_\mu)$.

At physical time 1, variation of constants gives

$$e^{is\xi}(e^{-H_s}v)(\xi) = e^{-|\xi|}\left[1 - b\int_0^s e^{|\xi|t/s + i\xi t} g_s(t)\, dt\right]. \tag{26}$$

The integral operator from $L^2(0,\infty)$ to $L^2(\mu)$ in brackets is uniformly bounded: expand $e^{|\xi|t/s}$ on $0 \le t \le s$ in powers of $|\xi|t/s$, use Plancherel for each power, and sum the majorant $\sum_{k\ge0} R^k/k! = e^R$. On every fixed input it converges strongly to the one-sided Fourier transform, since the $k = 0$ term converges by truncation while all $k \ge 1$ terms tend to zero by dominated convergence. Thus the integral converges in $L^2(\mu)$ to $\widehat{g}_\infty(-i\xi+0) = F_\mu(-i\xi+0)/(1+bF_\mu(-i\xi+0))$. Equation (26) proves (25). Symmetry and complex conjugation exchange H_s and H_s^* and fix v. □

B.3 The Cauchy spectral law makes the limit rational

Why the Cauchy law is natural.

The entire choice is motivated by one elementary identity: a Cauchy random variable of scale L has characteristic function $\int_{\mathbb{R}} e^{-it\xi}\rho_L(\xi)\, d\xi = e^{-L|t|}$. In the limiting fast-time rank-one equation, the scalar amplitude therefore obeys the convolution equation

$$g(t) = e^{-Lt} - b\int_0^t e^{-L(t-u)} g(u)\, du, \qquad \text{hence} \qquad g(t) = e^{-(L+b)t}.$$

Thus the only nontrivial analytic limit has an *exponential* kernel and a rational Laplace transform. The frequency cutoff below is used solely to return to bounded, finite-dimensional real models; it is not an additional dynamical mechanism.

Fix $L > 0$, and let

$$\rho_L(\xi) = \frac{L}{\pi(L^2+\xi^2)}, \qquad d\mu_{L,R}(\xi) = \frac{\mathbf{1}_{[-R,R]}(\xi)\rho_L(\xi)}{\int_{-R}^{R}\rho_L(\eta)d\eta}\,d\xi, \qquad b = L.$$

Write $C_{L,R}$ for the right side of (25) with $\mu = \mu_{L,R}$.

Lemma 12 (Explicit Cauchy limit)**.** *For every $L > 0$,*

$$\lim_{R\to\infty} C_{L,R} = C_L := \frac{L^2}{\pi}\int_{\mathbb{R}} \frac{e^{-2|\xi|}}{4L^2+\xi^2}\,d\xi, \qquad \lim_{L\to\infty} C_L = \frac{1}{4\pi}. \tag{27}$$

Proof For the full Cauchy law $\rho_L(\xi)d\xi$, its characteristic function on $t \geq 0$ is e^{-Lt}. Therefore its Cauchy transform is

$$F_L(z) = \int_0^\infty e^{-zt}e^{-Lt}\,dt = \frac{1}{z+L}.$$

The truncation limit requires no boundary-value theorem. Let $p_{L,R} = \int_{-R}^{R}\rho_L(\eta)\,d\eta$ and $E_{L,R}(z) = \int_{|\eta|>R}\rho_L(\eta)/(z+i\eta)\,d\eta$. For $\Re z > 0$,

$$F_{\mu_{L,R}}(z) = \frac{F_L(z) - E_{L,R}(z)}{p_{L,R}}.$$

For any fixed real ξ and $R > 2|\xi|$, the tail integral is analytic near $z = -i\xi$ and, uniformly for $\varepsilon \geq 0$,

$$|E_{L,R}(\varepsilon - i\xi)| \leq 2\int_{|\eta|>R}\frac{\rho_L(\eta)}{|\eta|}\,d\eta \leq \frac{2L}{\pi R^2}.$$

Since $p_{L,R} \to 1$ and $F_L(z) = (z+L)^{-1}$ extends continuously to the imaginary axis, we obtain directly $F_{\mu_{L,R}}(-i\xi+0) \to (L-i\xi)^{-1}$. As $\Re F_{\mu_{L,R}}(-i\xi+0) \geq 0$, the denominators in (25) have modulus at least one. For $R \geq L$, the densities $d\mu_{L,R}/d\xi$ are bounded by a fixed multiple of ρ_L, so dominated convergence applies. With $b = L$,

$$\frac{1}{|1+LF_L(-i\xi)|^2} = \frac{L^2+\xi^2}{4L^2+\xi^2}.$$

This proves the first formula. For the second, write the integrand as $\pi^{-1}e^{-2|\xi|}L^2/(4L^2+\xi^2)$ and apply dominated convergence, using $\int_{\mathbb{R}} e^{-2|\xi|}d\xi = 1$. □

B.4 Finite real models and completion

For a fixed L, R divide $[-R, R]$ into N equal cells (N even), with midpoints ξ_j and exact cell masses $w_j = \mu_{L,R}(I_j)$. Set $v_N = (\sqrt{w_j})_j$, $\Omega_N = \mathrm{diag}(\xi_j)$ and $D_{0,N} = |\Omega_N|$. The finite absorber is

$$D = D_{0,N} + sbv_Nv_N^*, \qquad H = D + is\Omega_N.$$

Embedding it as step functions in $L^2(\mu_{L,R})$ and using Duhamel's formula gives

$$\left|b\|e^{-H}v_N\|^2 - b\|e^{-H_s}v\|^2\right| \le \frac{2b(1+s)R}{N}. \tag{28}$$

Both matrices and the auxiliary generator are accretive; their semigroups are contractions. The bound follows from the uniform midpoint error R/N for Ω and $|\Omega|$.

Realify with

$$D_R = \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix}, \qquad K_R = \begin{pmatrix} 0 & -s\Omega_N \\ s\Omega_N & 0 \end{pmatrix}, \qquad S_R = D_R^{-1/2} K_R D_R^{-1/2}.$$

Since $D \succeq |\Omega_N|$, $\|S_R\| \le s$. Let $a_0 = v_N^* D_{0,N}^{-1} v_N \ge 1/R$, and set $\alpha = v_N^* D^{-1} v_N = a_0/(1+sba_0)$. For the real unit vector $u = (D^{-1/2}v_N/\sqrt{\alpha}, 0)$ and $v_R = (v_N, 0)$,

$$D_R^{1/2} u u^* D_R^{1/2} \succeq sb\, v_R v_R^*, \qquad \|S_R u\|^2 = \frac{s^2}{1+sba_0} \le \frac{sR}{b}. \tag{29}$$

The equality uses the symmetry of the quadrature: $w_j = w_{N+1-j}$ and $\xi_j = -\xi_{N+1-j}$, so $v_N^* D_{0,N}^{-1} \operatorname{sgn}(\Omega_N) v_N = 0$; the rank-one inverse identity then gives the formula. For clarity, write $A_0 = D_{0,N}^{-1}$ and $h = 1 + sba_0$. Sherman–Morrison gives $D^{-1}v_N = A_0 v_N/h$ and $\alpha = a_0/h$. The symmetry cancellation above also gives $D^{-1}\operatorname{sgn}(\Omega_N)v_N = A_0 \operatorname{sgn}(\Omega_N) v_N$. Consequently,

$$\|S_R u\|^2 = \frac{s^2}{\alpha}\|D^{-1/2}\Omega_N D^{-1} v_N\|^2 = \frac{s^2}{\alpha h^2}\|D^{-1/2}\operatorname{sgn}(\Omega_N)v_N\|^2 = \frac{s^2}{h}.$$

The observation dominance follows separately from $\alpha = a_0/h \le (sb)^{-1}$, which implies $D_R^{1/2} u u^* D_R^{1/2} = \alpha^{-1} v_R v_R^* \succeq sb\, v_R v_R^*$.

Proof of Theorem 9 Fix L, R. For large T, put $\delta_T = T^{-1/3}$ and $s_T = (1-\delta_T)\lambda_T$. Choose even N_T so large that (28) vanishes as $T \to \infty$, for example with $N_T/(1+s_T) \to \infty$ while L, R are fixed. For every individual T, this chooses a *finite* real matrix before any minimax limit in m is taken. Lemma 11 then gives

$$b\|e^{-H^*_{s_T,N_T}} v_{N_T}\|^2 \longrightarrow C_{L,R}. \tag{30}$$

By (29) and (23),

$$\frac{\|S_R u\|^2}{\|J_T g\|^2} \le \frac{256 s_T R}{bT^2} = O_{L,R}(T^{-1}) \le \delta_T,$$

for either of the two choices of g in Lemma 10. That lemma produces a finite real graph whose generator is $D_R + K_R$ and whose mean observation satisfies $B_X \succeq 2\kappa_T sb\, v_R v_R^*$. Using (21) and (30),

$$\liminf_{m\to\infty} m M_T(m) \ge 2\kappa_T s_T \left[b\|e^{-H^*_{s_T,N_T}} v_{N_T}\|^2\right].$$

Since $2\kappa_T s_T/T^2 \to 2/\pi^3$, this yields $\liminf_T \liminf_m m M_T(m)/T^2 \ge 2C_{L,R}/\pi^3$. First send $R \to \infty$, then $L \to \infty$; by Lemma 12 the result is $1/(2\pi^4)$.

For maximum-task loss, take $g = L_T^{-1/2} e_{\lfloor T/2 \rfloor} / \sqrt{\gamma_{\lfloor T/2 \rfloor}}$. The same argument gives

$$\liminf_m m N_T(m) \geq 2\gamma_{\lfloor T/2 \rfloor} s_T \left[b \| e^{-H^*_{s_T, N_T}} v_{N_T} \|^2 \right].$$

Because $\gamma_{\lfloor T/2 \rfloor}/T \to 1/4$ and $s_T/T \to 1/\pi$, the normalized prefactor tends to $2(1/4)(1/\pi) = 1/(2\pi)$, and the absorber factor tends (first in R, then in L) to $1/(4\pi)$. Thus the lower constant is $(1/2\pi)(1/4\pi) = 1/(8\pi^2)$. The graph is injective since $D_R > 0$ and the stacked X_i are injective; therefore its common fixed space is zero. □

C Sharp sectorial dissipation and the exact constants

The scalar regularization below closes the weighted contour estimate. The matching lower bounds are proved in Appendix B, and the graph reduction is proved in Appendix A. All upper estimates are uniform over arbitrary finite dimensions.

For a finite complex matrix $H = D + iC$, assume that D is positive definite and C is Hermitian. Put

$$S = D^{-1/2} C D^{-1/2}, \qquad \Gamma(s) = \sup_{\|S\| \leq s} \left\| D^{1/2} e^{-H} \right\|^2, \qquad s \geq 0,$$

where the supremum includes arbitrary finite dimensions and arbitrary positive definite D.

Theorem 13 (Sharp large-sector energy constant)**.** *One has*

$$\lim_{s \to \infty} \frac{\Gamma(s)}{s} = \frac{1}{4\pi}. \tag{31}$$

More precisely, the universal upper estimate is

$$\Gamma(s) \leq \frac{s}{4\pi} + O(s^{2/3}) \qquad (s \to \infty), \tag{32}$$

with an absolute implicit constant independent of matrix dimension. The limit is unchanged if the supremum is restricted to real $H = D + K$, with D real positive definite and K real skew symmetric.

C.1 Two dimension-free energy estimates

Why the contour is used.

The only purpose of the two-ray contour is to recover the *sharp* coefficient $1/(4\pi)$ uniformly in matrix dimension. The first lemma controls a time derivative by the two ordinary contraction energy identities. The second is a weighted resolvent identity: after Cauchy–Schwarz, the total energy along the two rays has an exact scalar value. The subsequent scalar regularization is needed because D may approach a singular matrix. No infinite-dimensional functional calculus or general scattering theorem is used here.

Lemma 14 (One derivative from two energy integrals)**.** *If* $H = D + iC$ *and* $\left\|D^{-1/2} C D^{-1/2}\right\| \le s$*, then*

$$\left\|H e^{-H}\right\| \le \frac{\sqrt{1+s^2}}{2}. \tag{33}$$

Proof The form representation $H = D^{1/2}(I + iS)D^{1/2}$ gives

$$|\langle y, Hx\rangle| \le \sqrt{1+s^2}\, \left\|D^{1/2} y\right\| \left\|D^{1/2} x\right\|.$$

The energy identities for the semigroup and its adjoint are

$$\int_0^1 \left\|D^{1/2} e^{-tH} x\right\|^2 dt = \tfrac{1}{2}(\|x\|^2 - \left\|e^{-H}x\right\|^2) \le \tfrac{1}{2}\|x\|^2,$$
$$\int_0^1 \left\|D^{1/2} e^{-tH^*} y\right\|^2 dt \le \tfrac{1}{2}\|y\|^2.$$

Since H commutes with its exponential,

$$H e^{-H} = \int_0^1 e^{-(1-t)H} H e^{-tH}\, dt.$$

Pair this identity with y, x, apply the form bound inside the integral, and use Cauchy–Schwarz in t. The two energy integrals prove (33). □

Define

$$c(a) = \frac{(1+a^2)(\pi + 2\arctan a)}{4\pi^2 a}, \qquad a > 0.$$

Lemma 15 (Weighted contour bound)**.** *For* $as < 1$*,*

$$D^{1/2} e^{-H} e^{-H^*} D^{1/2} \preceq c(a)(I - a^2 S^2)^{-1}. \tag{34}$$

Proof On the two half-rays put

$$z(\omega) = -a|\omega| + i\omega, \qquad R(\omega) = (H + z(\omega) I)^{-1}, \qquad D_\pm = D \pm aC.$$

The sector hypothesis makes $D_\pm$ positive definite and ensures that the resolvents exist. Direct multiplication gives

$$R^* D_\pm R = \mathrm{Re}[(1 \pm ia)R] \quad \text{on } \mathbb{R}_\pm.$$

Here is an elementary proof of the exact two-ray energy identity, using only the resolvent identity, a scalar angle, and one finite-dimensional homotopy. In particular, no matrix logarithm or choice of its branch is needed. Write $z_+ = (-a + i)R$, $z_- = (-a - i)R$, and let $\Gamma_R = \{z(\omega) : -R \le \omega \le R\}$, oriented upwards. For $0 \le \tau \le 1$ put $H_\tau = (1-\tau)I + \tau H$ and

$$\Phi_R(\tau) = \int_{\Gamma_R} (H_\tau + zI)^{-1}\, dz.$$

The Hermitian and skew-Hermitian parts of H_τ are $D_\tau = (1-\tau)I + \tau D > 0$ and τC, respectively, so that $-sD_\tau \preceq \tau C \preceq sD_\tau$. Every eigenvalue λ of H_τ therefore satisfies $\Re\lambda > 0$ and $|\Im\lambda| \le s\Re\lambda$. As $as < 1$, the poles $z = -\lambda$ lie strictly to the left of Γ_R. All inverses above exist throughout the homotopy.

Since H_τ and $H - I$ commute, differentiation under this *finite* contour integral and the elementary identity $d(H_\tau + zI)^{-1}/dz = -(H_\tau + zI)^{-2}$ give

$$\frac{d}{d\tau}\Phi_R(\tau) = (H - I)\big[(H_\tau + z_+ I)^{-1} - (H_\tau + z_- I)^{-1}\big].$$

For the fixed finite matrix H, the two endpoint resolvents are $O_H(R^{-1})$, uniformly in $\tau \in [0, 1]$. Thus $\Phi_R(1) = \Phi_R(0) + O_H(R^{-1})$. At $\tau = 0$, the integrand is scalar, and direct integration yields

$$\Phi_R(0) = \big[\log(1+z_+) - \log(1+z_-)\big] I = 2i \arg(1 - aR + iR)\, I \longrightarrow 2i\vartheta I, \qquad \vartheta = \frac{\pi}{2} + \arctan a.$$

Only the ordinary scalar logarithm is used here. Consequently,

$$\lim_{R\to\infty} \int_{\Gamma_R} (H + zI)^{-1}\, dz = 2i\vartheta I. \tag{35}$$

On the positive and negative rays one has, respectively, $-iz'(\omega) = 1 + ia$ and $-iz'(\omega) = 1 - ia$; direct multiplication now gives

$$R(\omega)^* D_\pm R(\omega) = \mathrm{Re}\,\big[-iz'(\omega) R(\omega)\big].$$

Both energy integrals converge absolutely because $R(\omega) = O_H(|\omega|^{-1})$ and therefore $R^* D_\pm R = O_H(|\omega|^{-2})$. Taking the Hermitian real part of $-i$ times (35) proves

$$\int_{\mathbb{R}_+} R^* D_+ R\, d\omega + \int_{\mathbb{R}_-} R^* D_- R\, d\omega = (\pi + 2\arctan a) I. \tag{36}$$

The inverse-Laplace contour formula is

$$e^{-H} = \frac{1}{2\pi} \sum_{\pm} \int_{\mathbb{R}_\pm} (1 \pm ia) e^{z(\omega)} R(\omega)\, d\omega.$$

To justify it, start with the standard vertical inverse-Laplace line $\Re z = \sigma > 0$ and deform it to the two rays. Every pole $z = -\lambda$ lies strictly to the left: indeed $\Re\lambda > a|\Im\lambda|$. Finite-dimensional resolvents are $O_H(|z|^{-1})$ at infinity; along the rays $|e^{z(\omega)}| = e^{-a|\omega|}$. To justify deformation on truncated contours, join the vertical line and the rays at imaginary heights $\pm R$. On each horizontal connector, $\|R(z)\| = O_H(R^{-1})$ and $\int_{-aR}^{\sigma} e^x\, dx \le e^\sigma$; hence both connector integrals are $O_H(R^{-1})$. Thus the ray integrals converge absolutely for each fixed finite matrix and the contour deformation is valid. For a matrix element against u, x, factor $D^{1/2}R$ through $D_\pm^{-1/2}$ and $D_\pm^{1/2}$. Cauchy–Schwarz on the union of both rays and (36) yield

$$\begin{aligned} |\left\langle u, D^{1/2} e^{-H} x \right\rangle|^2 &\le \frac{\pi + 2\arctan a}{4\pi^2} \|x\|^2 \frac{1 + a^2}{2a} \left\langle u, [(I + aS)^{-1} + (I - aS)^{-1}] u \right\rangle \\ &= c(a) \|x\|^2 \left\langle u, (I - a^2 S^2)^{-1} u \right\rangle. \end{aligned}$$

Taking the supremum over unit x proves the lemma. □

C.2 Vanishing scalar regularization

Lemma 16 (Quadratic bound after regularization)**.** *Let $r > 0$, $a > 0$, and $as < 1$. Set*

$$k = \frac{a^2}{1 - a^2 s^2}, \qquad A = c(a)(1 + 2k), \qquad B = \frac{c(a) k (1 + s^2)}{2r}.$$

Then every matrix in the definition of $\Gamma(s)$ satisfies

$$\left\| D^{1/2} e^{-H} \right\|^2 \le e^{2r} \frac{A + \sqrt{A^2 + 4B}}{2}. \tag{37}$$

Proof Put

$$H_r = H + rI, \qquad D_r = D + rI, \qquad S_r = D_r^{-1/2} C D_r^{-1/2}, \qquad G = D_r^{1/2} e^{-H_r}.$$

The inequalities $-sD \preceq C \preceq sD$ imply $-sD_r \preceq C \preceq sD_r$, so $\|S_r\| \le s$. Fix a unit vector x and set $E = \|Gx\|^2$. If $E = 0$ there is nothing to prove. Otherwise let $u = Gx/\sqrt{E}$ and $q = \|S_r G x\|^2$. Since

$$\left\|G^* u\right\| \ge |\,\langle x, G^* u\rangle\,| = \sqrt{E},$$

Lemma 15, applied to H_r, gives

$$\begin{aligned} E &\le c(a)\left\langle u, (I - a^2 S_r^2)^{-1} u\right\rangle \\ &\le c(a)\big(1 + k\,\|S_r u\|^2\big) = c(a)\left(1 + k\frac{q}{E}\right). \end{aligned} \tag{38}$$

The second inequality follows by the spectral theorem from

$$\frac{1}{1 - a^2 t^2} \le 1 + \frac{a^2}{1 - a^2 s^2} t^2 \quad (|t| \le s).$$

The key estimate uses the scalar floor $D_r \succeq rI$:

$$\begin{aligned} q &= \left\|D_r^{-1/2} C e^{-H_r} x\right\|^2 \\ &= \left\|D_r^{-1/2}(H_r - D_r) e^{-H_r} x\right\|^2 \\ &\le 2r^{-1}\left\|H_r e^{-H_r} x\right\|^2 + 2\left\|D_r^{1/2} e^{-H_r} x\right\|^2 \\ &\le \frac{1 + s^2}{2r} + 2E. \end{aligned} \tag{39}$$

The factor i in $H_r - D_r = iC$ disappears inside the norm. The final step uses Lemma 14, with the same sector bound s. Combining (38) and (39) gives

$$E^2 \le c(a)(1 + 2k)E + \frac{c(a)k(1 + s^2)}{2r} = AE + B.$$

Thus $E \le (A + \sqrt{A^2 + 4B})/2$, uniformly over unit x.

Finally, scalar commutation gives the exact identity

$$\left\|D_r^{1/2} e^{-H_r} x\right\|^2 = e^{-2r}\left(\left\|D^{1/2} e^{-H} x\right\|^2 + r\left\|e^{-H} x\right\|^2\right) \ge e^{-2r}\left\|D^{1/2} e^{-H} x\right\|^2.$$

Taking suprema over unit x proves (37). □

Proof of the upper bound in Theorem 13 For $s \ge 8$, choose

$$r = \delta = s^{-1/3}, \qquad a = \frac{1 - \delta}{s}.$$

Then $as = 1 - \delta < 1$ and

$$\begin{aligned} c(a) &= \frac{s}{4\pi}\big(1 + O(s^{-1/3})\big), \\ k &= \frac{(1 - \delta)^2}{s^2(2\delta - \delta^2)} = O(s^{-5/3}), \\ A &= \frac{s}{4\pi} + O(s^{2/3}), \qquad B = O(s^{5/3}). \end{aligned}$$

All constants are absolute. Since A is bounded below by a positive absolute multiple of s,

$$\frac{A + \sqrt{A^2 + 4B}}{2} \le A + \frac{B}{A} = \frac{s}{4\pi} + O(s^{2/3}).$$

Also $e^{2r} = 1 + O(s^{-1/3})$. Lemma 16 proves (32) and $\limsup_s \Gamma(s)/s \le 1/(4\pi)$. □

Why regularization is needed.

No claim is made that an unregularized extremal output has small normalized skew. Near-singular dilations can make that claim false. Estimate (39) uses the strictly positive scalar floor of $D + rI$ before r is sent to zero. Its cost is r^{-1}, which becomes negligible after the quadratic inequality is divided by s^2.

C.3 The matching lower bound

The Cauchy absorber of Appendix B also proves the lower half of the sectorial constant. For each fixed L, R, take $\mu = \mu_{L,R}$, $b = L$, and $H_s = |\Omega| + sbvv^* + is\Omega$. The finite midpoint discretizations have positive definite real parts and relative sector bound at most s. Because $D_s \succeq sbvv^*$, passage from the finite matrices to the auxiliary continuum gives

$$\Gamma(s) \geq sb\|e^{-H_s^*}v\|^2.$$

Lemmas 11 and 12 show

$$\liminf_{s\to\infty} \frac{\Gamma(s)}{s} \geq C_{L,R}, \qquad \lim_{L\to\infty}\lim_{R\to\infty} C_{L,R} = \frac{1}{4\pi}.$$

The finite matrices can be realified without changing norms or the sector bound. This proves the lower limit and the real assertion.

Completion of the constant argument.

The upper bound of Theorem 3 follows from the weighted contour estimate and scalar regularization above. Its lower bound is realized by the absorber of Appendix B. The exact projection constants then follow in the main text from the graph variational principle and its Dirichlet observation estimates.

D Unit-row mean squared error

Proposition 17. *Under the same exact fitting rule, replace the task loss by* $s_i^{-1}\|X_i w - b_i\|^2$ *and impose unit bounds on the row norms in place of the operator-norm constraint. The supremum of the endpoint task-average loss is* $\mathcal{F}(T, m, q)$.

Proof The covariance $X_i^\top X_i/s_i$ has trace at most one and is supported on $\operatorname{ran}(Q_i)$, so it is bounded above by Q_i. This bounds the task-average loss by the corresponding $F_m(x)$.

For the reverse direction fix any projection instance and let $e = A^m x$. For one task write $Q = I - P$ and $r = \operatorname{rank} Q$. For $r \geq 2$, choose $u = Qe/\|Qe\|$ if $Qe \neq 0$, and any unit vector in $\operatorname{ran}(Q)$ otherwise. Let B have orthonormal columns spanning $\operatorname{ran}(Q)$, starting with u, and let V be an $r \times r$ orthogonal matrix with first column $r^{-1/2}\mathbf{1}$. For $0 < \varepsilon < 1$, take

$$X = V \operatorname{diag}\left(\sqrt{r(1-\varepsilon)}, \sqrt{\frac{r\varepsilon}{r-1}}, \ldots, \sqrt{\frac{r\varepsilon}{r-1}}\right) B^\top.$$

Every row has squared norm $(1-\varepsilon) + \frac{r\varepsilon}{r-1}(1-1/r) = 1$. The covariance is $(1-\varepsilon)uu^\top + \frac{\varepsilon}{r-1}(Q - uu^\top)$, so the task MSE at e is $(1-\varepsilon)\|Qe\|^2$. For $r = 1$, take one unit row in $\operatorname{ran}(Q)$, and for $r = 0$, take one zero row. These choices preserve the entire projection trajectory

because each row space is preserved. The matrices and compatible labels are chosen before training for the prescribed horizon. The task-average MSE is at least $(1-\varepsilon)F_m(x)$. Letting ε decrease to zero after taking suprema gives the claimed equality. □

Statements and Declarations

Data availability

No external datasets were used in this theoretical study. Figure 1 presents numerical evaluations of the deterministic finite-dimensional construction specified in the manuscript; the parameters are given in the figure caption. The mathematical conclusions are established by analytical proofs and do not depend on empirical datasets.

Competing interests

The author declares no competing interests.

Funding

This research received no external funding.

Use of generative AI

ChatGPT was used as an auxiliary tool to support reasoning, proof checking, and manuscript preparation. The author reviewed the AI-assisted work and takes full responsibility for the results and content of the manuscript.

References

Meng Ding, Kaiyi Ji, Di Wang, and Jinhui Xu. Understanding forgetting in continual learning with linear regression. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 10978–11001, 2024. URL https://proceedings.mlr.press/v235/ding24c.html.

Thang Doan, Mehdi Abbana Bennani, Bogdan Mazoure, Guillaume Rabusseau, and Pierre Alquier. A theoretical analysis of catastrophic forgetting through the NTK overlap matrix. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1072–1080, 2021. URL https://proceedings.mlr.press/v130/doan21a.html.

Itay Evron, Edward Moroshko, Rachel Ward, Nathan Srebro, and Daniel Soudry. How catastrophic can catastrophic forgetting be in linear regression? In *Proceedings of the 35th Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 4028–4079, 2022. URL https://proceedings.mlr.press/v178/evron22a.html.

Itay Evron, Ran Levinstein, Matan Schliserman, Uri Sherman, Tomer Koren, Daniel Soudry, and Nathan Srebro. From continual learning to SGD and back: Better rates for continual linear models. In *Proceedings of the 37th International Conference on Algorithmic Learning Theory*, volume 313 of *Proceedings of Machine Learning Research*, pages 1–50, 2026. URL https://proceedings.mlr.press/v313/evron26a.html.

Mark Hay-Yat Kong. *Trajectories of Iterates of Orthogonal Projections*. PhD thesis, University of California, Los Angeles, 2026. URL https://escholarship.org/uc/item/0p25z0gt.

Sen Lin, Peizhong Ju, Yingbin Liang, and Ness Shroff. Theory on forgetting and generalization of continual learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 21078–21100, 2023. URL https://proceedings.mlr.press/v202/lin23f.html.

Timur Oikhberg. Products of orthogonal projections. *Proceedings of the American Mathematical Society*, 127(12):3659–3669, 1999. https://doi.org/10.1090/S0002-9939-99-05255-7.

Benjamin Recht and Christopher Ré. Toward a noncommutative arithmetic-geometric mean inequality: Conjectures, case-studies, and consequences. In *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 11.1–11.24, 2012. URL https://proceedings.mlr.press/v23/recht12.html.

Simeon Reich and Rafał Zalas. Polynomial estimates for the method of cyclic projections in Hilbert spaces. *Numerical Algorithms*, 94:1217–1242, 2023. https://doi.org/10.1007/s11075-023-01533-w.

William Swartworth, Deanna Needell, Rachel Ward, Mark Kong, and Halyun Jeong. Nearly optimal bounds for cyclic forgetting. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/d72ae75abaa70a3b19c5d4f436c680d1-Abstract-Conference.html.

Matan Tsipory, Ran Levinstein, Itay Evron, Mark Kong, Deanna Needell, and Daniel Soudry. Are greedy task orderings better than random in continual linear regression? In *Advances in Neural Information Processing Systems*, volume 38, 2025. https://doi.org/10.52202/085713-5192.