GEOSHRINK: ACCELERATING DIFFUSION TRANSFORMERS WITH TWO LINES OF CODE

Anonymous authors
Paper under double-blind review



Figure 1: **Qualitative comparison of GeoShrink.** Direct NFE reduction (left) vs. GeoShrink (right) under the same model-evaluation budget. GeoShrink better preserves visual structure and temporal evolution under aggressive acceleration.

ABSTRACT

Diffusion transformers incur substantial inference cost through repeated model evaluations along a sampling trajectory. We introduce GeoShrink, a training-free acceleration method that retains the original solver grid while evaluating the model only at a prescribed set of anchors. At skipped stages, GeoShrink predicts the solver-facing output by adding a geometrically retained fraction of the latest observed innovation to the most recent exact output. We derive this rule from chordal tangent transport and round-trip line projection, and establish a geometric anchor-spacing principle that minimizes the largest adjacent gap expansion under fixed coverage and first span. The analysis characterizes the geometric closure and propagation of prediction errors without assuming access to future model outputs. Experiments cover image, video, motion, and audio generation, together with adapted 3D backends. At approximately $5\times$ acceleration, GeoShrink improves FLUX PSNR by 3.10 dB over the strongest listed baseline. On HunyuanVideo, it achieves a reported $4.99\times$ speedup and improves ChronoMagic-Bench-150 PSNR by 5.44 dB over the strongest listed fidelity baseline. Comparisons at fixed evaluation budgets further show substantial gains on Motion, Audio, Music and 3d generation.

1 INTRODUCTION

Diffusion transformers (Peebles & Xie, 2023; Bao et al., 2023) have become a common backbone for generative modeling across images (Esser et al., 2024; Black Forest Labs, 2024), video (Kong et al., 2024), human motion (Wen et al., 2025), audio (Hung et al., 2024; Gong et al., 2025), and 3D content (Xiang et al., 2025; Zhao et al., 2025b). Despite their versatility, inference remains

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



Figure 2: **Motivation for adaptive innovation retention.** From left to right, we show the full-NFE reference, fixed-coefficient predictions with $\omega \in \{0, 0.5, 1.5, 2\}$, and GeoShrink. Small ω underuses the latest innovation, while aggressive reuse ($\omega > 1$) progressively introduces color, contrast, and appearance drift. GeoShrink instead adapts ω_t from observable trajectory geometry and more closely preserves the full-NFE reference.

expensive because a large neural network should be repeatedly evaluated along the sampling trajectory (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song & Ermon, 2019; Song et al., 2021b; Rombach et al., 2022). Reducing these evaluations without modifying or retraining the pretrained generator is therefore an important practical problem across generative domains.

Existing training-free acceleration methods exploit temporal redundancy through caching or forecasting intermediate features across denoising steps (Ma et al., 2024b; Selvaraju et al., 2024; Zhao et al., 2025a; Liu et al., 2025a; Zou et al., 2025; Lv et al., 2025; Liu et al., 2025b;c; Zhao et al., 2026). Feature caching reuses previously computed representations, while feature forecasting extrapolates them to approximate features at subsequent stages. While effective, these approaches rely on model-specific internal representations, making transfer to new backbones dependent on choices such as layers, tokens, cache locations, reuse intervals, or verification mechanisms (Zou et al., 2025; Liu et al., 2025c). We explore a simpler alternative: accelerating diffusion models entirely at the *solver interface*. The solver only consumes the model output at each sampling stage, rather than its internal representations. We therefore retain the original solver grid, query the pretrained model only at sparse anchor stages, and predict the missing solver-facing outputs from previous exact observations. The key question is then: *how much of the latest observed change should be reused?* As shown in Figure 2, small coefficients underuse the observed trajectory, whereas full reuse or further extrapolation quickly increases deviation from the full-NFE reference. This motivates shrinking the latest change rather than predicting a new high-dimensional representation.

Motivated by this observation, we introduce **GeoShrink**, a training-free acceleration method operating entirely at the solver interface. GeoShrink stores only the three most recent exact outputs and predicts along the latest observed change. Its key idea is to adaptively shrink this change using trajectory geometry: stronger turning or longer forecast horizons lead to more conservative reuse, while locally consistent changes over short horizons retain more of the latest direction. This reduces prediction from forecasting a new high-dimensional representation to estimating the reliability of an already observed direction. The same principle also determines where exact queries are placed. Since future turning is unknown, GeoShrink controls prediction exposure through anchor spacing. Under a fixed query budget, minimizing the worst-case exposure yields a geometric anchor schedule. Thus, GeoShrink unifies *how much* observed information to reuse with *how far* to propagate it, without training, learned predictors, intermediate-feature access, or architecture-specific design.

On standard image and video generation, GeoShrink substantially outperforms strong training-free acceleration baselines: at approximately $5\times$ acceleration, it improves PSNR by 3.10 dB on FLUX.1-dev and 5.44 dB on HunyuanVideo over the strongest listed fidelity baselines. Transferring the same solver-interface principle to human motion yields even larger gains, reducing key geometric errors by tens to hundreds of times over direct step reduction at matched NFE budgets. We further extend GeoShrink to 3D, audio, and music generation, demonstrating broad cross-modal applicability across diverse architectures and generative domains. Our contributions are threefold:

i. We formulate training-free acceleration as sparse prediction of solver-facing outputs while preserving the original solver grid, avoiding model-internal feature access.

ii. We derive a parameter-free geometric retention rule and a fixed-budget anchor schedule whose unique continuous solution is a geometric mesh.

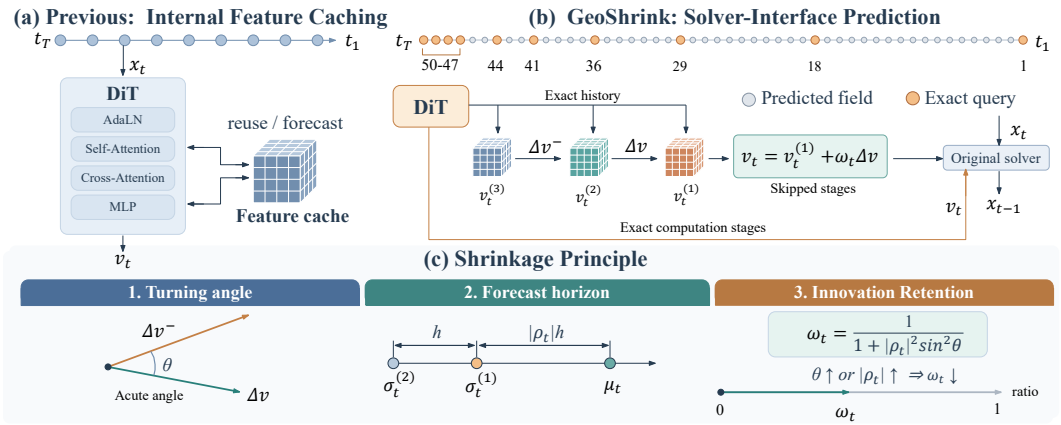

Figure 3: **Overview of GeoShrink.** (a) Previous methods rely on internal feature caching, reusing or forecasting intermediate DiT representations across denoising steps. (b) GeoShrink operates at the solver interface, allocating 10 exact model queries over a 50-step grid through geometric anchoring. At skipped stages, it predicts the solver-facing field from the three latest exact outputs while preserving the original solver and its grid. (c) The retention coefficient $\omega_t = (1+\rho_t^2 \sin^2\theta)^{-1}$ combines the acute angle θ between successive innovations with the normalized midpoint forecast horizon ρ_t. Stronger turning or a longer horizon reduces the retained innovation in $v_t = v_t^{(1)} + \omega_t \Delta v$.

iii. We analyze prediction and error propagation, and evaluate GeoShrink across image, video, motion, audio, music, and 3D generation.

2 Methodology

2.1 Preliminaries

Diffusion transformers (Peebles & Xie, 2023) process noisy latent tokens conditioned on time and auxiliary information. Rectified flow (Liu et al., 2023) learns a velocity field $V_\theta(x_\sigma, \sigma, c)$ with target $\epsilon - x_{\text{data}}$ along $x_\sigma = (1-\sigma)x_{\text{data}} + \sigma\epsilon$, where x_{data} is a clean latent, $\epsilon \sim \mathcal{N}(0, I)$, and c is the condition. Sampling integrates $\mathrm{d}x/\mathrm{d}\sigma = V_\theta(x, \sigma, c)$ from $\sigma = 1$ to $\sigma = 0$. On the grid $\{\sigma_t\}_{t=0}^T$, a solver uses $v_t = V_\theta(x_t, \sigma_t, c)$ to update x_t to x_{t-1}.

2.2 Geometric Retention from Exact Anchors

At a non-anchor stage $t \notin \mathcal{A}$, let $t^{(3)} > t^{(2)} > t^{(1)} > t$ index the three latest exact queries, with outputs $v_t^{(i)} = V_\theta(x_{t^{(i)}}, \sigma_t^{(i)}, c)$, $i \in \{1,2,3\}$, where $\sigma_t^{(i)} = \sigma_{t^{(i)}}$. Define the innovations $\Delta v = v_t^{(1)} - v_t^{(2)}$ and $\Delta v^- = v_t^{(2)} - v_t^{(3)}$. GeoShrink uses the prediction family

$$v_t(\omega) = v_t^{(1)} + \omega \Delta v, \qquad 0 \le \omega \le 1, \tag{1}$$

Let $\theta \in [0, \pi/2]$ be the angle between the innovation lines:

$$\sin^2\theta = \sin^2 \angle(\Delta v, \Delta v^-) = \frac{\|P^\perp_{\Delta v^-} \Delta v\|_2^2}{\|\Delta v\|_2^2} \in [0,1]. \tag{2}$$

With $\mu_t = (\sigma_t + \sigma_{t-1})/2$ and $h = |\sigma_t^{(1)} - \sigma_t^{(2)}|$, define the normalized horizon

$$\rho_t = \frac{\mu_t - \sigma_t^{(1)}}{\sigma_t^{(1)} - \sigma_t^{(2)}} = \frac{\sigma_t + \sigma_{t-1} - 2\sigma_t^{(1)}}{2(\sigma_t^{(1)} - \sigma_t^{(2)})}. \tag{3}$$

The geometric exposure is $R_t = \rho_t^2 \sin^2\theta$.

Proposition 1. *Assume $\Delta v \neq 0$ and $\Delta v^- \neq 0$. Let $u_t = \Delta v/\|\Delta v\|_2$ and $u_t^- = \Delta v^-/\|\Delta v^-\|_2$, choosing the sign of u_t^- such that $\langle u_t, u_t^-\rangle \geq 0$. Define $\xi_t = P_{u_t}^{\perp} u_t^-$, $\eta_t = -\rho_t \xi_t$, and $\widetilde{u}_t = (u_t + \eta_t)/\|u_t + \eta_t\|_2$. The line overlap is*

$$\omega_t = \mathrm{tr}(P_{u_t} P_{\widetilde{u}_t}) = \frac{1}{1 + \rho_t^2 \sin^2\theta} \tag{4}$$

Proof. By construction, $\langle u_t, \xi_t\rangle = 0$ and $\|\xi_t\|_2^2 = \sin^2\theta$. Hence

$$\|\eta_t\|_2^2 = \rho_t^2 \sin^2\theta = R_t, \qquad \|u_t + \eta_t\|_2^2 = 1 + R_t, \qquad \langle u_t, \widetilde{u}_t\rangle^2 = \frac{1}{1+R_t}.$$

Since P_{u_t} and $P_{\widetilde{u}_t}$ are rank-one projectors, $\omega_t = \mathrm{tr}(P_{u_t} P_{\widetilde{u}_t}) = \langle u_t, \widetilde{u}_t\rangle^2$, proving the claim.

Combining Proposition 1 and equation 1 yields

$$v_t = v_t^{(1)} + \frac{v_t^{(1)} - v_t^{(2)}}{1 + \left(\frac{\sigma_t + \sigma_{t-1} - 2\sigma_t^{(1)}}{2(\sigma_t^{(1)} - \sigma_t^{(2)})}\right)^2 \sin^2 \angle\left(v_t^{(1)} - v_t^{(2)}, v_t^{(2)} - v_t^{(3)}\right)} \tag{5}$$

2.3 Budget-Optimal Geometric Anchoring

With future turning unknown, anchor spacing controls exposure. Let $h_j, h_{j+1} > 0$ denote consecutive exact-query gaps in a forward progress coordinate.

Lemma 1. *For skipped updates in the cell following h_j,*

$$0 \leq |\rho_t^{(s)}| \leq \frac{h_{j+1}}{h_j}. \tag{6}$$

If $h_{j+1}/h_j \leq \Gamma$, then

$$R_t^{(s)} \leq \Gamma^2, \qquad \omega_t^{(s)} \geq \frac{1}{1+\Gamma^2}. \tag{7}$$

Proof. The distance from the latest anchor to any point in the next cell is at most h_{j+1}, while the normalization span is h_j, giving $|\rho_t^{(s)}| \leq h_{j+1}/h_j$. Since $0 \leq \sin^2\theta \leq 1$, $R_t^{(s)} = (\rho_t^{(s)})^2 \sin^2\theta \leq \Gamma^2$, and equation 4 gives the retention bound.

With B queries over T stages and bootstrap queries ($h_1 = h_2 = 1$), Lemma 1 motivates

$$\min_{\{h_j\}_{j=1}^{B-1}} \max_{2 \leq j \leq B-2} \frac{h_{j+1}}{h_j} \quad \text{s.t.} \quad h_1 = h_2 = 1, \qquad \sum_{j=1}^{B-1} h_j = T - 1. \tag{8}$$

Theorem 1. *Equation 8 has the unique continuous solution*

$$h_j = r_\star^{\max(j-2,0)}, \qquad j = 1, \ldots, B-1, \qquad \sum_{j=1}^{B-1} r_\star^{\max(j-2,0)} = T - 1, \quad r_\star \geq 1. \tag{9}$$

Proof. Let $\Gamma = \max_{2 \leq j \leq B-2} h_{j+1}/h_j$. Then $h_j \leq \Gamma^{\max(j-2,0)}$, so any feasible schedule satisfies

$$T - 1 = \sum_{j=1}^{B-1} h_j \leq \sum_{j=1}^{B-1} \Gamma^{\max(j-2,0)}.$$

The right-hand side is strictly increasing for $\Gamma \geq 1$, so $\Gamma \geq r_\star$. The geometric sequence attains equality and is therefore optimal. If an optimizer satisfied $h_j < r_\star^{\max(j-2,0)}$ for any j, its total coverage would be strictly less than $T - 1$; hence equality holds for every j, proving uniqueness.

The continuous anchor stages are

$$t_m = T - \sum_{j=1}^{m-1} r_\star^{\max(j-2,0)}, \qquad m = 1, \ldots, B. \tag{10}$$

On the solver grid, apply the guarded order-preserving projection

$$t_m^{\text{disc}} = \Pi_{\text{grid}}\left(T - \sum_{j=1}^{m-1} r_\star^{\max(j-2,0)} \right). \tag{11}$$

Algorithm 1 GeoShrink

Input: Model V_θ, total steps T, solver $\mathcal{D}_t$, condition c, noise x_T, noise schedule $\{\sigma_t\}_{t=0}^{T}$, and budget B
Output: Final sample x_0

for $t = T, T-1, \ldots, 1$ **do**

if $t \in \left\{ \Pi_{\text{grid}}\left(T - \sum_{j=1}^{m-1} r_\star^{\max(j-2,0)} \right) \,\middle|\, m = 1, \ldots, B,\ r_\star \geq 1,\ \sum_{j=1}^{B-1} r_\star^{\max(j-2,0)} = T-1 \right\}$ **then**

$$v_t \leftarrow V_\theta(x_t, \sigma_t, c), \quad t^{(3)} \leftarrow t^{(2)}, \quad t^{(2)} \leftarrow t^{(1)}, \quad t^{(1)} \leftarrow t$$

else

$$v_t \leftarrow v_t^{(1)} + \frac{v_t^{(1)} - v_t^{(2)}}{1 + \left(\frac{\sigma_t + \sigma_{t-1} - 2\sigma_t^{(1)}}{2\left(\sigma_t^{(1)} - \sigma_t^{(2)}\right)} \right)^2 \sin^2\left(\angle\left(v_t^{(1)} - v_t^{(2)}, v_t^{(2)} - v_t^{(3)} \right) \right)}$$

end if

$$x_{t-1} \leftarrow \mathcal{D}_t(x_t, v_t; \sigma_t, \sigma_{t-1})$$

end for

return x_0

3 Empirical Analysis

3.1 Experimental Settings

Models, benchmarks, and baselines. We evaluate SD3.5 Medium (Stability AI, 2024; Esser et al., 2024) and FLUX.1-dev (Black Forest Labs, 2024) on Pick-a-Pic (Kirstain et al., 2023), and HunyuanVideo (Kong et al., 2024) on ChronoMagic-Bench-150 (Yuan et al., 2024) and VBench (Huang et al., 2024b), with additional audiovisual evaluation on MiniMax-H3 (MiniMax, 2026). Cross-domain experiments use HY-Motion (Wen et al., 2025), TRELLIS-text-base (Xiang et al., 2025) on T^3Bench-300 (He et al., 2023), Hunyuan3D-2 (Zhao et al., 2025b) on GSO-100 (Downs et al., 2022), TangoFlux (Hung et al., 2024) on AudioCaps-100 (Kim et al., 2019), and ACE-Step (Gong et al., 2025) on MusicCaps-100 (Agostinelli et al., 2023). Image and video baselines include FORA (Selvaraju et al., 2024), TeaCache (Liu et al., 2025a), ToCa (Zou et al., 2025), TaylorSeer (Liu et al., 2025b), SpeCa (Liu et al., 2025c), and ResilPhase (Zhao et al., 2026).

Evaluation metrics. We report PSNR (Wang & Bovik, 2009), SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018) for visual fidelity, together with ImageReward (Xu et al., 2023), MTScore (Yuan et al., 2024), VBench Score (Huang et al., 2024b), CLIP (Radford et al., 2021), and FVD (Unterthiner et al., 2018). Motion evaluation uses joint, root, velocity, acceleration, and retrieval metrics (Petrovich et al., 2023); 3D evaluation uses Chamfer distance (Fan et al., 2017), Hausdorff distance (Huttenlocher et al., 1993), IoU, and normal error. Audio evaluation uses log-spectral distance (Gray & Markel, 1976), multi-resolution STFT distance (Yamamoto et al., 2020), SI-SDR (Le Roux et al., 2019), and CLAP (Wu et al., 2023), with mel-spectrogram cosine similarity and log-mel L1 for audiovisual evaluation.

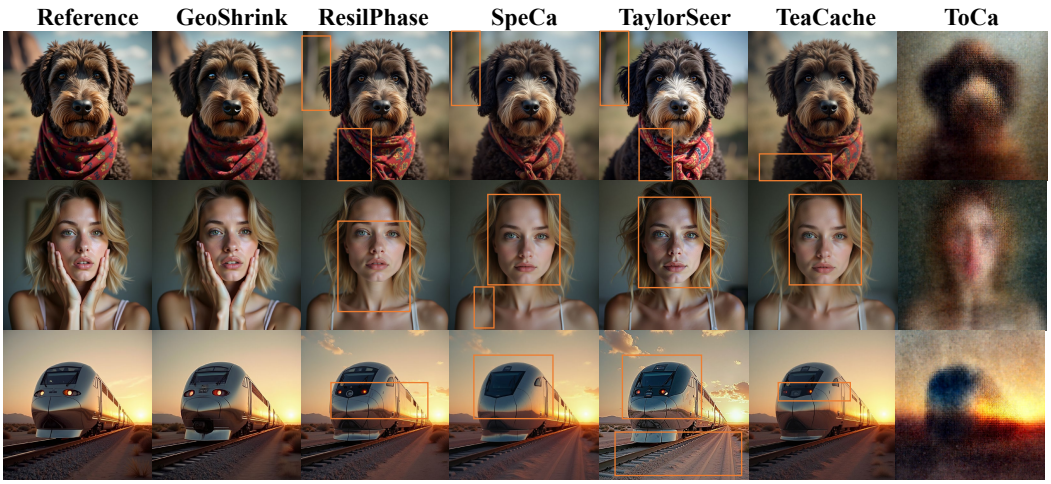

Figure 4: **Qualitative comparison on text-to-image generation.** GeoShrink better preserves the full-sampling reference than competing acceleration methods at comparable speedups.

Table 1: Quantitative comparison of acceleration methods across image and video generation models. For the task-specific metric, we report ImageReward on Pick-a-Pic, MTScore on ChronoMagic-Bench-150, and VBench Score on VBench.

Model	Benchmark	Method	Setting	Speedup	PSNR ↑	SSIM ↑	LPIPS ↓	Task Score ↑
SD3.5 Medium	Pick-a-Pic	SpeCa	$\tau_0 = 7, \beta = 0.3$	2.40×	20.759	0.7743	0.2704	0.7873
		TaylorSeer	$\mathcal{N} = 3, O = 1$	2.49×	18.493	0.7454	0.2918	0.7815
		TeaCache	$l = 0.4$	2.66×	20.872	0.7616	0.2942	0.7671
		ToCa	$\mathcal{N} = 10, R = 90\%$	2.09×	20.431	0.7057	0.3111	0.7662
		ResilPhase	$\mathcal{N} = 3, O = 1$	2.76×	18.483	0.7451	0.2921	0.7813
		GeoShrink	$B = 10$	**2.80**×	**24.744**	**0.8450**	**0.2078**	**0.7924**
FLUX.1-dev	Pick-a-Pic	SpeCa	$\tau_0 = 12, \beta = 0.3$	4.78×	17.4881	0.6768	0.4481	0.9031
		TaylorSeer	$\mathcal{N} = 11, O = 2$	4.65×	15.7641	0.6439	0.4398	0.9181
		TeaCache	$l = 1.4$	4.82×	18.0757	0.6850	0.4123	0.9207
		ToCa	$\mathcal{N} = 3, R = 93\%$	4.86×	14.8189	0.4608	0.7701	0.7310
		ResilPhase	$\mathcal{N} = 6, O = 1$	4.97×	18.5211	0.7083	0.3499	0.9219
		GeoShrink	$B = 10$	**4.98**×	**21.6227**	**0.7903**	**0.2552**	**0.9468**
HunyuanVideo	ChronoMagic-150	FORA	$\mathcal{N} = 6$	4.87×	14.181	0.5812	0.4924	41.4804
		TeaCache	$l = 0.4$	4.62×	17.620	0.6706	0.3492	41.1334
		ToCa	$\mathcal{N} = 10, R = 90\%$	4.86×	13.335	0.5166	0.5840	42.3887
		SpeCa	$\tau_0 = 1.5, \beta = 0.2$	4.65×	14.428	0.5867	0.4720	40.9493
		TaylorSeer	$\mathcal{N} = 7, O = 1$	4.63×	14.303	0.5872	0.4850	41.1623
		ResilPhase	$\mathcal{N} = 6, O = 1$	4.98×	19.881	0.6854	0.2878	42.7770
		GeoShrink	$B = 10$	**4.99**×	**25.317**	**0.8202**	**0.1699**	**42.9833**
HunyuanVideo	VBench	FORA	$\mathcal{N} = 6$	4.87×	15.871	0.6073	0.4584	78.70
		TeaCache	$l = 0.4$	4.62×	17.923	0.6547	0.3760	79.77
		ToCa	$\mathcal{N} = 10, R = 90\%$	4.86×	17.581	0.5857	0.4501	76.37
		SpeCa	$\tau_0 = 1.5, \beta = 0.2$	4.65×	16.461	0.5883	0.4219	79.59
		TaylorSeer	$\mathcal{N} = 7, O = 1$	4.63×	15.520	0.5641	0.4581	79.07
		ResilPhase	$\mathcal{N} = 6, O = 1$	4.98×	18.920	0.6709	0.3341	79.78
		GeoShrink	$B = 10$	**4.99**×	**22.8045**	**0.7925**	**0.1930**	**80.18**

3.2 Image and Video Generation

As shown in Table 1, GeoShrink improves reference fidelity at comparable speedups. On SD3.5 Medium and FLUX.1-dev, it improves PSNR over the strongest listed fidelity baselines by 3.87 dB and 3.10 dB, respectively, while also improving SSIM and LPIPS. On HunyuanVideo, it reaches $4.99\times$ speedup and improves ChronoMagic-Bench-150 PSNR by 5.44 dB. Figures 4 and 5 further show improved visual fidelity, while Table 2 confirms clear gains over direct step reduction on MiniMax-H3 at matched NFE. On MiniMax-H3, improvements extend to both visual and audio metrics across all tested NFE budgets. These results support the applicability of solver-interface prediction to joint audiovisual generation.

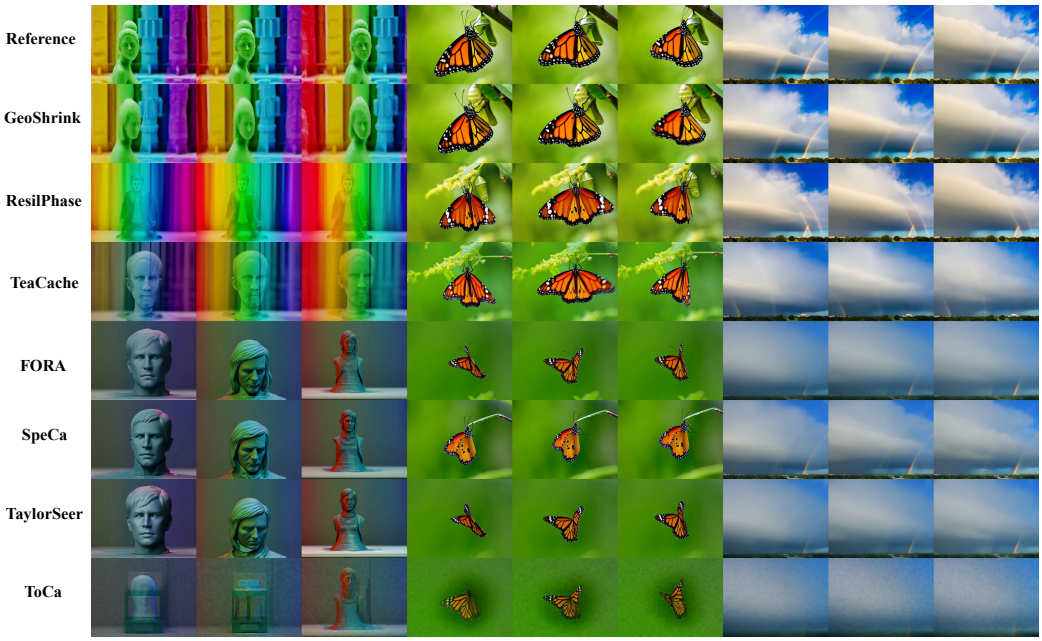

Figure 5: **Qualitative comparison on text-to-video generation.** GeoShrink preserves spatial details and temporal consistency more faithfully under aggressive acceleration.

Table 2: Quantitative comparison between **GeoShrink** and direct step reduction on MiniMax-H3.

NFE	Method	Visual					Audio		
		CLIP ↑	FVD ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Mel Cos ↑	LogMel L1 ↓	SI-SDR ↑
10	Direct	33.7164	166.0071	17.4910	0.6190	0.3632	0.7730	1.1660	-0.2793
	GeoShrink	**33.8711**	**164.5004**	**24.2122**	**0.7883**	**0.1678**	**0.8528**	**0.6998**	**1.5133**
15	Direct	33.6598	166.8756	19.2673	0.6732	0.2914	0.8537	0.8240	3.2408
	GeoShrink	**33.7885**	**166.1741**	**29.1320**	**0.8810**	**0.0818**	**0.9595**	**0.3645**	**7.6285**
20	Direct	33.6435	166.9863	21.2167	0.7250	0.2317	0.9011	0.5850	6.3888
	GeoShrink	**33.6454**	**164.4024**	**33.1394**	**0.9276**	**0.0451**	**0.9878**	**0.2134**	**12.585**

Table 3: Comparison between **GeoShrink** and direct step reduction on the HY-Motion.

NFE	Method	RA-MPJPE ↓	Global MPJPE ↓	RootErr ↓	VelErr ↓	AccErr ↓	uTMR R@3 ↑	Matching ↓
10	Direct	120.618	501.158	465.109	7.325	2.484	0.6260	49.8151
	GeoShrink	**5.303**	**26.009**	**25.024**	**0.694**	**0.301**	**0.7280**	**46.8595**
20	Direct	81.194	360.800	335.515	5.822	1.948	0.7105	47.9181
	GeoShrink	**1.252**	**6.004**	**5.786**	**0.199**	**0.102**	**0.7320**	**46.7981**
30	Direct	51.350	194.661	177.461	4.035	1.363	0.7175	47.2669
	GeoShrink	**0.394**	**2.337**	**2.279**	**0.093**	**0.063**	**0.7300**	**46.8036**
40	Direct	25.686	86.252	77.388	2.219	0.766	0.7155	46.9870
	GeoShrink	**0.100**	**0.799**	**0.789**	**0.046**	**0.047**	**0.7300**	**46.8060**

3.3 Human Motion Generation

As shown in Table 3 and Figure 6, the benefit of retaining the dense solver trajectory is particularly pronounced for human motion. Across all tested NFE budgets, GeoShrink sharply reduces geometric and kinematic errors over direct step reduction while improving retrieval-based semantic metrics. The large gap at matched NFE suggests that motion trajectories are especially sensitive to removing intermediate solver updates.

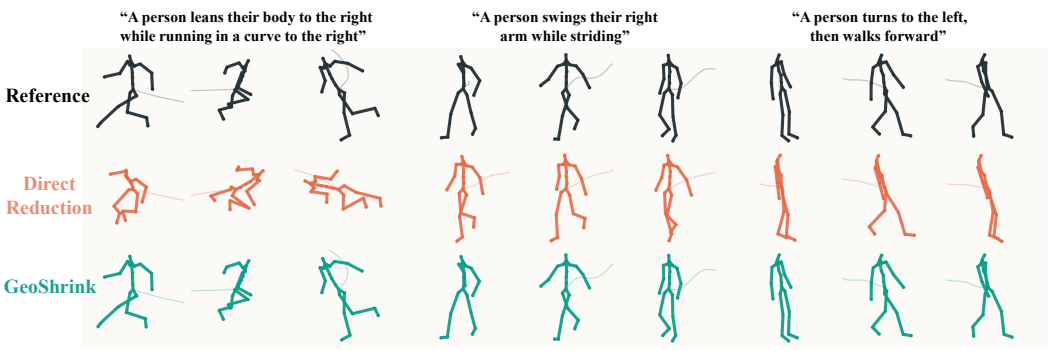

Figure 6: **Motion qualitative comparison.** GeoShrink better preserves full-sampling trajectories.

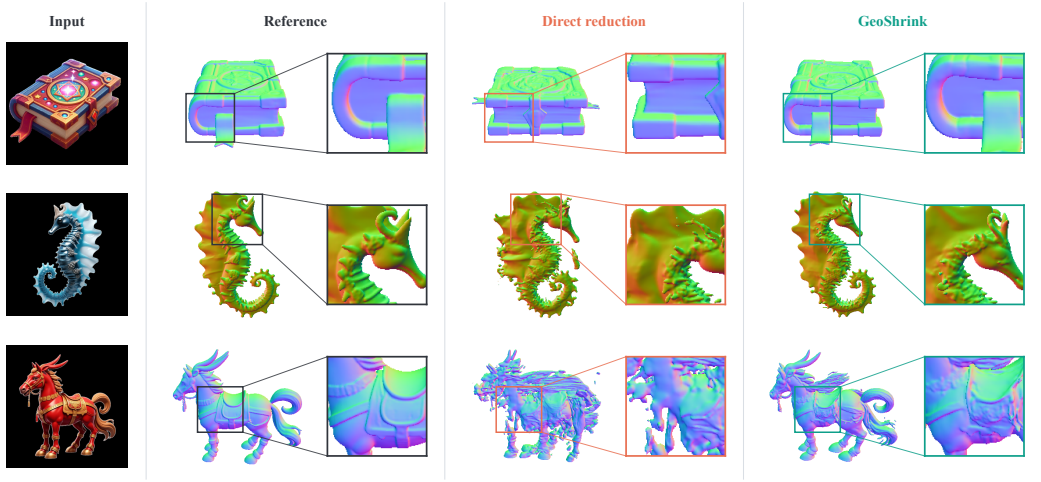

Figure 7: **3D qualitative comparison.** Left to right: input, reference (black), direct reduction (red), and GeoShrink (green). GeoShrink better preserves geometry.

Table 4: Comparison between **GeoShrink** implementations and direct step reduction for 3D generation. We evaluate TRELLIS-text-base on T^{3}Bench-300 and Hunyuan3D-2 on GSO-100.

TRELLIS-text-base / T^{3}Bench-300						Hunyuan3D-2 / GSO-100					
NFE	Method	CLIP ↑	Center CD^2 ↓	LPIPS ↓	Hausdorff ↓	NFE	Method	CLIP ↑	Surface CD^2 ↓	IoU ↑	Normal (°) ↓
10	Direct	26.6521	2.007×10^{-6}	0.009377	0.013083	10	Direct	74.3904	5.411×10^{-3}	0.89199	15.3606
	GeoShrink	**26.7623**	$\mathbf{1.005 \times 10^{-6}}$	**0.004232**	**0.010800**		**GeoShrink**	**74.4081**	$\mathbf{5.270 \times 10^{-3}}$	**0.89790**	**15.0119**
12	Direct	**26.7772**	1.529×10^{-6}	0.006723	0.012179	15	Direct	75.1844	$\mathbf{6.383 \times 10^{-4}}$	**0.95457**	**9.4850**
	GeoShrink	26.7711	$\mathbf{4.042 \times 10^{-7}}$	**0.001528**	**0.008603**		**GeoShrink**	**75.3055**	6.704×10^{-4}	0.95319	9.5296
15	Direct	26.6790	1.176×10^{-6}	0.004662	0.011269	20	Direct	75.2853	$\mathbf{4.383 \times 10^{-4}}$	0.96723	7.6636
	GeoShrink	**26.7704**	$\mathbf{3.845 \times 10^{-7}}$	**0.001484**	**0.008399**		**GeoShrink**	**75.3562**	4.505×10^{-4}	**0.96818**	**7.4691**
18	Direct	26.7687	5.040×10^{-7}	0.001729	0.008954	25	Direct	75.4099	3.750×10^{-4}	0.97453	6.5459
	GeoShrink	**26.7743**	$\mathbf{1.710 \times 10^{-7}}$	**0.000603**	**0.006766**		**GeoShrink**	**75.4237**	$\mathbf{1.796 \times 10^{-4}}$	**0.97631**	**6.1900**
20	Direct	**26.8150**	4.745×10^{-7}	0.001613	0.008815	30	Direct	75.4142	1.380×10^{-4}	0.98147	5.6879
	GeoShrink	26.7753	$\mathbf{1.593 \times 10^{-7}}$	**0.000584**	**0.006574**		**GeoShrink**	**75.4569**	$\mathbf{8.061 \times 10^{-5}}$	**0.98253**	**5.5189**

3.4 3D Generation

Table 4 and Figure 7 show that GeoShrink generally improves reference fidelity over direct step reduction, especially on TRELLIS, where Center Chamfer Distance, LPIPS, and Hausdorff distance are consistently reduced. Results on Hunyuan3D-2 are more heterogeneous, with clear gains at several budgets and small reversals on individual geometry metrics.

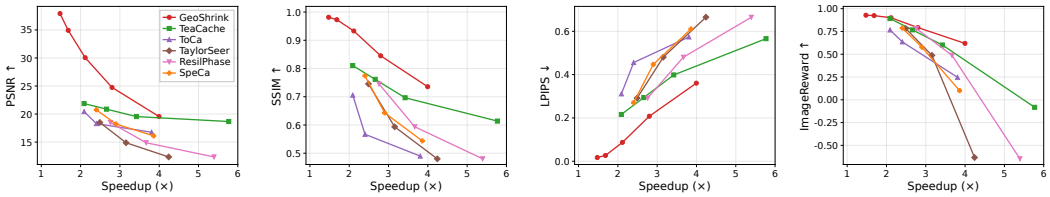

Figure 8: **Speed-quality trade-off.** GeoShrink yields a better Pareto frontier on SD3.5 Medium.

Table 5: Comparison between **GeoShrink** and direct step reduction for audio generation at matched NFE budgets. We evaluate TangoFlux on AudioCaps-100 and ACE-Step on MusicCaps-100.

NFE	Method	TangoFlux / AudioCaps-100				ACE-Step / MusicCaps-100			
		LSD ↓	MS-STFT ↓	SI-SDR ↑	CLAP ↑	LSD ↓	MS-STFT ↓	SI-SDR ↑	CLAP ↑
15	Direct	9.594	0.458	3.83	0.6857	8.524	0.454	**3.66**	0.4906
	GeoShrink	**6.233**	**0.228**	**11.06**	**0.6886**	**8.156**	**0.453**	3.58	**0.4947**
20	Direct	8.006	0.332	6.92	0.6902	7.989	0.395	5.45	0.4902
	GeoShrink	**4.852**	**0.090**	**18.52**	**0.6914**	**7.596**	**0.340**	**6.87**	**0.4955**
25	Direct	6.744	0.245	10.10	0.6922	7.513	**0.309**	**8.25**	0.4887
	GeoShrink	**4.295**	**0.045**	**24.39**	**0.6928**	**7.486**	0.322	7.47	**0.4923**
30	Direct	6.184	0.184	13.48	0.6913	7.333	0.285	9.18	0.4895
	GeoShrink	**3.922**	**0.016**	**33.55**	**0.6931**	**6.945**	**0.229**	**11.30**	**0.4922**

3.5 Audio and Music Generation

As shown in Table 5, GeoShrink transfers effectively to audio generation. On TangoFlux, it consistently improves spectral and waveform fidelity across all tested NFE budgets, with especially large gains in SI-SDR. Results on ACE-Step are more mixed: GeoShrink generally improves LSD and CLAP but does not dominate every spectral metric at every budget, indicating that music generation can be more metric-dependent.

3.6 Ablation Study

Component ablation. Table 6 shows that geometric adaptation and shrink interpolation improve quality on SD3.5 Medium at $4\times$ acceleration. Both components provide individual gains, while their combination performs best across all metrics, confirming their complementary effects.

Table 6: Ablation study of geometric adaptation and innovation shrinkage using SD3.5 Medium.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	ImageReward ↑
Fixed-Linear	9.882	0.404832	0.702306	-1.2556
Geo-Linear	15.457	0.621274	0.414758	0.278629
Fixed-Shrink	19.556	0.692327	0.388319	0.559383
GeoShrink	**19.851**	**0.735519**	**0.360551**	**0.619176**

Budget robustness. As shown in Figure 8, GeoShrink maintains a stronger speed-quality trade-off across the tested acceleration range. At comparable speedups, it achieves higher PSNR and SSIM and lower LPIPS, while maintaining competitive ImageReward scores. Its fidelity advantage remains evident as acceleration increases, suggesting greater robustness to reduced computation budgets. These results show that the gains extend across multiple metrics and operating points, allowing GeoShrink to accommodate different inference budgets while preserving output quality.

4 Conclusion

We introduced **GeoShrink**, a training-free acceleration method that preserves the original solver grid while sparsifying expensive model evaluations. GeoShrink predicts skipped solver-facing outputs through geometry-aware innovation shrinkage and places exact queries using a geometric anchor schedule. Experiments across image, video, motion, audio, music, and 3D generation show that GeoShrink consistently improves fidelity under aggressive acceleration and yields a stronger speed-quality trade-off. Overall, GeoShrink demonstrates that preserving the solver grid can be more effective than directly shortening it.

REFERENCES

Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. MusicLM: Generating Music From Text. *arXiv preprint arXiv:2301.11325*, 2023. URL https://arxiv.org/abs/2301.11325.

Fan Bao, Shen Nie, Kaiwen Xue, Yue Cao, Chongxuan Li, Hang Su, and Jun Zhu. All are Worth Words: A ViT Backbone for Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. URL https://arxiv.org/abs/2209.12152.

Black Forest Labs. FLUX.1 [dev]. Official model card, 2024. URL https://huggingface.co/black-forest-labs/FLUX.1-dev. Accessed September 22, 2026.

Peiliang Cai, Jiacheng Liu, Haowen Xu, Xinyu Wang, Chang Zou, and Linfeng Zhang. LESA: Learnable Stage-Aware Predictors for Diffusion Model Acceleration. *arXiv preprint arXiv:2602.20497*, 2026. URL https://arxiv.org/abs/2602.20497.

Pengtao Chen, Mingzhu Shen, Peng Ye, Jianjian Cao, Chongjun Tu, Christos-Savvas Bouganis, Yiren Zhao, and Tao Chen. Δ-DiT: A Training-Free Acceleration Method Tailored for Diffusion Transformers. *arXiv preprint arXiv:2406.01125*, 2024. URL https://arxiv.org/abs/2406.01125.

Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B. McHugh, and Vincent Vanhoucke. Google Scanned Objects: A High-Quality Dataset of 3D Scanned Household Items. In *IEEE International Conference on Robotics and Automation*, 2022. URL https://arxiv.org/abs/2204.11918.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12606–12633. PMLR, 2024. URL https://proceedings.mlr.press/v235/esser24a.html.

Haoqiang Fan, Hao Su, and Leonidas Guibas. A Point Set Generation Network for 3D Object Reconstruction from a Single Image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. URL https://arxiv.org/abs/1612.00603.

Junmin Gong, Sean Zhao, Sen Wang, Shengyuan Xu, and Joe Guo. ACE-Step: A Step Towards Music Generation Foundation Model. *arXiv preprint arXiv:2506.00045*, 2025. URL https://arxiv.org/abs/2506.00045.

A. Gray and J. Markel. Distance measures for speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(5):380–391, 1976. doi: 10.1109/tassp.1976.1162849. URL https://doi.org/10.1109/tassp.1976.1162849.

Yuze He, Yushi Bai, Matthieu Lin, Wang Zhao, Yubin Hu, Jenny Sheng, Ran Yi, Juanzi Li, and Yong-Jin Liu. T^3Bench: Benchmarking Current Progress in Text-to-3D Generation. *arXiv preprint arXiv:2310.02977*, 2023. URL https://arxiv.org/abs/2310.02977.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, 2020. URL https://arxiv.org/abs/2006.11239.

Yushi Huang, Zining Wang, Ruihao Gong, Jing Liu, Xinjie Zhang, Jinyang Guo, Xianglong Liu, and Jun Zhang. HarmoniCa: Harmonizing Training and Inference for Better Feature Caching in Diffusion Transformer Acceleration. *arXiv preprint arXiv:2410.01723*, 2024a. URL https://arxiv.org/abs/2410.01723.

Ziqi Huang, Yinan He, Jiashuo Yu, Fan Zhang, Chenyang Si, Yuming Jiang, Yuanhan Zhang, Tianxing Wu, Qingyang Jin, Nattapol Chanpaisit, Yaohui Wang, Xinyuan Chen, Limin Wang, Dahua Lin, Yu Qiao, and Ziwei Liu. VBench: Comprehensive Benchmark Suite for Video Generative Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024b. URL https://arxiv.org/abs/2311.17982.

Chia-Yu Hung, Navonil Majumder, Zhifeng Kong, Ambuj Mehrish, Amir Ali Bagherzadeh, Chuan Li, Rafael Valle, Bryan Catanzaro, and Soujanya Poria. TangoFlux: Super Fast and Faithful Text to Audio Generation with Flow Matching and Clap-Ranked Preference Optimization. *arXiv preprint arXiv:2412.21037*, 2024. URL https://arxiv.org/abs/2412.21037.

D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993. doi: 10.1109/34.232073. URL https://doi.org/10.1109/34.232073.

Kumara Kahatapitiya, Haozhe Liu, Sen He, Ding Liu, Menglin Jia, Chenyang Zhang, Michael S. Ryoo, and Tian Xie. Adaptive Caching for Faster Video Generation with Diffusion Transformers. *arXiv preprint arXiv:2411.02397*, 2024. URL https://arxiv.org/abs/2411.02397.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the Design Space of Diffusion-Based Generative Models. In *Advances in Neural Information Processing Systems*, 2022. URL https://arxiv.org/abs/2206.00364.

Chris Dongjoo Kim, Byeongchang Kim, Hyunmin Lee, and Gunhee Kim. AudioCaps: Generating Captions for Audios in The Wild. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 119–132. Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-1011. URL https://aclanthology.org/N19-1011/.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-Pic: An Open Dataset of User Preferences for Text-to-Image Generation. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2305.01569.

Weijie Kong, Qi Tian, Zijian Zhang, Rox Min, Zuozhuo Dai, Jin Zhou, Jiangfeng Xiong, Xin Li, Bo Wu, Jianwei Zhang, Kathrina Wu, Qin Lin, Junkun Yuan, Yanxin Long, Aladdin Wang, Andong Wang, Changlin Li, Duojun Huang, Fang Yang, Hao Tan, Hongmei Wang, Jacob Song, Jiawang Bai, Jianbing Wu, Jinbao Xue, Joey Wang, Kai Wang, Mengyang Liu, Pengyu Li, Shuai Li, Weiyan Wang, Wenqing Yu, Xinchi Deng, Yang Li, Yi Chen, Yutao Cui, Yuanbo Peng, Zhentao Yu, Zhiyu He, Zhiyong Xu, Zixiang Zhou, Zunnan Xu, Yangyu Tao, Qinglin Lu, Songtao Liu, Dax Zhou, Hongfa Wang, Yong Yang, Di Wang, Yuhong Liu, Jie Jiang, and Caesar Zhong. HunyuanVideo: A Systematic Framework For Large Video Generative Models. *arXiv preprint arXiv:2412.03603*, 2024. URL https://arxiv.org/abs/2412.03603.

Jonathan Le Roux, Scott Wisdom, Hakan Erdogan, and John R. Hershey. SDR – Half-baked or Well Done? In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 626–630, 2019. doi: 10.1109/icassp.2019.8683855. URL https://doi.org/10.1109/icassp.2019.8683855.

Senmao Li, Taihang Hu, Joost van de Weijer, Fahad Shahbaz Khan, Tao Liu, Linxuan Li, Shiqi Yang, Yaxing Wang, Ming-Ming Cheng, and Jian Yang. Faster Diffusion: Rethinking the Role of the Encoder for Diffusion Model Inference. *arXiv preprint arXiv:2312.09608*, 2023. URL https://arxiv.org/abs/2312.09608.

Feng Liu, Shiwei Zhang, Xiaofeng Wang, Yujie Wei, Haonan Qiu, Yuzhong Zhao, Yingya Zhang, Qixiang Ye, and Fang Wan. Timestep Embedding Tells: It's Time to Cache for Video Diffusion Model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025a. URL https://arxiv.org/abs/2411.19108.

Jiacheng Liu, Chang Zou, Yuanhuiyi Lyu, Junjie Chen, and Linfeng Zhang. From Reusing to Forecasting: Accelerating Diffusion Models with TaylorSeers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025b. URL https://arxiv.org/abs/2503.06923.

Jiacheng Liu, Chang Zou, Yuanhuiyi Lyu, Fei Ren, Shaobo Wang, Kaixin Li, and Linfeng Zhang. SpeCa: Accelerating Diffusion Transformers with Speculative Feature Caching. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 10024–10033, 2025c. doi: 10.1145/3746027.3755331. URL https://doi.org/10.1145/3746027.3755331.

Joseph Liu, Joshua Geddes, Ziyu Guo, Haomiao Jiang, and Mahesh Kumar Nandwana. SmoothCache: A Universal Inference Acceleration Technique for Diffusion Transformers. *arXiv preprint arXiv:2411.10510*, 2024. URL https://arxiv.org/abs/2411.10510.

Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo Numerical Methods for Diffusion Models on Manifolds. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2202.09778.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.03003.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. In *Advances in Neural Information Processing Systems*, 2022. URL https://arxiv.org/abs/2206.00927.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast Solver for Guided Sampling of Diffusion Probabilistic Models. *Machine Intelligence Research*, 22 (4):730–751, 2025. doi: 10.1007/s11633-025-1562-4. URL https://doi.org/10.1007/s11633-025-1562-4.

Zhengyao Lv, Chenyang Si, Junhao Song, Zhenyu Yang, Yu Qiao, Ziwei Liu, and Kwan-Yee K. Wong. FasterCache: Training-Free Video Diffusion Model Acceleration with High Quality. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.19355.

Xinyin Ma, Gongfan Fang, Michael Bi Mi, and Xinchao Wang. Learning-to-Cache: Accelerating Diffusion Transformer via Layer Caching. *arXiv preprint arXiv:2406.01733*, 2024a. URL https://arxiv.org/abs/2406.01733.

Xinyin Ma, Gongfan Fang, and Xinchao Wang. DeepCache: Accelerating Diffusion Models for Free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15762–15772, 2024b. URL https://openaccess.thecvf.com/content/CVPR2024/html/Ma_DeepCache_Accelerating_Diffusion_Models_for_Free_CVPR_2024_paper.html.

Zehong Ma, Longhui Wei, Feng Wang, Shiliang Zhang, and Qi Tian. MagCache: Fast Video Generation with Magnitude-Aware Cache. *arXiv preprint arXiv:2506.09045*, 2025. URL https://arxiv.org/abs/2506.09045.

MiniMax. MiniMax H3. Official model card, 2026. URL https://huggingface.co/MiniMaxAI/MiniMax-H3. Accessed September 22, 2026.

William Peebles and Saining Xie. Scalable Diffusion Models with Transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4195–4205, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Peebles_Scalable_Diffusion_Models_with_Transformers_ICCV_2023_paper.html.

Mathis Petrovich, Michael J. Black, and Gül Varol. TMR: Text-to-Motion Retrieval Using Contrastive 3D Human Motion Synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. URL https://arxiv.org/abs/2305.00976.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. URL https://arxiv.org/abs/2112.10752.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align Your Steps: Optimizing Sampling Schedules in Diffusion Models. *arXiv preprint arXiv:2404.14507*, 2024. URL https://arxiv.org/abs/2404.14507.

Pratheba Selvaraju, Tianyu Ding, Tianyi Chen, Ilya Zharkov, and Luming Liang. FORA: Fast-Forward Caching in Diffusion Transformer Acceleration. *arXiv preprint arXiv:2407.01425*, 2024. URL https://arxiv.org/abs/2407.01425.

Junhyuk So, Jungwon Lee, and Eunhyeok Park. FRDiff : Feature Reuse for Universal Training-free Acceleration of Diffusion Models. *arXiv preprint arXiv:2312.03517*, 2023. URL https://arxiv.org/abs/2312.03517.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265. PMLR, 2015. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*, 2021a. URL https://arxiv.org/abs/2010.02502.

Yang Song and Stefano Ermon. Generative Modeling by Estimating Gradients of the Data Distribution. In *Advances in Neural Information Processing Systems*, 2019. URL https://arxiv.org/abs/1907.05600.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021b. URL https://arxiv.org/abs/2011.13456.

Stability AI. Stable Diffusion 3.5 Medium. Official model card, 2024. URL https://huggingface.co/stabilityai/stable-diffusion-3.5-medium. Accessed September 22, 2026.

Thomas Unterthiner, Sjoerd van Steenkiste, Karol Kurach, Raphael Marinier, Marcin Michalski, and Sylvain Gelly. Towards Accurate Generative Models of Video: A New Metric & Challenges. *arXiv preprint arXiv:1812.01717*, 2018. URL https://arxiv.org/abs/1812.01717.

Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009. doi: 10.1109/msp.2008.930649. URL https://doi.org/10.1109/msp.2008.930649.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. doi: 10.1109/tip.2003.819861. URL https://doi.org/10.1109/tip.2003.819861.

Yuxin Wen, Qing Shuai, Di Kang, Jing Li, Cheng Wen, Yue Qian, Ningxin Jiao, Changhai Chen, Weijie Chen, Yiran Wang, Jinkun Guo, Dongyue An, Han Liu, Yanyu Tong, Chao Zhang, Qing Guo, Juan Chen, Qiao Zhang, Youyi Zhang, Zihao Yao, Cheng Zhang, Hong Duan, Xiaoping Wu, Qi Chen, Fei Cheng, Liang Dong, Peng He, Hao Zhang, Jiaxin Lin, Chao Zhang, Zhongyi Fan, Yifan Li, Zhichao Hu, Yuhong Liu, Linus, Jie Jiang, Xiaolong Li, and Linchao Bao. HY-Motion 1.0: Scaling Flow Matching Models for Text-To-Motion Generation. *arXiv preprint arXiv:2512.23464*, 2025. URL https://arxiv.org/abs/2512.23464.

Yusong Wu, Ke Chen, Tianyu Zhang, Yuchen Hui, Marianna Nezhurina, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. Large-scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2023. URL https://arxiv.org/abs/2211.06687.

Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3D Latents for Scalable and Versatile 3D Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21469–21480, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Xiang_Structured_3D_Latents_for_Scalable_and_Versatile_3D_Generation_CVPR_2025_paper.html.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2304.05977.

Shuchen Xue, Mingyang Yi, Weijian Luo, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, and Zhi-Ming Ma. SA-Solver: Stochastic Adams Solver for Fast Sampling of Diffusion Models. *arXiv preprint arXiv:2309.05019*, 2023. URL https://arxiv.org/abs/2309.05019.

Ryuichi Yamamoto, Eunwoo Song, and Jae-Min Kim. Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6199–6203, 2020. doi: 10.1109/icassp40776.2020.9053795. URL https://doi.org/10.1109/icassp40776.2020.9053795.

Shenghai Yuan, Jinfa Huang, Yongqi Xu, Yaoyang Liu, Shaofeng Zhang, Yujun Shi, Ruijie Zhu, Xinhua Cheng, Jiebo Luo, and Li Yuan. ChronoMagic-Bench: A Benchmark for Metamorphic Evaluation of Text-to-Time-lapse Video Generation. In *Advances in Neural Information Processing Systems*, 2024. URL https://arxiv.org/abs/2406.18522.

Qinsheng Zhang and Yongxin Chen. Fast Sampling of Diffusion Models with Exponential Integrator. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2204.13902.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. URL https://arxiv.org/abs/1801.03924.

Qicheng Zhao, Yu Li, Qi Sun, and Zheyu Yan. ResilPhase: Plug-and-Play Phase Mapping and Noise-Resilient Macro-Trajectory Extrapolation for Diffusion Acceleration. *arXiv preprint arXiv:2606.26769*, 2026. URL https://arxiv.org/abs/2606.26769.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A Unified Predictor-Corrector Framework for Fast Sampling of Diffusion Models. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2302.04867.

Xuanlei Zhao, Xiaolong Jin, Kai Wang, and Yang You. Real-Time Video Generation with Pyramid Attention Broadcast. In *International Conference on Learning Representations*, 2025a. URL https://arxiv.org/abs/2408.12588.

Zibo Zhao, Zeqiang Lai, Qingxiang Lin, Yunfei Zhao, Haolin Liu, Shuhui Yang, Yifei Feng, Mingxin Yang, Sheng Zhang, Xianghui Yang, Huiwen Shi, Sicong Liu, Junta Wu, Yihang Lian, Fan Yang, Ruining Tang, Zebin He, Xinzhou Wang, Jian Liu, Xuhui Zuo, Zhuo Chen, Biwen Lei, Haohan Weng, Jing Xu, Yiling Zhu, Xinhai Liu, Lixin Xu, Changrong Hu, Shaoxiong Yang, Song Zhang, Yang Liu, Tianyu Huang, Lifu Wang, Jihong Zhang, Meng Chen, Liang Dong, Yiwen Jia, Yulin Cai, Jiaao Yu, Yixuan Tang, Hao Zhang, Zheng Ye, Peng He, Runzhou Wu, Chao Zhang, Yonghao Tan, Jie Xiao, Yangyu Tao, Jianchen Zhu, Jinbao Xue, Kai Liu, Chongqing Zhao, Xinming Wu, Zhichao Hu, Lei Qin, Jianbing Peng, Zhan Li, Minghui Chen, Xipeng Zhang, Lin Niu, Paige Wang, Yingkai Wang, Haozhao Kuang, Zhongyi Fan, Xu Zheng, Weihao Zhuang, YingPing

He, Tian Liu, Yong Yang, Di Wang, Yuhong Liu, Jie Jiang, Jingwei Huang, and Chunchao Guo. Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation. *arXiv preprint arXiv:2501.12202*, 2025b. URL https://arxiv.org/abs/2501.12202.

Kaiwen Zheng, Cheng Lu, Jianfei Chen, and Jun Zhu. DPM-Solver-v3: Improved Diffusion ODE Solver with Empirical Model Statistics. *arXiv preprint arXiv:2310.13268*, 2023. URL https://arxiv.org/abs/2310.13268.

Xin Zhou, Dingkang Liang, Kaijin Chen, Tianrui Feng, Xiwu Chen, Hongkai Lin, Yikang Ding, Feiyang Tan, Hengshuang Zhao, and Xiang Bai. Less is Enough: Training-Free Video Diffusion Acceleration via Runtime-Adaptive Caching. *arXiv preprint arXiv:2507.02860*, 2025. URL https://arxiv.org/abs/2507.02860.

Chang Zou, Shikang Zheng, Evelyn Zhang, Runlin Guo, Haohang Xu, Zhengyi Shi, Conghui He, Xuming Hu, and Linfeng Zhang. Rethinking Token-wise Feature Caching: Accelerating Diffusion Transformers with Dual Feature Caching. *arXiv preprint arXiv:2412.18911*, 2024. URL https://arxiv.org/abs/2412.18911.

Chang Zou, Xuyang Liu, Ting Liu, Siteng Huang, and Linfeng Zhang. Accelerating Diffusion Transformers with Token-wise Feature Caching. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.05317.

A Related Works

Diffusion acceleration has been explored through improved numerical solvers, temporal reuse, computation reduction, parallel execution, and learned few-step generation.

Fast numerical sampling. Early work reduces the number of denoising evaluations by designing more accurate discretizations of diffusion ODEs or SDEs. DDIM (Song et al., 2021a) enables non-Markovian sampling with fewer denoising steps, while PNDM (Liu et al., 2022), DEIS (Zhang & Chen, 2023), DPM-Solver (Lu et al., 2022), DPM-Solver++ (Lu et al., 2025), UniPC (Zhao et al., 2023), DPM-Solver-v3 (Zheng et al., 2023), and SA-Solver (Xue et al., 2023) improve low-budget sampling through higher-order or structure-aware integration. EDM (Karras et al., 2022) jointly studies parameterization and discretization, while Align Your Steps (Sabour et al., 2024) directly optimizes the sampling schedule for a fixed model and solver.

Temporal reuse and feature caching. A large body of work exploits the temporal redundancy of intermediate diffusion representations. DeepCache (Ma et al., 2024b), Faster Diffusion (Li et al., 2023), and FRDiff (So et al., 2023) reuse slowly varying U-Net features across timesteps. For diffusion transformers, Learning-to-Cache (Ma et al., 2024a), HarmoniCa (Huang et al., 2024a), Δ-DiT (Chen et al., 2024), FORA (Selvaraju et al., 2024), Pyramid Attention Broadcast (Zhao et al., 2025a), FasterCache (Lv et al., 2025), TeaCache (Liu et al., 2025a), ToCa (Zou et al., 2025), AdaCache (Kahatapitiya et al., 2024), SmoothCache (Liu et al., 2024), DuCa (Zou et al., 2024), EasyCache (Zhou et al., 2025), and MagCache (Ma et al., 2025) develop increasingly adaptive policies for deciding what and when to reuse. More recently, TaylorSeer (Liu et al., 2025b), SpeCa (Liu et al., 2025c), ResilPhase (Zhao et al., 2026), and LESA (Cai et al., 2026) go beyond direct caching by forecasting future intermediate representations from historical features.

B Implementation details

All experiments are conducted on NVIDIA A100 GPUs, with each experiment repeated using five random seeds. We report sampling latency, excluding encoding and decoding. Speedups are computed relative to the corresponding full-NFE reference using the same timing protocol. For each example and seed, all methods and the corresponding reference use identical inputs, conditioning, and initial noise, with all other inference settings held fixed except for the acceleration configuration.

We evaluate SD3.5 Medium and FLUX.1-dev on 100 prompts from the Pick-a-Pic test set, using 28-step and 50-step reference grids, respectively. Both models use an image resolution of 1024×1024,

with guidance scales of 4.5 and 3.5, respectively. HunyuanVideo is evaluated on ChronoMagic-Bench-150 and the full VBench prompt set using a 50-step reference grid. It produces 65-frame videos at 640×480 resolution and 24 fps, with a guidance scale of 5.0.

For MiniMax-H3, we evaluate all 150 prompts from ChronoMagic-Bench-150 using a 30-step reference grid and evaluation budgets of $B \in \{10, 15, 20\}$. Videos contain 124 frames at 960×544 resolution and 24 fps. Inference uses BF16 precision and automatic offloading with a 12 GB memory reserve.

For HY-Motion-1.0, we select 100 examples through stratified sampling from the official 1,983-example SSAE set. We generate five-second motions with the Euler solver, a guidance scale of 5.0, and a 50-step reference grid, evaluating budgets of $B \in \{10, 20, 30, 40\}$. Prompt rewriting and duration estimation are disabled.

For TRELLIS-text-base, we evaluate all 300 examples from T³Bench, comprising 100 examples per category. The sparse-structure stage remains fixed at 25 steps, while the SLAT stage uses a 25-step reference grid. We set the guidance scale to 7.5, the guidance interval to $[0.5, 0.95]$, and the time-rescaling parameter to 3.0. For Hunyuan3D-2, we evaluate shape generation on 100 objects selected from GSO by a fixed hash-based rule. Inference uses FP16 precision, a 50-step reference grid, a guidance scale of 5.0, a latent shape of 3072×64, and a mesh extraction resolution of 256.

For TangoFlux and ACE-Step, we select 100 examples at fixed index intervals from the AudioCaps test split and MusicCaps, respectively. Both models generate ten-second audio using a 40-step reference grid. TangoFlux uses a sampling rate of 44.1 kHz and a guidance scale of 4.5. ACE-Step uses 48 kHz, BF16 precision, a guidance scale of 7.0, and a shift parameter of 3.0, with empty lyrics in text-to-music mode.

C Proofs and Additional Analysis

In the appendix, q indexes the latest exact query, a_q is its solver stage, and n is the current skipped stage. Write $v_n^{(i)} = V_\theta(\widetilde{x}_{a_{q-i+1}}, \sigma_n^{(i)}, c)$ and $\sigma_n^{(i)} = \sigma_{a_{q-i+1}}$ for $i \in \{1, 2, 3\}$. The anchor-indexed innovations are $\Delta v_q = v_n^{(1)} - v_n^{(2)}$ and $\Delta v_q^- = v_n^{(2)} - v_n^{(3)}$; $\theta_q \in [0, \pi/2]$ is the acute angle between their lines, and $\omega_{q,n}$ denotes the corresponding retention coefficient.

C.1 Directional Prediction Loss and the Benefit of Shrinkage

Proposition 2 (Directional prediction loss)**.** *For a skipped stage, define the unavailable target innovation at the same current state by*

$$I_{q,n} = V_\theta(\widetilde{x}_n, \sigma_n, c) - v_n^{(1)}.$$

Suppose $A_q = \|\Delta v_q\|_2^2 > 0$*, and define*

$$c_{q,n} = \frac{\langle \Delta v_q, I_{q,n} \rangle}{A_q}. \tag{12}$$

For every real ω*,*

$$\|\omega \Delta v_q - I_{q,n}\|_2^2 = \left\| P_{\Delta v_q}^\perp I_{q,n} \right\|_2^2 + A_q(\omega - c_{q,n})^2. \tag{13}$$

Consequently, the optimal coefficient in equation 1 is $\omega_{q,n}^\star = \mathrm{clip}_{[0,1]}(c_{q,n})$*. For* $0 < \omega < 1$*, the prediction strictly improves on both* $\omega = 0$ *and* $\omega = 1$ *if and only if*

$$\frac{\omega}{2} < c_{q,n} < \frac{1+\omega}{2}. \tag{14}$$

Proof of Proposition 2. Fix an anchor-target pair and abbreviate $\Delta v = \Delta v_q$, $I = I_{q,n}$, $A = \|\Delta v\|_2^2 > 0$, and $c = \langle \Delta v, I \rangle / A$. Orthogonal decomposition gives

$$I = c\Delta v + P_{\Delta v}^\perp I.$$

Therefore

$$\omega \Delta v - I = (\omega - c)\Delta v - P_{\Delta v}^\perp I.$$

The two terms are orthogonal, so the Pythagorean theorem gives equation 13. The only term depending on ω is $A(\omega - c)^2$. Its minimizer over $[0, 1]$ is the Euclidean projection $\mathrm{clip}_{[0,1]}(c)$.

Let $\mathcal{L}(\omega) = \|\omega\Delta v - I\|_2^2$. Direct subtraction yields

$$\mathcal{L}(\omega) - \mathcal{L}(0) = A\,\omega(\omega - 2c), \tag{15}$$

$$\mathcal{L}(\omega) - \mathcal{L}(1) = A\,(\omega - 1)(\omega + 1 - 2c). \tag{16}$$

For $0 < \omega < 1$, the first difference is negative precisely when $c > \omega/2$, and the second is negative precisely when $c < (1 + \omega)/2$. Their intersection is equation 14.

If $\Delta v_q = 0$, all members of equation 1 coincide with $v_n^{(1)}$. No coefficient needs to be identified, and equation 12 is left undefined. The implemented predictor automatically returns $v_n^{(1)}$ in this case.

C.2 Conditional Calibration of a Scalar Exposure

The scalar R need not contain all useful information in the query history. The following statement specifies the optimum only within the class of retention rules that are functions of R.

Proposition 3 (Best retention rule based on one scalar)**.** *Consider a distribution of anchor-target pairs, with $\mathbb{E}\,\|\Delta v\|_2^2 < \infty$ and $\mathbb{E}\,\|I\|_2^2 < \infty$. Let R be any measurable scalar statistic. Where $a(r) = \mathbb{E}[\|\Delta v\|_2^2 \mid R = r] > 0$, the minimizer over measurable $f : \mathbb{R} \to [0, 1]$ of $\mathbb{E}\,\|f(R)\Delta v - I\|_2^2$ is, almost surely,*

$$f^\star(r) = \mathrm{clip}_{[0,1]}\left(\frac{\mathbb{E}[\langle \Delta v, I\rangle \mid R = r]}{\mathbb{E}[\|\Delta v\|_2^2 \mid R = r]}\right). \tag{17}$$

On a conditional set with $a(r) = 0$, every coefficient has the same loss.

Proof. Let

$$b(r) = \mathbb{E}[\langle \Delta v, I\rangle \mid R = r], \qquad d(r) = \mathbb{E}[\|I\|_2^2 \mid R = r].$$

For a deterministic coefficient ω, expand the squared norm before taking the conditional expectation:

$$\begin{aligned}\|\omega\Delta v - I\|_2^2 &= \langle \omega\Delta v - I, \omega\Delta v - I\rangle \\ &= \omega^2\,\|\Delta v\|_2^2 - 2\omega\,\langle \Delta v, I\rangle + \|I\|_2^2\,.\end{aligned}$$

Linearity of conditional expectation gives

$$\mathbb{E}[\|\omega\Delta v - I\|_2^2 \mid R = r] = a(r)\omega^2 - 2b(r)\omega + d(r). \tag{18}$$

If $a(r) > 0$, complete the square:

$$\begin{aligned}a(r)\omega^2 - 2b(r)\omega + d(r) &= a(r)\left[\omega^2 - 2\frac{b(r)}{a(r)}\omega\right] + d(r) \\ &= a(r)\left(\omega - \frac{b(r)}{a(r)}\right)^2 + d(r) - \frac{b(r)^2}{a(r)}.\end{aligned}$$

The last two terms do not depend on ω. The closest point in the closed interval $[0, 1]$ to the unconstrained minimizer $b(r)/a(r)$ is its Euclidean projection $\mathrm{clip}_{[0,1]}(b(r)/a(r))$, which proves equation 17 on $\{a(r) > 0\}$.

To pass from conditional to global risk, let f be any measurable rule with values in $[0, 1]$, and let $f^\star$ be the conditional minimizer. The pointwise conditional inequality gives, almost surely,

$$\mathbb{E}[\|f^\star(R)\Delta v - I\|_2^2 \mid R] \le \mathbb{E}[\|f(R)\Delta v - I\|_2^2 \mid R].$$

Taking expectations of both sides and using the tower property gives

$$\begin{aligned}\mathbb{E}\,\|f^\star(R)\Delta v - I\|_2^2 &= \mathbb{E}\Big[\mathbb{E}[\|f^\star(R)\Delta v - I\|_2^2 \mid R]\Big] \\ &\le \mathbb{E}\Big[\mathbb{E}[\|f(R)\Delta v - I\|_2^2 \mid R]\Big] \\ &= \mathbb{E}\,\|f(R)\Delta v - I\|_2^2\,.\end{aligned}$$

If $a(r) = 0$, then the nonnegative random variable $\|\Delta v\|_2^2$ has conditional expectation zero. It is therefore zero conditionally almost surely. Both $\omega^2 \|\Delta v\|_2^2$ and $\omega \langle \Delta v, I \rangle$ vanish, so equation 18 reduces to $d(r)$, independently of ω.

If $A = \|\Delta v\|_2^2 > 0$ and $c = \langle \Delta v, I \rangle / A$, the unclipped optimum is $\mathbb{E}[Ac \mid R]/\mathbb{E}[A \mid R]$. Thus calibration for ordinary field MSE is innovation-energy weighted; it is generally not the unweighted conditional average of c. An unweighted average instead corresponds to a per-sample loss normalized by A, under the integrability conditions needed for that loss.

C.3 Why Smoothness Cannot Determine the Retention Coefficient

Proposition 4 (Smoothness alone does not identify retention)**.** *Fix three distinct anchor times and their vector-valued outputs, with $\Delta v_q \neq 0$, and fix a distinct target time. For every $c_0 \in \mathbb{R}$, there exists a C^∞ model-output trace that matches all three anchor outputs and whose target coefficient in equation 12 equals c_0. The trace can be realized by a smooth time-dependent flow field that is independent of the state.*

Proof of Proposition 4. Let the three anchor times be t_0, t_1, t_2, let the target time be t_3, and let the corresponding fixed anchor outputs be v_0, v_1, v_2. All four times are distinct. Define

$$\Delta v = v_2 - v_1 \neq 0$$

and fix an arbitrary desired coefficient $c_0 \in \mathbb{R}$. Choose the target output

$$v_3 = v_2 + c_0 \Delta v.$$

For $i \in \{0, 1, 2, 3\}$, define the scalar Lagrange basis polynomial

$$L_i(t) = \prod_{\substack{0 \leq j \leq 3 \\ j \neq i}} \frac{t - t_j}{t_i - t_j}.$$

Every denominator is nonzero because the times are distinct. Moreover, $L_i(t_j) = 1$ when $i = j$ and $L_i(t_j) = 0$ when $i \neq j$. The vector-valued polynomial

$$g_{c_0}(t) = \sum_{i=0}^{3} v_i L_i(t) \tag{19}$$

therefore satisfies

$$g_{c_0}(t_i) = v_i, \qquad i = 0, 1, 2, 3.$$

It is a polynomial of degree at most three in each coordinate, hence it is C^∞. At the target,

$$I = g_{c_0}(t_3) - g_{c_0}(t_2) = v_3 - v_2 = c_0 \Delta v.$$

Substitution into equation 12 gives

$$c = \frac{\langle \Delta v, c_0 \Delta v \rangle}{\|\Delta v\|_2^2} = c_0.$$

The three anchor outputs and all statistics computed from them, including $\sin^2 \theta$, are identical for every c_0; the times and hence ρ are also identical. Finally, define a state-independent time-dependent field

$$V_{c_0}(x, t) = g_{c_0}(t).$$

This field is C^∞ in (x, t), and its output along every state trace is exactly $g_{c_0}(t)$. Thus the construction lies inside the class of smooth flow fields and completes the proof.

The preceding construction can be strengthened: even revealing any fixed finite number of derivatives at every anchor does not identify the target coefficient.

Proposition 5 (Finite anchor jets do not identify retention). *Fix an integer $m \geq 0$, distinct anchor times t_0, t_1, t_2, a distinct target time t_3, and a C^∞ trace g_0 with $\Delta v = g_0(t_2) - g_0(t_1) \neq 0$. There is a one-parameter family of C^∞ traces g_α such that*

$$g_\alpha^{(k)}(t_j) = g_0^{(k)}(t_j), \qquad j \in \{0, 1, 2\}, \quad 0 \leq k \leq m, \tag{20}$$

while the oracle coefficient at t_3 ranges over all of $\mathbb{R}$. Each trace is realizable by a state-independent smooth flow field.

Proof. Define the scalar polynomial

$$\phi(t) = \prod_{j=0}^{2} (t - t_j)^{m+1} \tag{21}$$

and the vector-valued trace

$$g_\alpha(t) = g_0(t) + \alpha\phi(t)\Delta v, \qquad \alpha \in \mathbb{R}. \tag{22}$$

Fix an anchor t_j. Factoring the corresponding term gives

$$\phi(t) = (t - t_j)^{m+1}\psi_j(t), \qquad \psi_j(t) = \prod_{\substack{0 \leq i \leq 2 \\ i \neq j}} (t - t_i)^{m+1}.$$

For $0 \leq k \leq m$, Leibniz' rule gives

$$\phi^{(k)}(t) = \sum_{\ell=0}^{k} \binom{k}{\ell} \frac{d^\ell}{dt^\ell}(t - t_j)^{m+1} \psi_j^{(k-\ell)}(t).$$

Every ℓ in this sum satisfies $\ell \leq k \leq m$, and hence

$$\frac{d^\ell}{dt^\ell}(t - t_j)^{m+1} = \frac{(m+1)!}{(m+1-\ell)!}(t - t_j)^{m+1-\ell}$$

still contains a positive power of $t - t_j$. Therefore $\phi^{(k)}(t_j) = 0$. Differentiating equation 22 now proves equation 20 term by term.

In particular, $g_\alpha(t_j) = g_0(t_j)$ at all three anchors, so the latest innovation remains Δv. Since $\phi(t_2) = 0$, the target innovation is

$$\begin{aligned} I_\alpha &= g_\alpha(t_3) - g_\alpha(t_2) \\ &= g_0(t_3) - g_0(t_2) + \alpha\phi(t_3)\Delta v. \end{aligned}$$

Let $c_0 = \langle \Delta v, g_0(t_3) - g_0(t_2) \rangle \,/\, \|\Delta v\|_2^2$. Then

$$\begin{aligned} c_\alpha &= \frac{\langle \Delta v, I_\alpha \rangle}{\|\Delta v\|_2^2} \\ &= c_0 + \alpha\phi(t_3). \end{aligned}$$

The target differs from every anchor, so every factor in $\phi(t_3)$ is nonzero and $\phi(t_3) \neq 0$. As α ranges over $\mathbb{R}$, so does c_α. Finally, $V_\alpha(x, t) = g_\alpha(t)$ is a state-independent C^∞ field realizing the trace.

The interpolation proofs are global and exact. A local expansion shows more specifically why projective turning and the oracle coefficient contain different differential information.

Proposition 6 (Tangential-normal separation). *Let $g \in C^3$, fix a time t with $v = g'(t) \neq 0$, and put $a = g''(t)$. Use equally spaced anchor times $t - 2h, t - h, t$, and a target time $t + \lambda h$ with fixed $\lambda > 0$. Let c_h be the oracle coefficient and $\sin^2 \theta_h$ the projective statistic. Then*

$$c_h = \lambda + \frac{\lambda(1+\lambda)}{2} h \frac{\langle v, a \rangle}{\|v\|_2^2} + O(h^2), \tag{23}$$

$$\sin^2 \theta_h = h^2 \frac{\left\|P_v^\perp a\right\|_2^2}{\|v\|_2^2} + O(h^3). \tag{24}$$

Thus the first correction to c_h depends on tangential log-speed $\langle v, a \rangle \,/\, \|v\|_2^2$, whereas $\sin^2 \theta_h$ measures normal turning.

Proof. Taylor expansion about t gives

$$g(t-h) = g(t) - hv + \frac{h^2}{2}a + O(h^3),$$
$$g(t-2h) = g(t) - 2hv + 2h^2 a + O(h^3),$$
$$g(t+\lambda h) = g(t) + \lambda hv + \frac{\lambda^2 h^2}{2}a + O(h^3).$$

Therefore the latest innovation, previous innovation, and target innovation are

$$\Delta v_h = g(t) - g(t-h) = hv - \frac{h^2}{2}a + O(h^3), \tag{25}$$

$$\Delta v_h^- = g(t-h) - g(t-2h) = hv - \frac{3h^2}{2}a + O(h^3), \tag{26}$$

$$I_h = g(t+\lambda h) - g(t) = \lambda hv + \frac{\lambda^2 h^2}{2}a + O(h^3). \tag{27}$$

Using bilinearity of the inner product in equation 25 and equation 27,

$$\langle \Delta v_h, I_h \rangle = \lambda h^2 \|v\|_2^2 + \frac{\lambda(\lambda-1)}{2} h^3 \langle v, a \rangle + O(h^4),$$
$$\|\Delta v_h\|_2^2 = h^2 \|v\|_2^2 - h^3 \langle v, a \rangle + O(h^4).$$

After cancelling h^2, use $(A+hB)^{-1} = A^{-1} - hB/A^2 + O(h^2)$ with $A = \|v\|_2^2$ to obtain

$$\begin{aligned} c_h &= \frac{\lambda \|v\|_2^2 + \frac{\lambda(\lambda-1)}{2} h \langle v, a \rangle + O(h^2)}{\|v\|_2^2 - h \langle v, a \rangle + O(h^2)} \\ &= \lambda + \left[\frac{\lambda(\lambda-1)}{2} + \lambda \right] h \frac{\langle v, a \rangle}{\|v\|_2^2} + O(h^2) \\ &= \lambda + \frac{\lambda(1+\lambda)}{2} h \frac{\langle v, a \rangle}{\|v\|_2^2} + O(h^2), \end{aligned}$$

which proves equation 23.

It remains to expand the angle. Let $u = v/\|v\|_2$. For any fixed vector b, direct expansion of the norm shows

$$\frac{v + hb + O(h^2)}{\|v + hb + O(h^2)\|_2} = u + \frac{h}{\|v\|_2} P_v^\perp b + O(h^2). \tag{28}$$

Indeed,

$$\left\| v + hb + O(h^2) \right\|_2 = \|v\|_2 + h \langle u, b \rangle + O(h^2),$$

and substituting the reciprocal of this scalar yields $b - u \langle u, b \rangle = P_v^\perp b$ in the first-order term. Divide equation 25 and equation 26 by h and apply equation 28. Their unit directions obey

$$u_{\Delta v} = u - \frac{h}{2\|v\|_2} P_v^\perp a + O(h^2),$$
$$u_{\Delta v^-} = u - \frac{3h}{2\|v\|_2} P_v^\perp a + O(h^2).$$

Hence

$$u_{\Delta v} - u_{\Delta v^-} = \frac{h}{\|v\|_2} P_v^\perp a + O(h^2).$$

Because $P^\perp_{u_{\Delta v^-}} u_{\Delta v^-} = 0$ and $u_{\Delta v^-} = u + O(h)$,

$$\begin{aligned} \sin\theta_h &= \left\| P^\perp_{u_{\Delta v^-}} u_{\Delta v} \right\|_2 \\ &= \left\| P^\perp_{u_{\Delta v^-}} (u_{\Delta v} - u_{\Delta v^-}) \right\|_2 \\ &= \frac{h}{\|v\|_2} \left\| P_v^\perp a \right\|_2 + O(h^2). \end{aligned}$$

Squaring this identity proves equation 24.

Even when tangential speed is constant, the observable closure is not the oracle coefficient of every smooth curved trace. The following exact example keeps this distinction visible.

Proposition 7 (Constant-speed circular sanity check)**.** *Let $g(t) = (\cos(\Omega t), \sin(\Omega t))$, use anchors at $-2h, -h, 0$, and predict the target at h. Assume $\sin(\Omega h/2) \neq 0$. Then*

$$\sin^2 \theta_h = \sin^2(\Omega h), \qquad c_h = \cos(\Omega h). \tag{29}$$

At unit normalized horizon, GeoShrink instead uses

$$\omega_h = \frac{1}{1 + \sin^2(\Omega h)}. \tag{30}$$

Consequently,

$$\begin{aligned} c_h &= 1 - \frac{\Omega^2 h^2}{2} + O(h^4), \\ \omega_h &= 1 - \Omega^2 h^2 + O(h^4). \end{aligned} \tag{31}$$

so the two coefficients already differ at order h^2.

Proof. For arbitrary u, the angle-addition identities give

$$\begin{aligned} g(u+h) - g(u) &= \big(\cos(\Omega(u+h)) - \cos(\Omega u), \sin(\Omega(u+h)) - \sin(\Omega u)\big) \\ &= 2\sin\left(\frac{\Omega h}{2}\right)\left(-\sin\left(\Omega u + \frac{\Omega h}{2}\right), \cos\left(\Omega u + \frac{\Omega h}{2}\right)\right). \end{aligned}$$

Thus every length-h chord has magnitude $2|\sin(\Omega h/2)|$, and its oriented unit direction is the unit tangent at the chord midpoint, up to a common sign. The previous innovation, latest innovation, and target innovation have midpoints $-3h/2, -h/2, h/2$, respectively. The angle between the previous and latest chord lines is therefore Ωh modulo sign, and the squared projective sine is $\sin^2(\Omega h)$. This proves the first identity in equation 29.

The latest and target innovations have equal magnitude, and their oriented directions differ by Ωh. Hence

$$c_h = \frac{\langle \Delta v_h, I_h \rangle}{\|\Delta v_h\|_2^2} = \cos(\Omega h),$$

which proves the second identity. At unit normalized horizon, $R = \sin^2 \theta_h$, so equation 4 gives equation 30. Finally, with $x = \Omega h$,

$$\cos x = 1 - \frac{x^2}{2} + O(x^4), \qquad \sin^2 x = x^2 + O(x^4),$$

and

$$\frac{1}{1 + \sin^2 x} = 1 - \sin^2 x + O(\sin^4 x) = 1 - x^2 + O(x^4).$$

Substituting $x = \Omega h$ proves equation 31.

C.4 Projective Residual and Projector Identities

For nonzero vectors a, b,

$$P_b = \frac{bb^\top}{\|b\|_2^2}, \qquad \|P_b a\|_2^2 = \frac{(a^\top b)^2}{\|b\|_2^2}.$$

The orthogonality of $P_b a$ and $P_b^\perp a$ implies

$$\frac{\left\|P_b^\perp a\right\|_2^2}{\|a\|_2^2} = 1 - \frac{(a^\top b)^2}{\|a\|_2^2 \, \|b\|_2^2}. \tag{32}$$

Let $u = a/\|a\|_2$ and $v = b/\|b\|_2$. Since u and v are unit vectors,

$$\begin{aligned} \|uu^\top - vv^\top\|_F^2 &= \operatorname{tr}(uu^\top uu^\top) + \operatorname{tr}(vv^\top vv^\top) - 2\operatorname{tr}(uu^\top vv^\top) \\ &= 2 - 2(u^\top v)^2. \end{aligned}$$

Taking $a = \Delta v_q$, $b = \Delta v_q^-$ proves equation 32. Nonzero scalar multiplication does not change a line projector. Orthogonal coordinate transformations conjugate the projectors and preserve their Frobenius distance.

C.5 Why the Update Midpoint Appears

The midpoint horizon used in equation 3 satisfies

$$\int_0^1 r(u)\,du = \rho_{q,n}, \qquad \int_0^1 r(u)^2\,du = \rho_{q,n}^2 + \frac{1}{12}\left(\frac{\sigma_n - \sigma_{n-1}}{\sigma_n^{(1)} - \sigma_n^{(2)}}\right)^2. \tag{33}$$

Let

$$H_q = \sigma_n^{(1)} - \sigma_n^{(2)} \neq 0, \qquad r_0 = \frac{\sigma_n - \sigma_n^{(1)}}{H_q}, \qquad r_1 = \frac{\sigma_{n-1} - \sigma_n^{(1)}}{H_q}.$$

The normalized displacement at the affine interval coordinate $u \in [0,1]$ is

$$\begin{aligned} r(u) &= \frac{(1-u)\sigma_n + u\sigma_{n-1} - \sigma_n^{(1)}}{H_q} \\ &= (1-u)r_0 + ur_1. \end{aligned}$$

Define its midpoint and endpoint difference by

$$\bar{r} = \frac{r_0 + r_1}{2}, \qquad \Delta r = r_1 - r_0.$$

Then

$$\begin{aligned} r(u) &= (1-u)\left(\bar{r} - \frac{\Delta r}{2}\right) + u\left(\bar{r} + \frac{\Delta r}{2}\right) \\ &= \bar{r} + \left(u - \frac{1}{2}\right)\Delta r. \end{aligned} \tag{34}$$

By equation 3, $\bar{r} = \rho_{q,n}$. Also,

$$\begin{aligned} \int_0^1 \left(u - \frac{1}{2}\right)\,du &= \left[\frac{u^2}{2} - \frac{u}{2}\right]_0^1 = 0, \\ \int_0^1 \left(u - \frac{1}{2}\right)^2\,du &= \int_0^1 \left(u^2 - u + \frac{1}{4}\right)\,du \\ &= \left[\frac{u^3}{3} - \frac{u^2}{2} + \frac{u}{4}\right]_0^1 \\ &= \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}. \end{aligned}$$

Integrating equation 34 therefore gives

$$\begin{aligned} \int_0^1 r(u)\,du &= \bar{r} + \Delta r \int_0^1 \left(u - \frac{1}{2}\right)\,du \\ &= \bar{r} = \rho_{q,n}. \end{aligned}$$

Squaring the same identity first gives

$$r(u)^2 = \bar{r}^2 + 2\bar{r}\Delta r\left(u - \frac{1}{2}\right) + (\Delta r)^2\left(u - \frac{1}{2}\right)^2.$$

After integration, the cross term vanishes and the last integral equals $1/12$, so

$$\begin{aligned} \int_0^1 r(u)^2\,du &= \bar{r}^2 + \frac{(\Delta r)^2}{12} \\ &= \rho_{q,n}^2 + \frac{1}{12}\left(\frac{\sigma_{n-1} - \sigma_n}{H_q}\right)^2 \\ &= \rho_{q,n}^2 + \frac{1}{12}\left(\frac{\sigma_n - \sigma_{n-1}}{\sigma_n^{(1)} - \sigma_n^{(2)}}\right)^2. \end{aligned}$$

This proves both identities in equation 33. The implemented rule uses the square of the mean horizon, whereas the exact mean squared horizon contains the nonnegative within-update variance. The difference vanishes quadratically as the solver interval becomes small relative to the latest anchor span.

C.6 Continuum Consistency under C^2 Regularity

Theorem 2 (Continuum limit of the chordal tangent and exposure)**.** *Let g be C^2 near $t_\star$, with $g'(t_\star) \neq 0$. Let $t_0 < t_1 < t_2$ converge to $t_\star$, write $h_1 = t_1 - t_0$, $h_2 = t_2 - t_1$, and suppose $h_1/h_2 \to \eta \in (0, \infty)$. Let u_1, u_2 be the two unit secant directions, let $m_j = (t_{j-1} + t_j)/2$, and orient each sign so that $\langle u_j, U(m_j)\rangle > 0$, where $U(t) = g'(t)/\left\|g'(t)\right\|_2$. Let θ_h be their projective angle. Then*

$$\begin{aligned} -\frac{P_{u_2}^{\perp} u_1}{h_2} &\longrightarrow \frac{1+\eta}{2}\, U'(t_\star), \\ \frac{\sin\theta_h}{h_2} &\longrightarrow \frac{1+\eta}{2}\, \omega(t_\star). \end{aligned} \tag{35}$$

where

$$\omega(t) = \left\|U'(t)\right\|_2 = \frac{\left\|P_{g'(t)}^{\perp} g''(t)\right\|_2}{\left\|g'(t)\right\|_2}. \tag{36}$$

If $\delta_h \to \delta$ and $\rho_h = \delta_h/h_2$, then

$$\begin{aligned} -\rho_h P_{u_2}^{\perp} u_1 &\longrightarrow \delta \frac{1+\eta}{2} U'(t_\star), \\ \rho_h^2 \sin^2\theta_h &\longrightarrow \delta^2 \left(\frac{1+\eta}{2}\right)^2 \omega(t_\star)^2. \end{aligned} \tag{37}$$

Proof of Theorem 2. Define the normalized tangent

$$U(t) = \frac{g'(t)}{\left\|g'(t)\right\|_2}.$$

Because g' is continuous and $g'(t_\star) \neq 0$, there are a neighborhood $\mathcal{N}$ of $t_\star$ and a number $b > 0$ such that $\left\|g'(t)\right\|_2 \geq b$ for every $t \in \mathcal{N}$. Hence normalization is well defined on $\mathcal{N}$. Put $s(t) = \left\|g'(t)\right\|_2$. Direct differentiation gives

$$s'(t) = \frac{\langle g'(t), g''(t)\rangle}{\left\|g'(t)\right\|_2}.$$

Applying the quotient rule to $U(t) = g'(t)/s(t)$ yields

$$\begin{aligned} U'(t) &= \frac{g''(t)}{s(t)} - \frac{g'(t)s'(t)}{s(t)^2} \\ &= \frac{g''(t)}{\left\|g'(t)\right\|_2} - \frac{g'(t)\,\langle g'(t), g''(t)\rangle}{\left\|g'(t)\right\|_2^3} \\ &= \frac{1}{\left\|g'(t)\right\|_2}\left(I - \frac{g'(t)g'(t)^\top}{\left\|g'(t)\right\|_2^2}\right) g''(t) \\ &= \frac{P_{g'(t)}^{\perp} g''(t)}{\left\|g'(t)\right\|_2}. \end{aligned}$$

Therefore $U(t)^\top U'(t) = 0$ and $\left\|U'(t)\right\|_2 = \omega(t)$.

For $j \in \{1, 2\}$, define the interval midpoint and secant velocity

$$m_j = \frac{t_{j-1} + t_j}{2}, \qquad s_j = \frac{g(t_j) - g(t_{j-1})}{h_j}.$$

The fundamental theorem of calculus gives

$$\begin{aligned} s_j &= \frac{1}{h_j}\int_{t_{j-1}}^{t_j} g'(t)\, dt \\ &= \frac{1}{h_j}\int_{-h_j/2}^{h_j/2} g'(m_j + u)\, du. \end{aligned}$$

Since g'' is continuous, uniformly for $|u| \le h_j/2$,

$$g'(m_j + u) = g'(m_j) + ug''(m_j) + ur_j(u), \qquad \sup_{|u|\le h_j/2} \|r_j(u)\|_2 = o(1).$$

Substitution into the integral gives

$$\begin{aligned} s_j - g'(m_j) &= \frac{g''(m_j)}{h_j} \int_{-h_j/2}^{h_j/2} u\,du + \frac{1}{h_j} \int_{-h_j/2}^{h_j/2} ur_j(u)\,du \\ &= \frac{1}{h_j} \int_{-h_j/2}^{h_j/2} ur_j(u)\,du, \end{aligned}$$

because the integral of the odd function u over the symmetric interval is zero. Consequently,

$$\begin{aligned} \|s_j - g'(m_j)\|_2 &\le \frac{1}{h_j} \sup_{|u|\le h_j/2} \|r_j(u)\|_2 \int_{-h_j/2}^{h_j/2} |u|\,du \\ &= \frac{h_j}{4} \sup_{|u|\le h_j/2} \|r_j(u)\|_2 = o(h_j). \end{aligned} \tag{38}$$

The map $N(x) = x/\|x\|_2$ is continuously differentiable wherever $x \neq 0$, with derivative

$$DN(x) = \frac{1}{\|x\|_2} \left(I - \frac{xx^\top}{\|x\|_2^2} \right).$$

Equation 38 and continuity give $s_j \to g'(t_\star)$. Hence, for all sufficiently small spans, $\|s_j - g'(m_j)\|_2 \le b/2$. Every point on the segment joining them has the form $x_\lambda = g'(m_j) + \lambda[s_j - g'(m_j)]$, with $0 \le \lambda \le 1$, and satisfies

$$\|x_\lambda\|_2 \ge \|g'(m_j)\|_2 - \lambda \|s_j - g'(m_j)\|_2 \ge b/2.$$

Thus the displayed derivative satisfies $\|DN(x_\lambda)\|_{\mathrm{op}} \le 2/b$ along the entire segment. The mean-value inequality therefore yields

$$\begin{aligned} \left\| \frac{s_j}{\|s_j\|_2} - \frac{g'(m_j)}{\|g'(m_j)\|_2} \right\| &\le \frac{2}{b} \|s_j - g'(m_j)\|_2 \\ &= o(h_j). \end{aligned}$$

Therefore

$$u_j := \frac{s_j}{\|s_j\|_2} = U(m_j) + o(h_j).$$

In particular, both u_1 and u_2 converge to $U(t_\star)$, so their inner product is positive for all sufficiently small intervals. The ordinary angle and the projective angle then have the same sine.

The midpoint separation is

$$d_h = m_2 - m_1 = \frac{h_1 + h_2}{2}.$$

The assumption $h_1/h_2 \to \eta \in (0, \infty)$ implies $d_h \asymp h_1 \asymp h_2$. Hence $o(h_1) + o(h_2) = o(d_h)$. By the fundamental theorem of calculus applied to U,

$$\begin{aligned} u_2 - u_1 &= U(m_2) - U(m_1) + o(d_h) \\ &= \int_{m_1}^{m_2} U'(t)\,dt + o(d_h) \\ &= d_h U'(t_\star) + o(d_h), \end{aligned}$$

where the last equality follows from continuity of U' and $m_1, m_2 \to t_\star$.

The vector used by the chordal retraction is based at the latest secant u_2. Since $P_{u_2}^\perp u_2 = 0$,

$$\begin{aligned} P_{u_2}^\perp u_1 &= P_{u_2}^\perp (u_1 - u_2) \\ &= -d_h P_{u_2}^\perp U'(t_\star) + o(d_h). \end{aligned}$$

Because $u_2 \to U(t_\star)$ and $U'(t_\star) \perp U(t_\star)$, division by d_h gives the vector limit

$$-\frac{P_{u_2}^{\perp} u_1}{d_h} \longrightarrow U'(t_\star).$$

The two projected residuals have the same norm:

$$\begin{aligned}\left\|P_{u_2}^{\perp} u_1\right\|_2^2 &= 1 - \langle u_1, u_2\rangle^2 \\ &= \left\|P_{u_1}^{\perp} u_2\right\|_2^2 = \sin^2\theta_h.\end{aligned}$$

Taking norms in the vector limit therefore gives

$$\frac{\sin\theta_h}{d_h} \longrightarrow \|U'(t_\star)\|_2 = \omega(t_\star).$$

Finally,

$$\frac{d_h}{h_2} = \frac{h_1 + h_2}{2h_2} = \frac{1 + h_1/h_2}{2} \longrightarrow \frac{1+\eta}{2}.$$

Multiplying both the vector limit and its norm limit by d_h/h_2 proves the two limits in equation 35.

For the transported tangent, use $\rho_h = \delta_h/h_2$ to write

$$\begin{aligned}-\rho_h P_{u_2}^{\perp} u_1 &= \delta_h \left(-\frac{P_{u_2}^{\perp} u_1}{h_2}\right) \\ &\longrightarrow \delta \frac{1+\eta}{2} U'(t_\star).\end{aligned}$$

For its squared magnitude,

$$\begin{aligned}\rho_h^2 \sin^2\theta_h &= \delta_h^2 \left(\frac{\sin\theta_h}{h_2}\right)^2 \\ &\longrightarrow \delta^2 \left(\frac{1+\eta}{2}\right)^2 \omega(t_\star)^2.\end{aligned}$$

These are the two limits in equation 37. The square in the GeoShrink exposure is used because the transported directional deviation enters a squared-error objective; the theorem does not claim that a remote-horizon asymptotic uniquely determines this exponent.

Corollary 1 (Shape-regular bounds without a limiting gap ratio)**.** *Keep the definitions, C^2 regularity, nonvanishing derivative, and convergence $t_0, t_1, t_2 \to t_\star$ from Theorem 2, but do not assume that h_1/h_2 converges. Instead, suppose that for some $\Gamma \geq 1$,*

$$\Gamma^{-1} \leq \frac{h_1}{h_2} \leq \Gamma$$

for all sufficiently small intervals. Then

$$\frac{\sin\theta_h}{h_2} = \frac{1 + h_1/h_2}{2}\,\omega(t_\star) + o(1). \tag{39}$$

If additionally $\delta_h \to \delta$, then

$$\delta^2 \left(\frac{1+\Gamma^{-1}}{2}\right)^2 \omega(t_\star)^2 \leq \liminf_h \rho_h^2 \sin^2\theta_h, \tag{40}$$

$$\limsup_h \rho_h^2 \sin^2\theta_h \leq \delta^2 \left(\frac{1+\Gamma}{2}\right)^2 \omega(t_\star)^2. \tag{41}$$

Proof. The proof above established $\sin\theta_h/d_h \to \omega(t_\star)$ before using convergence of h_1/h_2. The shape-regular bounds imply $d_h \asymp h_2$, so the same remainder remains $o(1)$ after multiplication by $d_h/h_2 = (1 + h_1/h_2)/2$. This proves equation 39. Squaring, multiplying by δ_h^2, and using

$$\frac{1+\Gamma^{-1}}{2} \leq \frac{1 + h_1/h_2}{2} \leq \frac{1+\Gamma}{2}$$

gives equation 40 and equation 41.

Corollary 2 (A quantitative local refinement)**.** *Under the conditions of Theorem 2, write* $h = \max\{h_1, h_2\}$. *Suppose additionally that* $g \in C^3$ *near* $t_\star$ *and* $\max_{0\le j\le 2} |t_j - t_\star| = O(h)$. *Then*

$$\sin\theta_h = d_h\omega(t_\star) + O(h^2). \tag{42}$$

For bounded δ_h *and* $\rho_h = \delta_h/h_2$,

$$\rho_h^2 \sin^2\theta_h = \delta_h^2 \left(\frac{1 + h_1/h_2}{2}\right)^2 \omega(t_\star)^2 + O(\delta_h^2 h). \tag{43}$$

Proof. A midpoint Taylor expansion now gives $s_j = g'(m_j) + O(h_j^2)$. The normalized secants therefore obey $u_j = U(m_j) + O(h_j^2)$. Since $U \in C^2$ and the observation midpoints are $O(h)$ from $t_\star$,

$$u_2 - u_1 = d_h U'(t_\star) + O(h^2), \qquad u_1 = U(t_\star) + O(h).$$

Because $P_{u_2}^{\perp} u_2 = 0$,

$$\begin{aligned} P_{u_2}^{\perp} u_1 &= P_{u_2}^{\perp}(u_1 - u_2) \\ &= -d_h P_{u_2}^{\perp} U'(t_\star) + O(h^2). \end{aligned}$$

Moreover, $u_2 = U(t_\star) + O(h)$ and $U'(t_\star) \perp U(t_\star)$. Since $P_{u_2}^{\perp} = I - u_2 u_2^\top$,

$$\begin{aligned} P_{u_2}^{\perp} U'(t_\star) - U'(t_\star) &= -u_2 u_2^\top U'(t_\star) \\ &= -u_2 \langle u_2 - U(t_\star), U'(t_\star)\rangle \\ &= O(h). \end{aligned}$$

Consequently,

$$P_{u_2}^{\perp} u_1 = -d_h U'(t_\star) + O(h^2),$$

where $d_h = O(h)$. Taking norms and using $d_h > 0$ gives

$$\begin{aligned} \sin\theta_h &= \left\| P_{u_2}^{\perp} u_1 \right\|_2 \\ &= d_h \left\| U'(t_\star) \right\|_2 + O(h^2) \\ &= d_h \omega(t_\star) + O(h^2), \end{aligned}$$

which proves equation 42. Squaring gives

$$\begin{aligned} \sin^2\theta_h &= \left[d_h\omega(t_\star) + O(h^2) \right]^2 \\ &= d_h^2 \omega(t_\star)^2 + O(d_h h^2) + O(h^4) \\ &= d_h^2 \omega(t_\star)^2 + O(h^3). \end{aligned}$$

Because $h_1/h_2 \to \eta \in (0, \infty)$, we have $h_2 \asymp h$. Using $\rho_h = \delta_h/h_2$,

$$\begin{aligned} \rho_h^2 \sin^2\theta_h &= \frac{\delta_h^2}{h_2^2} \left[d_h^2 \omega(t_\star)^2 + O(h^3) \right] \\ &= \delta_h^2 \left(\frac{d_h}{h_2}\right)^2 \omega(t_\star)^2 + O(\delta_h^2 h) \\ &= \delta_h^2 \left(\frac{1 + h_1/h_2}{2}\right)^2 \omega(t_\star)^2 + O(\delta_h^2 h), \end{aligned}$$

which proves equation 43.

The localization condition in this corollary is needed to state the $O(h^2)$ rate at a fixed $t_\star$. Convergence of all observation times to $t_\star$ alone suffices for the main theorem's limit, but does not specify their distance from $t_\star$ relative to their spacing. Neither result predicts a finite remote tangent from local information. Applying the interpretation to a sampler additionally requires a smooth model-output trace; the lemmas alone do not establish smoothness of an accelerated trajectory.

C.7 Complete Derivation of Chordal Projective Retention

Theorem 3 (Chordal transport and cache projection)**.** *Under the chordal transport construction of Proposition 1, let* $\Delta v = \Delta v_q$, $\Delta v^- = \Delta v_q^-$, $L_q = \operatorname{span}(\Delta v)$, *and* $\widetilde{L}_{q,n} = \operatorname{span}(\widetilde{u}_{q,n})$. *The two projection problems*

$$y_{q,n} = \arg\min_{y \in \widetilde{L}_{q,n}} \|y - \Delta v\|_2^2, \tag{44}$$

$$\widehat{\Delta v}_{q,n} = \arg\min_{x \in L_q} \|x - y_{q,n}\|_2^2 \tag{45}$$

have unique solutions, and

$$\widehat{\Delta v}_{q,n} = P_u \widetilde{P}_{q,n} \Delta v = \operatorname{tr}(P_u \widetilde{P}_{q,n}) \Delta v = \frac{1}{1 + \rho_{q,n}^2 \sin^2 \theta_q} \Delta v. \tag{46}$$

The same scalar obeys

$$\begin{aligned} \omega_{q,n} &= \frac{\left\| \widetilde{P}_{q,n} \Delta v \right\|_2^2}{\|\Delta v\|_2^2} = \operatorname{tr}(P_u \widetilde{P}_{q,n}) \\ &= 1 - \frac{1}{2} \left\| P_u - \widetilde{P}_{q,n} \right\|_F^2. \end{aligned} \tag{47}$$

It is also the Hilbert-Schmidt compression coefficient,

$$\begin{aligned} \omega_{q,n} &= \arg\min_{\gamma \in \mathbb{R}} \left\| \widetilde{P}_{q,n} - \gamma P_u \right\|_F^2, \\ P_u \widetilde{P}_{q,n} P_u &= \omega_{q,n} P_u. \end{aligned} \tag{48}$$

Finally, for every $\omega \in \mathbb{R}$,

$$(1 + R_{q,n}) \left\| \omega \Delta v - \widetilde{P}_{q,n} \Delta v \right\|_2^2 = \|\Delta v - \omega \Delta v\|_2^2 + \rho_{q,n}^2 \left\| P_{\Delta v^-}^{\perp} (\omega \Delta v) \right\|_2^2. \tag{49}$$

Therefore the unique cache-compatible coefficient is exactly equation 4.

This section proves Theorem 3 for the deterministic closure stated in the main text. The closure has four explicit parts: the scalar correction family in equation 1, the observable chordal tangent between the last two innovation lines, the declared constant-tangent ambient continuation followed by normalization, and two Euclidean orthogonal projections. Equivalently, the last operation is the Hilbert-Schmidt compression of the transported line projector onto the one-dimensional cache-operator family. Once these parts are fixed, the coefficient and the quadratic identity follow without a stochastic model, a fitted constant, or an assumed future output. Smoothness supplies only the local scaling result in Theorem 2; it does not uniquely select this closure.

Step 1: orient the two line representatives. Abbreviate

$$\Delta v = \Delta v_q, \qquad \Delta v^- = \Delta v_q^-, \qquad A = \|\Delta v\|_2^2 > 0,$$

and define the unit vectors

$$u = \frac{\Delta v}{\sqrt{A}}, \qquad v = \frac{\Delta v^-}{\|\Delta v^-\|_2}.$$

A projective line is unchanged when its unit representative changes sign. Choose $\bar{v} \in \{v, -v\}$ so that

$$\alpha = \langle u, \bar{v} \rangle = |\langle u, v \rangle| \in [0, 1].$$

When $\langle u, v \rangle = 0$, either sign may be used. Define the component of the previous direction tangent to the unit sphere at u by

$$\xi = P_u^{\perp} \bar{v} = (I - uu^{\top}) \bar{v} = \bar{v} - \alpha u. \tag{50}$$

Its orthogonality to u follows term by term:

$$\begin{aligned} \langle u, \xi \rangle &= \langle u, \bar{v} \rangle - \alpha \langle u, u \rangle \\ &= \alpha - \alpha = 0. \end{aligned}$$

Its squared norm is

$$\begin{aligned}\|\xi\|_2^2 &= \langle \bar{v} - \alpha u, \bar{v} - \alpha u \rangle \\ &= \|\bar{v}\|_2^2 - 2\alpha \langle u, \bar{v} \rangle + \alpha^2 \|u\|_2^2 \\ &= 1 - 2\alpha^2 + \alpha^2 \\ &= 1 - \alpha^2 \\ &= 1 - \langle u, v \rangle^2 \\ &= 1 - \frac{\langle \Delta v, \Delta v^- \rangle^2}{\|\Delta v\|_2^2 \|\Delta v^-\|_2^2} \\ &= \sin^2 \theta_q.\end{aligned} \tag{51}$$

This establishes the chordal-tangent identity used in Proposition 1. In particular, $\sqrt{\sin^2 \theta_q}$ is exactly the chordal tangent magnitude, not a postulated curvature variable.

Step 2: verify the normalization retraction. For a tangent vector $\eta \in u^\perp$, the normalization map

$$\mathrm{Retr}_u(\eta) = \frac{u + \eta}{\|u + \eta\|_2}$$

is a retraction on the unit sphere. Indeed,

$$\mathrm{Retr}_u(0) = u.$$

For a scalar s, orthogonality gives

$$\|u + s\eta\|_2^2 = \|u\|_2^2 + 2s \langle u, \eta \rangle + s^2 \|\eta\|_2^2 = 1 + s^2 \|\eta\|_2^2,$$

so

$$\mathrm{Retr}_u(s\eta) = \frac{u + s\eta}{\sqrt{1 + s^2 \|\eta\|_2^2}}.$$

Differentiating at $s = 0$ gives

$$\begin{aligned}\frac{d}{ds} \mathrm{Retr}_u(s\eta)\Big|_{s=0} &= \frac{\eta}{\sqrt{1 + s^2 \|\eta\|_2^2}} - \frac{s \|\eta\|_2^2 (u + s\eta)}{(1 + s^2 \|\eta\|_2^2)^{3/2}}\Bigg|_{s=0} \\ &= \eta.\end{aligned}$$

Thus the map has the required first-order velocity. Taking $\eta = -\rho_{q,n}\xi$, and using equation 51, gives

$$\begin{aligned}\|u - \rho_{q,n}\xi\|_2^2 &= \|u\|_2^2 - 2\rho_{q,n} \langle u, \xi \rangle + \rho_{q,n}^2 \|\xi\|_2^2 \\ &= 1 + \rho_{q,n}^2 \sin^2 \theta_q.\end{aligned}$$

Set $a_\rho = u - \rho_{q,n}\xi$. The preceding norm identity shows that $a_\rho \neq 0$. For every unit vector y, expansion followed by the Cauchy-Schwarz inequality gives

$$\begin{aligned}\|y - a_\rho\|_2^2 &= \|y\|_2^2 + \|a_\rho\|_2^2 - 2 \langle y, a_\rho \rangle \\ &= 1 + \|a_\rho\|_2^2 - 2 \langle y, a_\rho \rangle \\ &\geq 1 + \|a_\rho\|_2^2 - 2 \|a_\rho\|_2.\end{aligned}$$

Equality in Cauchy-Schwarz holds with the required positive inner product if and only if $y = a_\rho / \|a_\rho\|_2$. Therefore the unique unit vector closest to the affine first-order continuation is

$$\widetilde{u} = \mathrm{Retr}_u(-\rho_{q,n}\xi) = \arg\min_{\|y\|_2 = 1} \|y - a_\rho\|_2^2 = \frac{u - \rho_{q,n}\xi}{\sqrt{1 + \rho_{q,n}^2 \sin^2 \theta_q}}.$$

It defines the orthogonal line projector

$$Q = \widetilde{u}\widetilde{u}^\top, \qquad Q^\top = Q, \qquad Q^2 = Q.$$

The last identity follows explicitly from

$$Q^2 = \widetilde{u}\widetilde{u}^\top \widetilde{u}\widetilde{u}^\top = \widetilde{u}(\widetilde{u}^\top \widetilde{u})\widetilde{u}^\top = \widetilde{u}\widetilde{u}^\top = Q.$$

This establishes the normalization retraction used in Proposition 1.

Step 3: solve the projection onto the extrapolated line. Let $\widetilde{L} = \mathrm{span}(\widetilde{u})$. Every vector $y \in \widetilde{L}$ satisfies $Qy = y$. Moreover,

$$\widetilde{u}^\top(\Delta v - Q\Delta v) = \widetilde{u}^\top \Delta v - \widetilde{u}^\top \widetilde{u}\widetilde{u}^\top \Delta v = 0,$$

so $\Delta v - Q\Delta v$ is orthogonal to $\widetilde{L}$. For arbitrary $y \in \widetilde{L}$, write

$$\Delta v - y = (\Delta v - Q\Delta v) + (Q\Delta v - y).$$

The first summand is orthogonal to $\widetilde{L}$, while the second belongs to $\widetilde{L}$. The Pythagorean theorem therefore gives

$$\|\Delta v - y\|_2^2 = \|\Delta v - Q\Delta v\|_2^2 + \|Q\Delta v - y\|_2^2 \geq \|\Delta v - Q\Delta v\|_2^2 . \tag{52}$$

Equality holds exactly when $y = Q\Delta v$. Hence

$$y_{q,n} = Q\Delta v$$

is the unique solution of equation 44.

Step 4: project back to the cache-compatible line. Let $L = \mathrm{span}(u) = \mathrm{span}(\Delta v)$ and $P = uu^\top$. The same orthogonal projection argument, now applied to $Q\Delta v$ and the closed subspace L, gives the unique solution

$$\widehat{\Delta v}_{q,n} = PQ\Delta v$$

of equation 45. It remains to evaluate this vector. Orthogonality of u and ξ gives

$$\begin{aligned}
\langle u, \widetilde{u}\rangle &= \frac{\langle u, u - \rho_{q,n}\xi\rangle}{\sqrt{1 + \rho_{q,n}^2 \sin^2\theta_q}} \\
&= \frac{\langle u, u\rangle - \rho_{q,n}\langle u, \xi\rangle}{\sqrt{1 + \rho_{q,n}^2 \sin^2\theta_q}} \\
&= \frac{1}{\sqrt{1 + \rho_{q,n}^2 \sin^2\theta_q}}.
\end{aligned} \tag{53}$$

Because $\Delta v = \sqrt{A}\, u$, first applying Q gives

$$\begin{aligned}
Q\Delta v &= \widetilde{u}\widetilde{u}^\top(\sqrt{A}\, u) \\
&= \sqrt{A}\ \langle \widetilde{u}, u\rangle\, \widetilde{u}.
\end{aligned}$$

Applying P gives

$$\begin{aligned}
PQ\Delta v &= uu^\top \left(\sqrt{A}\ \langle \widetilde{u}, u\rangle\, \widetilde{u}\right) \\
&= \sqrt{A}\ \langle \widetilde{u}, u\rangle \langle u, \widetilde{u}\rangle\, u \\
&= \sqrt{A}\ \langle u, \widetilde{u}\rangle^2\, u \\
&= \frac{\sqrt{A}}{1 + \rho_{q,n}^2 \sin^2\theta_q} u \\
&= \frac{1}{1 + \rho_{q,n}^2 \sin^2\theta_q} \Delta v.
\end{aligned}$$

The transported component has squared norm

$$\begin{aligned}
\|Q\Delta v\|_2^2 &= (Q\Delta v)^\top (Q\Delta v) \\
&= \Delta v^\top Q^\top Q \Delta v \\
&= \Delta v^\top Q \Delta v \\
&= A\, u^\top \widetilde{u}\widetilde{u}^\top u \\
&= A\ \langle u, \widetilde{u}\rangle^2 .
\end{aligned}$$

Dividing by $\|\Delta v\|_2^2 = A$ shows that the retained energy fraction is $\langle u, \widetilde{u}\rangle^2$. The same scalar is the overlap of the two rank-one projectors. Cyclicity of the trace gives

$$\begin{aligned}\operatorname{tr}(PQ) &= \operatorname{tr}(uu^\top \widetilde{u}\widetilde{u}^\top) \\ &= \operatorname{tr}(u^\top \widetilde{u}\widetilde{u}^\top u) \\ &= \langle u, \widetilde{u}\rangle^2 \\ &= \frac{1}{1+\rho_{q,n}^2 \sin^2\theta_q}.\end{aligned}$$

Moreover, since $P^2 = P$, $Q^2 = Q$, and both projectors have trace one,

$$\begin{aligned}\|P-Q\|_F^2 &= \operatorname{tr}[(P-Q)^\top (P-Q)] \\ &= \operatorname{tr}(P^2) + \operatorname{tr}(Q^2) - 2\operatorname{tr}(PQ) \\ &= 2 - 2\operatorname{tr}(PQ).\end{aligned}$$

Rearrangement proves equation 47. This also proves equation 46.

The same scalar follows from an operator-space least-squares problem. Equip matrices with the Hilbert-Schmidt inner product

$$\langle A, B\rangle_F = \operatorname{tr}(A^\top B).$$

Because P and Q are symmetric rank-one projectors,

$$\|P\|_F^2 = \operatorname{tr}(P^2) = \operatorname{tr}(P) = 1, \qquad \|Q\|_F^2 = 1.$$

For any $\gamma \in \mathbb{R}$, bilinearity of this inner product gives

$$\begin{aligned}\|Q-\gamma P\|_F^2 &= \|Q\|_F^2 - 2\gamma\langle Q, P\rangle_F + \gamma^2\|P\|_F^2 \\ &= 1 - 2\gamma\operatorname{tr}(PQ) + \gamma^2 \\ &= [\gamma - \operatorname{tr}(PQ)]^2 + 1 - \operatorname{tr}(PQ)^2.\end{aligned}$$

The final term does not depend on γ, so the unique minimizer is

$$\gamma^\star = \operatorname{tr}(PQ) = \omega_{q,n}.$$

Moreover,

$$\begin{aligned}PQP &= uu^\top Quu^\top \\ &= \left(u^\top Qu\right) uu^\top \\ &= \operatorname{tr}(PQ)P \\ &= \omega_{q,n}P.\end{aligned}$$

Since $P\Delta v = \Delta v$, applying this compressed operator to the latest innovation gives

$$PQP\Delta v = \omega_{q,n}P\Delta v = \omega_{q,n}\Delta v,$$

while $PQP\Delta v = PQ\Delta v$. This proves every statement in equation 48 and shows that the sequential vector projections are the action of the Hilbert-Schmidt compression on the cache line.

Therefore the scalar coefficient ω, the retained-energy fraction, and the scalar projection loss are unchanged by the sign of $\bar{v}$, by the sign of $\rho_{q,n}$, or by reversing either observed line. When $\langle u, v\rangle = 0$, the two admissible choices of $\bar{v}$ may produce different oriented transported lines, but they have the same squared line overlap with L and hence the same coefficient.

Step 5: derive the scalar least-squares problem. Every cache-compatible correction has the form $x = \omega\Delta v$. Since $PQ\Delta v$ is the orthogonal projection of $Q\Delta v$ onto L, it is also the unique minimizer of

$$\min_{\omega\in\mathbb{R}} \|\omega\Delta v - Q\Delta v\|_2^2.$$

For completeness, this follows by direct expansion. Put

$$R = \rho_{q,n}^2 \sin^2\theta_q, \qquad a^2 = \langle u, \widetilde{u}\rangle^2 = \frac{1}{1+R}.$$

Because Q is symmetric and idempotent,

$$\langle \Delta v, Q\Delta v\rangle = \Delta v^\top Q\Delta v,$$
$$\|Q\Delta v\|_2^2 = (Q\Delta v)^\top (Q\Delta v) = \Delta v^\top Q^\top Q\Delta v = \Delta v^\top Q^2\Delta v = \Delta v^\top Q\Delta v.$$

Equation 53 further gives

$$\Delta v^\top Q\Delta v = A\, u^\top \widetilde{u}\widetilde{u}^\top u = Aa^2 = \frac{A}{1+R}.$$

Consequently,

$$\begin{aligned}\|\omega\Delta v - Q\Delta v\|_2^2 &= \|\omega\Delta v\|_2^2 - 2\,\langle \omega\Delta v, Q\Delta v\rangle + \|Q\Delta v\|_2^2\\ &= A\omega^2 - 2A\omega a^2 + Aa^2\\ &= A\left[(\omega - a^2)^2 + a^2 - a^4\right].\end{aligned}$$

The last term is independent of ω, and $A > 0$, so the unique minimizer is

$$\omega = a^2 = \frac{1}{1+R}.$$

Since $R \geq 0$, this value lies in $(0, 1]$, so imposing the bounded family $0 \leq \omega \leq 1$ does not change the solution.

Step 6: prove the exact variational identity. Multiply the preceding expansion before completing the square by $1 + R$:

$$\begin{aligned}&(1+R)\,\|\omega\Delta v - Q\Delta v\|_2^2\\ &\quad= (1+R)A\left(\omega^2 - \frac{2\omega}{1+R} + \frac{1}{1+R}\right)\\ &\quad= A\left[(1+R)\omega^2 - 2\omega + 1\right].\end{aligned} \tag{54}$$

The first term on the right side of equation 49 is

$$\|\Delta v - \omega\Delta v\|_2^2 = (1-\omega)^2 A = A(1 - 2\omega + \omega^2).$$

Linearity of $P^{\perp}_{\Delta v^-}$ and equation 32 give

$$\begin{aligned}\rho_{q,n}^2 \left\|P^{\perp}_{\Delta v^-}(\omega\Delta v)\right\|_2^2 &= \rho_{q,n}^2\omega^2 \left\|P^{\perp}_{\Delta v^-}\Delta v\right\|_2^2\\ &= \rho_{q,n}^2\omega^2 A\sin^2\theta_q\\ &= AR\omega^2.\end{aligned}$$

Adding these two expressions yields

$$\begin{aligned}&\|\Delta v - \omega\Delta v\|_2^2 + \rho_{q,n}^2 \left\|P^{\perp}_{\Delta v^-}(\omega\Delta v)\right\|_2^2\\ &\quad= A(1 - 2\omega + \omega^2) + AR\omega^2\\ &\quad= A\left[(1+R)\omega^2 - 2\omega + 1\right],\end{aligned}$$

which equals equation 54. This proves equation 49 exactly. In particular, the objective on the right-hand side of equation 49 is a positive rescaling of one euclidean projection loss, and its two displayed terms are its exact scalar expansion. They need not be interpreted as orthogonal components of the original output-space residual. Completing the square gives the equivalent certificate

$$\mathcal{J}_{q,n}(\omega) = A(1+R)\left(\omega - \frac{1}{1+R}\right)^2 + \frac{AR}{1+R}. \tag{55}$$

The nonnegative square vanishes only at the GeoShrink coefficient, proving optimality and uniqueness a second way.

Sensitivity to reference amplitude and tangent scale. The same projection geometry identifies precisely how two deliberate method changes alter the coefficient. Let $\beta \in \mathbb{R}$ multiply the reference innovation, and rescale the chordal tangent by $\sqrt{\kappa}$, where $\kappa \geq 0$. Define

$$\widetilde{u}_\kappa = \frac{u - \sqrt{\kappa}\rho_{q,n}\xi}{\sqrt{1+\kappa R}}, \qquad Q_\kappa = \widetilde{u}_\kappa \widetilde{u}_\kappa^\top.$$

Repeating equation 53 gives

$$\langle u, \widetilde{u}_\kappa \rangle^2 = \frac{1}{1+\kappa R}.$$

Therefore

$$\arg\min_{\omega \in \mathbb{R}} \|\omega \Delta v - \beta Q_\kappa \Delta v\|_2^2 = \frac{\beta}{1+\kappa R}. \tag{56}$$

If the candidate family is constrained to $0 \leq \omega \leq 1$, the solution is its Euclidean projection onto that interval:

$$\omega^\star_{[0,1]} = \mathrm{clip}_{[0,1]}\left(\frac{\beta}{1+\kappa R}\right). \tag{57}$$

The corresponding exact identity is

$$\begin{aligned}(1+\kappa R)\,\|\omega \Delta v - \beta Q_\kappa \Delta v\|_2^2 \\ = A(\beta-\omega)^2 + \kappa \rho_{q,n}^2 \left\|P^\perp_{\Delta v^-}(\omega \Delta v)\right\|_2^2.\end{aligned} \tag{58}$$

This is a counterfactual sensitivity calculation, rather than an additional assumption used by GeoShrink. GeoShrink uses the latest observed innovation without amplitude rescaling and the unscaled ambient chordal tangent, hence $(\beta, \kappa) = (1, 1)$. A time-scaled reference sets $\beta = \lambda_{q,n}$. A value $\kappa \neq 1$ changes the tangent transport scale and is a different method; it is not required to derive the implemented rule.

Relation to intrinsic projective geometry. Let

$$\theta_q = \arccos |\, \langle u, v \rangle \,| \in [0, \pi/2].$$

Then

$$\|\xi\|_2 = \sin\theta_q = \theta_q + O(\theta_q^3) \qquad \text{as } \theta_q \to 0.$$

Thus the chordal tangent and the intrinsic projective logarithm agree to first order in the fine-anchor regime covered by Theorem 2. They differ at finite angles. If the intrinsic logarithm is followed by the same normalization retraction and round-trip projection, define, for $\theta_q > 0$, the tangent pointing away from the aligned previous line by

$$\zeta = -\frac{\theta_q}{\sin\theta_q}\xi.$$

Because $\xi \perp u$ and $\|\xi\|_2 = \sin\theta_q$,

$$\zeta \perp u, \qquad \|\zeta\|_2 = \theta_q.$$

The normalization retraction therefore gives

$$u_{\log} = \frac{u + \rho_{q,n}\zeta}{\sqrt{1+\rho_{q,n}^2\theta_q^2}}, \qquad \langle u, u_{\log}\rangle^2 = \frac{1}{1+\rho_{q,n}^2\theta_q^2}.$$

Step 4 showed that the round-trip projection coefficient equals this squared line overlap. Hence

$$\omega_{\log} = \frac{1}{1+\rho_{q,n}^2\theta_q^2}.$$

At $\theta_q = 0$, the same formula follows by continuity and equals one.

If instead one transports by the exact projective exponential along the same local geodesic branch, its unit-sphere representative is

$$u_{\exp} = \cos(\rho_{q,n}\theta_q)u + \sin(\rho_{q,n}\theta_q)\frac{\zeta}{\theta_q}.$$

Orthogonality of u and ζ gives

$$\langle u, u_{\exp} \rangle = \cos(\rho_{q,n}\theta_q).$$

Applying the same round-trip projection identity then yields

$$\omega_{\exp} = \cos^2(\rho_{q,n}\theta_q),$$

for horizons that remain on that branch. These are different finite rules. We call the present construction chordal or extrinsic throughout because its normalization retraction and round-trip projection give exactly $(1 + \rho_{q,n}^2 \sin^2 \theta_q)^{-1}$ in the zero-regularization model.

Degenerate cases and scope. If $\Delta v = 0$, every vector $\omega \Delta v$ is zero, so the correction is zero and the prediction is $v_n^{(1)}$; no current line or scalar coefficient needs to be defined. If $\Delta v^- = 0$, the preceding line and its chordal tangent are unidentified, so the ideal theorem does not assign a geometric coefficient; a numerical fallback in this event is an operational convention. If $\sin^2 \theta_q = 0$, then $\xi = 0$, $\widetilde{L} = L$, and $\omega = 1$. If $\rho_{q,n} = 0$, the retracted line also equals L and $\omega = 1$.

The theorem establishes exact optimality for the observable chordal transport-and-projection closure. It does not identify the true future line. Proposition 4 shows why no such identification follows from finite anchor data and smoothness alone; the causal bound in equation 64 states separately what standard $C^{1,1}$ regularity does guarantee about the actual trace error.

C.8 Blockwise Direct-Sum Extension

Let the output space be an orthogonal direct sum $\mathcal{H} = \bigoplus_{b=1}^{G} \mathcal{H}_b$, and write $\Delta v = (\Delta v^{(1)}, \ldots, \Delta v^{(G)})$. A modality, a batch element, or any orthogonal coordinate/token group may define one block. No statistical independence assumption is required. For block-specific horizons ρ_b, previous innovations $\Delta v^{(b),-}$, and coefficients ω_b, the ideal objective below assumes $\Delta v^{(b),-} \neq 0$ for every block. If $\Delta v^{(b),-} = 0$, the ideal projector term is undefined and the residual convention of Appendix C.15 should be used. If $A_b = \left\| \Delta v^{(b)} \right\|_2^2 = 0$, the corresponding block correction is zero for every coefficient, so that coefficient is immaterial. Under the nondegenerate ideal convention, consider

$$\mathcal{J}_{\text{block}}(\omega_1, \ldots, \omega_G) = \sum_{b=1}^{G} \left[\left\| \Delta v^{(b)} - \omega_b \Delta v^{(b)} \right\|_2^2 + \rho_b^2 \left\| P^{\perp}_{\Delta v^{(b),-}} (\omega_b \Delta v^{(b)}) \right\|_2^2 \right]. \tag{59}$$

Proposition 8 (Blockwise GeoShrink)**.** *For every block with $A_b = \left\| \Delta v^{(b)} \right\|_2^2 > 0$ and $\Delta v^{(b),-} \neq 0$, the unique minimizer of equation 59 is*

$$\omega_b^\star = \frac{A_b}{A_b + \rho_b^2 N_b} = \frac{1}{1 + \rho_b^2 \sin^2 \theta_b}, \quad N_b = \left\| P^{\perp}_{\Delta v^{(b),-}} \Delta v^{(b)} \right\|_2^2, \quad \sin^2 \theta_b = N_b / A_b. \tag{60}$$

If a single shared coefficient ω is imposed across all blocks, its unique minimizer, provided $\sum_b A_b > 0$, is

$$\omega_{\text{shared}}^\star = \frac{\sum_b A_b}{\sum_b A_b + \sum_b \rho_b^2 N_b}. \tag{61}$$

Proof. Orthogonality of the direct sum makes the squared norm of a concatenated vector equal to the sum of its block squared norms. Therefore equation 59 contains no cross-block terms. For each block, the expansion in equation 49 gives

$$\mathcal{J}_b(\omega_b) = A_b(1 - \omega_b)^2 + \rho_b^2 N_b \omega_b^2.$$

Differentiation gives

$$\mathcal{J}_b'(\omega_b) = 2(A_b + \rho_b^2 N_b)\omega_b - 2A_b,$$

and the second derivative $2(A_b + \rho_b^2 N_b)$ is positive. Solving the first-order equation proves equation 60. If all blocks share ω, summing the displayed quadratic over b gives

$$\mathcal{J}_{\text{shared}}(\omega) = \left(\sum_b A_b \right) (1 - \omega)^2 + \omega^2 \sum_b \rho_b^2 N_b.$$

Its derivative is

$$2\left(\sum_b A_b + \sum_b \rho_b^2 N_b\right)\omega - 2\sum_b A_b.$$

Setting it to zero proves equation 61; the positive second derivative proves uniqueness.

C.9 Innovation Retention and Time-Scaled Extrapolation

The same dimensionless time displacement can play two different roles. GeoShrink uses the update-midpoint displacement $\rho_{q,n}$ only as the strength of the transported-turning penalty. An affine extrapolator for the model output at the stage time σ_n would instead multiply the latest innovation by

$$\lambda_{q,n} = \frac{\sigma_n - \sigma_n^{(1)}}{\sigma_n^{(1)} - \sigma_n^{(2)}} \tag{62}$$

to account for the target displacement. The two quantities share the same anchor-span denominator, but $\rho_{q,n}$ uses the solver-interval midpoint whereas $\lambda_{q,n}$ uses the field-evaluation time. More fundamentally, GeoShrink uses the displacement to control confidence and deliberately does not use it as an extrapolation amplitude.

For example, let $g(t) = b + tv$, $v \neq 0$, with ordered anchor times $t_{q-2} < t_{q-1} < t_q$. Both innovations are parallel to v, so the ideal $\sin^2\theta_q = 0$. GeoShrink then predicts

$$\widehat{g}(t) = g(t_q) + g(t_q) - g(t_{q-1}) = g(t_q) + (t_q - t_{q-1})v.$$

This equals $g(t)$ only when $t - t_q = t_q - t_{q-1}$. The ideal predictor therefore does not reproduce an arbitrary affine trace at every target time. This is a statement about affine exactness, not a proof that the method cannot converge when all anchor gaps and forecast distances tend to zero.

A time-scaled comparison rule would be

$$\widehat{v}_n^{\text{time}} = v_n^{(1)} + \frac{\lambda_{q,n}\Delta v_q}{1 + R_{q,n}}, \tag{63}$$

which is affine-exact in the ideal zero-turning case. It belongs to a different correction family, based on $\lambda_{q,n}\Delta v_q$. For $\lambda_{q,n} > 1$, it can exceed one observed innovation even though its retention coefficient is at most one. It is not the algorithm specified in this manuscript.

Affine exactness by itself does not establish a smaller error on nonlinear generative trajectories. A comparison between the two rules should keep the exact-query schedule fixed. For either family, the loss decomposition in Proposition 2 applies after choosing the corresponding nominal correction vector. Purely tangential acceleration remains invisible to the projective statistic: any scalar trace $g(t) = b(t)v$ has zero ideal projective angle whenever its two innovations are nonzero.

C.10 Causal Secant Remainder and Retention Bias

Under the $C^{1,1}$ assumptions stated below, let $\Delta v = g(a) - g(a-h)$, $\lambda = \delta/h$, and let $\omega \in \mathbb{R}$. Then

$$\|g(a+\delta) - g(a) - \omega\Delta v\|_2 \leq \frac{M}{2}\delta(\delta + h) + |\lambda - \omega|\,\|\Delta v\|_2\,. \tag{64}$$

We now prove equation 64 directly from the standard $C^{1,1}$ regularity used in ODE error analysis. Let g take values in a finite-dimensional Euclidean space, or more generally in a Hilbert space, and suppose

$$\|g'(t) - g'(s)\|_2 \leq M|t - s| \qquad \text{for all } s, t \in [a-h, a+\delta]. \tag{65}$$

Assume $h > 0$ and $\delta \geq 0$, and define

$$\Delta v = g(a) - g(a-h), \qquad \lambda = \frac{\delta}{h}.$$

The fundamental theorem of calculus gives

$$\Delta v = \int_{a-h}^{a} g'(u)\,du = \int_0^h g'(a-u)\,du, \tag{66}$$

$$g(a+\delta) - g(a) = \int_a^{a+\delta} g'(u)\,du = \int_0^\delta g'(a+s)\,ds. \tag{67}$$

Multiply equation 66 by $\lambda = \delta/h$, subtract it from equation 67, and add and subtract $\delta g'(a)$. This yields the exact identity

$$\begin{aligned} &g(a+\delta) - g(a) - \lambda\Delta v \\ &\quad = \int_0^\delta \big[g'(a+s) - g'(a)\big]\, ds + \frac{\delta}{h}\int_0^h \big[g'(a) - g'(a-u)\big]\, du. \end{aligned} \tag{68}$$

Apply the triangle inequality for vector-valued integrals and then equation 65 to each term:

$$\begin{aligned} &\|g(a+\delta) - g(a) - \lambda\Delta v\|_2 \\ &\le \int_0^\delta \|g'(a+s) - g'(a)\|_2\, ds + \frac{\delta}{h}\int_0^h \|g'(a) - g'(a-u)\|_2\, du \\ &\le \int_0^\delta Ms\, ds + \frac{\delta}{h}\int_0^h Mu\, du \\ &= \frac{M\delta^2}{2} + \frac{M\delta h}{2} \\ &= \frac{M}{2}\delta(\delta+h). \end{aligned}$$

This is the causal secant remainder: it uses only past observations and is exact for every affine trace.

For an arbitrary correction coefficient ω, add and subtract $\lambda\Delta v$:

$$g(a+\delta) - g(a) - \omega\Delta v = \big[g(a+\delta) - g(a) - \lambda\Delta v\big] + (\lambda - \omega)\Delta v.$$

Taking norms, using the triangle inequality, and substituting the bound just proved gives

$$\|g(a+\delta) - g(a) - \omega\Delta v\|_2 \le \frac{M}{2}\delta(\delta+h) + |\lambda - \omega|\, \|\Delta v\|_2\,,$$

which is equation 64. No probabilistic independence, Gaussian noise, monotone curvature, or assumed relation between tangential and normal acceleration is used.

For the ideal GeoShrink coefficient $\omega = 1/(1+R)$, another triangle inequality on the real line gives

$$\begin{aligned} |\lambda - \omega| &\le |\lambda - 1| + |1 - \omega| \\ &= |\lambda - 1| + \frac{R}{1+R}. \end{aligned}$$

Therefore

$$\|g(a+\delta) - g(a) - \omega\Delta v\|_2 \le \frac{M}{2}h^2\lambda(1+\lambda) + \left(|\lambda - 1| + \frac{R}{1+R}\right)\|\Delta v\|_2\,. \tag{69}$$

The three displayed contributions have separate meanings: local secant truncation, the one-span reference mismatch, and shrinkage bias. When g is affine, $M = 0$; when it is also collinear, $R = 0$. The remaining error is then exactly $|\lambda - 1|\,\|\Delta v\|_2$, recovering the affine counterexample in Appendix C.9.

C.11 Relative-Horizon Certificates

Let $s_{j-1} < s_j < s_{j+1}$ be consecutive anchors in an increasing affine progress coordinate and let

$$h_j = s_j - s_{j-1} > 0, \qquad h_{j+1} = s_{j+1} - s_j > 0.$$

For any update interval wholly contained in the next anchor cell, its midpoint μ satisfies

$$s_j \le \mu \le s_{j+1}.$$

Subtracting s_j from all three terms gives

$$0 \le \mu - s_j \le s_{j+1} - s_j = h_{j+1}.$$

Division by the positive previous span h_j yields

$$0 \le \frac{\mu - s_j}{h_j} \le \frac{h_{j+1}}{h_j},$$

which is equation 6.

Likewise, for any field-evaluation coordinate ξ in the same cell, define

$$\lambda_j^{(s)} = \frac{\xi - s_j}{h_j}.$$

The inequalities $s_j \le \xi \le s_{j+1}$ give

$$0 \le \lambda_j^{(s)} \le \frac{h_{j+1}}{h_j}.$$

If the adjacent ratio is at most Γ, then $\lambda_j^{(s)} \le \Gamma$. Since the derivative of $x(1+x)$ is $1+2x > 0$ for $x \ge 0$, it also follows that

$$\lambda_j^{(s)}(1+\lambda_j^{(s)}) \le \Gamma(1+\Gamma).$$

On an integer progress grid, if the cell length is the integer h_{j+1} and both endpoints are exact anchors, the skipped stage midpoints are

$$s_j + \frac{3}{2}, s_j + \frac{5}{2}, \dots, s_{j+1} - \frac{1}{2}$$

whenever $h_{j+1} \ge 2$. The exact largest skipped midpoint horizon is therefore

$$\max_{\text{skips in cell}} |\rho_j^{(s)}| = \frac{h_{j+1} - 1/2}{h_j} < \frac{h_{j+1}}{h_j}. \tag{70}$$

If there is no skipped stage in the cell, the maximum over skips is zero.

Finally, $0 \le \sin^2\theta_q \le 1$ implies in this progress coordinate

$$R_{q,n}^{(s)} = (\rho_{q,n}^{(s)})^2 \sin^2\theta_q \le \left(\frac{h_{j+1}}{h_j}\right)^2.$$

Since $\omega^{(s)} = 1/(1+R^{(s)})$ is decreasing in $R^{(s)}$, if $h_{j+1}/h_j \le \Gamma$, then

$$\omega_{q,n}^{(s)} \ge \frac{1}{1+\Gamma^2}.$$

This proves equation 7 and establishes the direct connection between adjacent gap expansion and the worst model-free retention certificate.

C.12 The Geometric Mesh at Fixed First Span and Coverage

Theorem 4 (Minimax normalized horizon at fixed coverage)**.** *Fix $N \ge 2$, $d > 0$, and $H \ge Nd$. Among positive spans $\ell_0, \dots, \ell_{N-1}$ satisfying $\ell_0 = d$ and $\sum_{j=0}^{N-1} \ell_j = H$, the unique minimizer of*

$$\Gamma(\ell) = \max_{0 \le j < N-1} \frac{\ell_{j+1}}{\ell_j} \tag{71}$$

is

$$\ell_j = dr^j, \qquad d\sum_{j=0}^{N-1} r^j = H, \qquad r \ge 1. \tag{72}$$

The closure equation has exactly one root in $[1, \infty)$.

Corollary 3 (Canonical terminal-closed mesh after minimal bootstrap)**.** *Fix $B \ge 4$, a total progress length S, and a bootstrap span $d > 0$, with $S \ge (B-1)d$. Among terminal-closed meshes with positive gaps $h_1, \dots, h_{B-1}$,*

$$h_1 = h_2 = d, \qquad \sum_{j=1}^{B-1} h_j = S, \tag{73}$$

the unique minimizer of $\max_{2\le j\le B-2} h_{j+1}/h_j$ *is*

$$h_1 = d, \qquad h_j = dr^{j-2}\ (2 \le j \le B-1), \qquad d\left(1 + \sum_{k=0}^{B-3} r^k\right) = S, \quad r \ge 1. \tag{74}$$

The corresponding continuous anchors are

$$\widetilde{s}_0 = 0, \qquad \widetilde{s}_1 = d, \qquad \widetilde{s}_q = d\left(1 + \sum_{k=0}^{q-2} r^k\right), \quad 2 \le q \le B-1. \tag{75}$$

Proof of Theorem 4. For $N \ge 2$, the polynomial

$$S_N(r) = \sum_{j=0}^{N-1} r^j$$

is continuous on $[1, \infty)$. Its derivative is

$$S_N'(r) = \sum_{j=1}^{N-1} j r^{j-1}.$$

Every summand is nonnegative on $[1, \infty)$, and the $j = 1$ summand equals one. Hence $S_N'(r) > 0$, so S_N is strictly increasing. At the left endpoint,

$$S_N(1) = \sum_{j=0}^{N-1} 1 = N.$$

The last term r^{N-1} diverges as $r \to \infty$, so $S_N(r) \to \infty$. Because $H/d \ge N = S_N(1)$, the intermediate value theorem gives at least one $r \in [1, \infty)$ satisfying $S_N(r) = H/d$. Strict monotonicity gives at most one such r. Thus the closure root exists and is unique.

For any feasible spans, let $\gamma = \max_{0\le j<N-1} \ell_{j+1}/\ell_j$. By the definition of a maximum,

$$\frac{\ell_{j+1}}{\ell_j} \le \gamma \quad \Longrightarrow \quad \ell_{j+1} \le \gamma \ell_j \qquad (0 \le j < N-1).$$

We now prove $\ell_j \le d\gamma^j$ by induction. For $j = 0$, feasibility gives $\ell_0 = d = d\gamma^0$. If the bound holds at index j, then

$$\ell_{j+1} \le \gamma \ell_j \le \gamma(d\gamma^j) = d\gamma^{j+1}.$$

Thus it holds for every $j = 0, \dots, N-1$. Summing the componentwise bounds yields

$$H = \sum_{j=0}^{N-1} \ell_j \le \sum_{j=0}^{N-1} d\gamma^j = dS_N(\gamma). \tag{76}$$

If $\gamma < 1$, the sum is strictly smaller than Nd, contradicting $H \ge Nd$. Explicitly, $\gamma^0 = 1$ and $\gamma^j < 1$ for every $j \ge 1$, so

$$dS_N(\gamma) = d\left(1 + \sum_{j=1}^{N-1} \gamma^j\right) < d[1 + (N-1)] = Nd.$$

Equation 76 would then imply $H < Nd$, a contradiction. Hence $\gamma \ge 1$. On this interval, strict monotonicity of S_N, together with

$$dS_N(\gamma) \ge H = dS_N(r),$$

implies $S_N(\gamma) \ge S_N(r)$ and therefore $\gamma \ge r$.

Now set $\ell_j^\star = dr^j$. Its first span is $\ell_0^\star = d$, and the closure equation gives

$$\sum_{j=0}^{N-1} \ell_j^\star = dS_N(r) = H,$$

so it is feasible. Every adjacent ratio is

$$\frac{\ell^\star_{j+1}}{\ell^\star_j} = \frac{dr^{j+1}}{dr^j} = r,$$

so its objective value is r. The lower bound $\gamma \geq r$ proves global optimality.

For uniqueness, suppose another feasible sequence has objective value r. Repeating the induction above with $\gamma = r$ gives

$$\ell_j \leq dr^j = \ell^\star_j \qquad \text{for every } j.$$

If at least one of these inequalities were strict, summing all coordinates would give

$$\sum_{j=0}^{N-1} \ell_j < \sum_{j=0}^{N-1} dr^j = H,$$

contradicting feasibility. Therefore equality holds at every coordinate, and the geometric sequence is the unique minimizer.

Proof of Corollary 3. The first gap is fixed at $h_1 = d$, so the remaining gaps

$$\ell_0 = h_2, \quad \ell_1 = h_3, \quad \ldots, \quad \ell_{B-3} = h_{B-1}$$

contain $N = B - 2$ positive spans. Their first span is $\ell_0 = h_2 = d$, and their total coverage is

$$\begin{aligned}\sum_{k=0}^{B-3} \ell_k &= \sum_{j=2}^{B-1} h_j \\ &= S - h_1 \\ &= S - d.\end{aligned}$$

The assumption $S \geq (B-1)d$ implies

$$S - d \geq (B-2)d = Nd,$$

so Theorem 4 applies with $H = S - d$. It gives the unique minimizing branch

$$\ell_k = dr^k, \qquad 0 \leq k \leq B-3,$$

where $r \geq 1$ is the unique solution of

$$d \sum_{k=0}^{B-3} r^k = S - d.$$

Restoring $h_1 = d$ and using $h_j = \ell_{j-2}$ for $j \geq 2$ yields

$$h_1 = d, \qquad h_j = dr^{j-2} \quad (2 \leq j \leq B-1).$$

Moving the fixed first gap to the left side of the closure equation gives

$$d\left(1 + \sum_{k=0}^{B-3} r^k\right) = S,$$

which is equation 74. Starting from $\widetilde{s}_0 = 0$, the cumulative locations are

$$\begin{aligned}\widetilde{s}_1 &= h_1 = d, \\ \widetilde{s}_q &= h_1 + \sum_{j=2}^{q} h_j \\ &= d + \sum_{j=2}^{q} dr^{j-2} \\ &= d\left(1 + \sum_{k=0}^{q-2} r^k\right), \qquad 2 \leq q \leq B-1,\end{aligned}$$

which proves equation 75; at $q = B - 1$, the closure equation gives $\widetilde{s}_{B-1} = S$. The objective in the corollary is exactly the adjacent-ratio objective of the ℓ-branch, because

$$\max_{2\le j\le B-2} \frac{h_{j+1}}{h_j} = \max_{0\le k<B-3} \frac{\ell_{k+1}}{\ell_k}.$$

Existence, attainment, and uniqueness therefore follow directly from the corresponding parts of Theorem 4.

C.13 Integer Projection of the Canonical Mesh

For the canonical integer mesh, compute the continuous anchors from equation 75 with $S = T - 1$ and $d = 1$, fix

$$s_0 = 0, \qquad s_1 = 1, \qquad s_2 = 2, \qquad s_{B-1} = T - 1, \tag{77}$$

and sequentially for $3 \le q \le B - 2$ set

$$s_q = \min\{T - B + q, \max(s_{q-1} + 1, Q_{\text{even}}(\widetilde{s}_q))\}. \tag{78}$$

For any resulting integer mesh, define the realized whole-mesh certificate

$$\widehat{\Gamma}(\mathcal{A}) = \max_{2\le j<B-1} \frac{s_{j+1} - s_j}{s_j - s_{j-1}}. \tag{79}$$

We prove that equation 77– equation 78 always produce a feasible integer mesh. For every index q, define the largest value that leaves enough terminal slots by

$$U_q = T - 1 - (B - 1 - q) = T - B + q.$$

Because $T \ge B$, the last bootstrap offset satisfies

$$s_2 = 2 \le T - B + 2 = U_2.$$

Now fix $q \in \{3, \dots, B - 2\}$ and suppose inductively that $s_{q-1} \le U_{q-1}$. Since $U_q = U_{q-1} + 1$,

$$\begin{aligned} s_{q-1} + 1 &\le U_{q-1} + 1 \\ &= U_q \\ &= T - B + q. \end{aligned}$$

Thus the interval between the lower and upper guards in equation 78 is nonempty. The rounded value $Q_{\text{even}}(\widetilde{s}_q)$ is an integer, and both guards preserve integrality, so s_q is an integer satisfying

$$s_{q-1} + 1 \le s_q \le U_q. \tag{80}$$

This proves by induction that every interior offset is strictly larger than its predecessor and leaves the required number of unused integer stages. At the final interior index, if it exists,

$$s_{B-2} \le U_{B-2} = T - 2 < T - 1 = s_{B-1}.$$

When the interior range is empty, $B = 4$, and the same terminal inequality follows directly from $s_2 = 2 \le T - 2$. Consequently,

$$0 = s_0 < s_1 = 1 < s_2 = 2 < \cdots < s_{B-1} = T - 1$$

contains exactly B distinct integer offsets. The mapping $a_{q+1} = T - s_q$ reverses their order and therefore gives exactly B distinct solver-stage anchors containing $T, T - 1, T - 2$, and 1.

The full-grid boundary case is also contained in the same construction. If $B = T$, then $U_q = q$. Starting from $s_2 = 2$, equation 80 forces $s_q = q$ at every interior index, and the terminal condition gives $s_{T-1} = T - 1$. Hence every solver stage is an anchor.

For completeness, the realized certificate applies to any strictly increasing integer mesh satisfying $s_0 = 0, s_1 = 1, s_2 = 2, s_{B-1} = T - 1$, independently of how it was obtained. Define $\widehat{h}_j = s_j - s_{j-1} > 0$. By equation 79,

$$\frac{\widehat{h}_{j+1}}{\widehat{h}_j} \le \widehat{\Gamma}(\mathcal{A}) \qquad (2 \le j < B - 1).$$

Applying the cell calculation in Appendix C.11 at every eligible anchor gives

$$\lambda_j^{(s)} \le \widehat{\Gamma}(\mathcal{A}), \qquad \lambda_j^{(s)}(1+\lambda_j^{(s)}) \le \widehat{\Gamma}(\mathcal{A})[1+\widehat{\Gamma}(\mathcal{A})],$$

and, because $0 \le \sin^2\theta_j \le 1$,

$$R_j^{(s)} \le \widehat{\Gamma}(\mathcal{A})^2, \qquad \omega_j^{(s)} \ge \frac{1}{1+\widehat{\Gamma}(\mathcal{A})^2}.$$

This proof uses the realized gaps themselves and therefore remains valid when rounding changes the continuous geometric ratios.

C.14 Comparison under a Bounded Time Warp

Suppose a monotone time coordinate can be written as $t = \psi(s)$, with $0 < m \le \psi'(s) \le M_\psi < \infty$ over the relevant intervals. For a decreasing solver time, use $t = -\sigma$; this leaves the ratio in equation 3 unchanged. Let r_0, r_1 be the progress endpoints of a skipped raw interval after s_j, and let s_{j-1}, s_j be the latest anchors. The two midpoint-based ratios are

$$\rho^{(s)} = \frac{r_0 + r_1 - 2s_j}{2(s_j - s_{j-1})},$$
$$\rho^{(t)} = \frac{\psi(r_0) + \psi(r_1) - 2\psi(s_j)}{2[\psi(s_j) - \psi(s_{j-1})]}.$$

All increments in the decomposition of the numerator as

$$[\psi(r_0) - \psi(s_j)] + [\psi(r_1) - \psi(s_j)]$$

are nonnegative. For any $a \le b$, the fundamental theorem of calculus and the derivative bounds give

$$m(b-a) \le \psi(b) - \psi(a) = \int_a^b \psi'(u)\,du \le M_\psi(b-a).$$

Apply this inequality separately to the two numerator increments. With

$$N_s = (r_0 - s_j) + (r_1 - s_j), \qquad N_t = [\psi(r_0) - \psi(s_j)] + [\psi(r_1) - \psi(s_j)],$$

we obtain

$$mN_s \le N_t \le M_\psi N_s.$$

For the denominator increments

$$D_s = s_j - s_{j-1}, \qquad D_t = \psi(s_j) - \psi(s_{j-1}),$$

the same argument gives

$$mD_s \le D_t \le M_\psi D_s.$$

Since all denominators are positive,

$$\rho^{(t)} = \frac{N_t}{2D_t}$$
$$\ge \frac{mN_s}{2M_\psi D_s} = \frac{m}{M_\psi}\rho^{(s)},$$
$$\rho^{(t)} \le \frac{M_\psi N_s}{2mD_s} = \frac{M_\psi}{m}\rho^{(s)}.$$

Therefore

$$\frac{m}{M_\psi}\rho^{(s)} \le \rho^{(t)} \le \frac{M_\psi}{m}\rho^{(s)}. \tag{81}$$

In particular, a logical-progress gap-ratio certificate Γ transfers to at most $(M_\psi/m)\Gamma$ in physical solver time under this additional bounded-distortion assumption. Combining this with equation 7 gives the physical-time floor

$$\omega \ge \frac{1}{1+(M_\psi\Gamma/m)^2}.$$

Without a bound on the time warp, no such uniform transfer follows from logical geometric spacing alone.

C.15 Regularization and Degenerate Innovations

Let

$$\Delta v = \Delta v_q, \qquad \Delta v^- = \Delta v_q^-, \qquad S = \|\Delta v\|_2^2, \qquad P = \left\|\Delta v^-\right\|_2^2, \qquad D = \left\langle \Delta v, \Delta v^- \right\rangle.$$

When $P > 0$, the ideal least-squares residual is

$$r_0 = \Delta v - \frac{D}{P}\Delta v^-, \qquad N_0 = \|r_0\|_2^2 = S - \frac{D^2}{P}.$$

The regularized residual uses

$$r_{\epsilon_p} = \Delta v - \frac{D}{P + \epsilon_p}\Delta v^-. \tag{82}$$

Expanding its squared norm one term at a time gives

$$\begin{aligned} N_{\epsilon_p} &= \left\|r_{\epsilon_p}\right\|_2^2 \\ &= \|\Delta v\|_2^2 - \frac{2D}{P+\epsilon_p}\left\langle \Delta v, \Delta v^- \right\rangle + \frac{D^2}{(P+\epsilon_p)^2}\left\|\Delta v^-\right\|_2^2 \\ &= S - \frac{2D^2}{P+\epsilon_p} + \frac{D^2 P}{(P+\epsilon_p)^2} \\ &= S - \frac{D^2(P+2\epsilon_p)}{(P+\epsilon_p)^2}. \end{aligned}$$

Subtracting the ideal residual energy yields the exact identity

$$N_{\epsilon_p} - N_0 = \frac{D^2\epsilon_p^2}{P(P+\epsilon_p)^2} \geq 0. \tag{83}$$

Thus regularizing the projection denominator can only increase the residual energy relative to exact projection.

The finite weight is itself an exact minimizer. Define

$$\mathcal{J}_\epsilon(\omega) = S(1-\omega)^2 + (\rho^2 N_{\epsilon_p} + \epsilon_s)\omega^2.$$

Expanding and differentiating gives

$$\mathcal{J}'_\epsilon(\omega) = 2[S + \rho^2 N_{\epsilon_p} + \epsilon_s]\omega - 2S,$$

and

$$\mathcal{J}''_\epsilon(\omega) = 2[S + \rho^2 N_{\epsilon_p} + \epsilon_s].$$

This is strictly positive unless all three nonnegative terms vanish. Hence the unique minimizer whenever the denominator is positive is

$$\omega_\epsilon = \frac{S}{S + \rho^2 N_{\epsilon_p} + \epsilon_s}. \tag{84}$$

This shows precisely that ϵ_s adds a small numerical ridge to the transported-turning penalty.

The degenerate cases require an explicit convention only when the displayed ratio becomes $0/0$. If $\Delta v = 0$, then $S = N_{\epsilon_p} = 0$. When $\epsilon_s > 0$, the formula gives $\omega_\epsilon = 0$; when $\epsilon_s = 0$, every ω has the same objective and the implementation sets $\omega_\epsilon = 0$. In both cases the correction $\omega_\epsilon \Delta v$ is zero. If $\Delta v^- = 0$, then $P = D = 0$, so $r_{\epsilon_p} = \Delta v$: without a previous line, the full latest innovation is treated as unsupported. If $\Delta v^- \neq 0$ and $\epsilon_p, \epsilon_s \downarrow 0$, then $N_{\epsilon_p} \to N_0 = S\sin^2\theta$ and $\omega_\epsilon \to 1/(1+\rho^2\sin^2\theta)$.

A backend that regularizes the direct angle instead may use

$$\sin^2\widetilde{\theta}_{\epsilon_a} = 1 - \frac{D^2}{SP + \epsilon_a}.$$

Cauchy-Schwarz gives $0 \le D^2 \le SP$, so this quantity lies in $[0, 1]$. For $S, P > 0$, its exact relation to the ideal statistic is

$$\sin^2 \widetilde{\theta}_{\epsilon_a} = \sin^2 \theta + (1 - \sin^2 \theta) \frac{\epsilon_a}{SP + \epsilon_a}. \tag{85}$$

Both finite-precision forms converge to the same ideal rule as their regularizers vanish, but they are not algebraically identical at fixed positive regularization.

Fixed absolute regularizers break the exact scale invariance of equation 2 and are outside the continuum theorem. Dot products and norm accumulations should use sufficient precision, and a direct-angle implementation should clamp its final statistic to $[0, 1]$ to suppress roundoff outside the mathematical range. All invariance and continuum claims in the paper concern the zero-regularization equations; reported inference uses the finite rule of its stated backend.

C.16 Full-Grid Error Propagation

For an exact anchor the supplied field is exact at the current accelerated state. Define the additional field defect by

$$\varepsilon_n = \widetilde{v}_n - V_\theta(\widetilde{x}_n, \sigma_n, c),$$

so that $\varepsilon_n = 0$ at every anchor. For a solver with history, let X_n denote the augmented numerical state and let $\pi X_n = x_n$ extract the latent.

Proposition 9 (Full-grid perturbation bound)**.** *Let $\mathcal{U}_n(X, v)$ be the solver update and*

$$F_n(X) = \mathcal{U}_n(X, V_\theta(\pi X, \sigma_n, c)).$$

Assume

$$\|F_n(X) - F_n(Y)\| \le \alpha_n \|X - Y\|, \qquad \|\mathcal{U}_n(X, v) - \mathcal{U}_n(X, w)\| \le \beta_n \|v - w\|_2,$$

with $\alpha_n, \beta_n \ge 0$. If the exact-field and GeoShrink runs have identical initial solver states, then

$$\|\widetilde{X}_0 - X_0^\star\| \le \sum_{n=1}^{T} \beta_n \|\varepsilon_n\|_2 \prod_{k=1}^{n-1} \alpha_k. \tag{86}$$

If $h_n = |\sigma_n - \sigma_{n-1}|$, $\alpha_n \le 1 + Lh_n$, and $\beta_n \le C_{\mathcal{D}} h_n$, then

$$\|\widetilde{X}_0 - X_0^\star\| \le C_{\mathcal{D}} e^{L \sum_n h_n} \sum_{n=1}^{T} h_n \|\varepsilon_n\|_2. \tag{87}$$

If latent extraction is nonexpansive and the exact-field full-grid method has latent discretization error E_{grid}, then

$$E_{\mathrm{GeoShrink}} \le E_{\mathrm{grid}} + C_{\mathcal{D}} e^{L \sum_n h_n} \sum_n h_n \|\varepsilon_n\|_2. \tag{88}$$

Proof of Proposition 9. Let $e_n = \|\widetilde{X}_n - X_n^\star\|$. At stage n, insert the exact-field update evaluated at the GeoShrink solver state:

$$\begin{aligned} e_{n-1} &\le \|\mathcal{U}_n(\widetilde{X}_n, \widetilde{v}_n) - \mathcal{U}_n(\widetilde{X}_n, V_\theta(\pi \widetilde{X}_n, \sigma_n, c))\| \\ &\quad + \|F_n(\widetilde{X}_n) - F_n(X_n^\star)\| \\ &\le \beta_n \|\varepsilon_n\|_2 + \alpha_n e_n. \end{aligned}$$

The initial states coincide, so $e_T = 0$. Repeated substitution in descending stage order can be written explicitly. Put $b_n = \beta_n \|\varepsilon_n\|_2$. The first two substitutions are

$$\begin{aligned} e_0 &\le b_1 + \alpha_1 e_1 \\ &\le b_1 + \alpha_1 b_2 + \alpha_1 \alpha_2 e_2. \end{aligned}$$

We claim that after substituting through stage m,

$$e_0 \le \sum_{n=1}^{m} b_n \prod_{k=1}^{n-1} \alpha_k + e_m \prod_{k=1}^{m} \alpha_k. \tag{89}$$

For $m = 1$, this is the one-step recurrence, with the empty product in the first term equal to one. Suppose it holds for $m < T$. The next recurrence is $e_m \le b_{m+1} + \alpha_{m+1} e_{m+1}$. Multiplying it by the nonnegative product $\prod_{k=1}^{m} \alpha_k$ and substituting into equation 89 gives

$$\begin{aligned} e_0 &\le \sum_{n=1}^{m} b_n \prod_{k=1}^{n-1} \alpha_k + b_{m+1} \prod_{k=1}^{m} \alpha_k + e_{m+1} \prod_{k=1}^{m+1} \alpha_k \\ &= \sum_{n=1}^{m+1} b_n \prod_{k=1}^{n-1} \alpha_k + e_{m+1} \prod_{k=1}^{m+1} \alpha_k. \end{aligned}$$

This proves the claim by induction. Taking $m = T$ and using $e_T = 0$ gives

$$e_0 \le \sum_{n=1}^{T} \beta_n \left\|\varepsilon_n\right\|_2 \prod_{k=1}^{n-1} \alpha_k,$$

which is equation 86. If $\alpha_n \le 1 + L h_n$ and $\beta_n \le C_{\mathcal{D}} h_n$, then

$$\begin{aligned} \prod_{k=1}^{n-1} \alpha_k &\le \prod_{k=1}^{n-1} (1 + L h_k) \\ &\le \prod_{k=1}^{n-1} e^{L h_k} \\ &= \exp\left(L \sum_{k=1}^{n-1} h_k \right) \\ &\le e^{L \sum_{k=1}^{T} h_k}, \end{aligned}$$

where $1 + x \le e^x$ was applied to each nonnegative $x = L h_k$. Substitution into equation 86, together with $\beta_n \le C_{\mathcal{D}} h_n$, gives

$$\begin{aligned} e_0 &\le \sum_{n=1}^{T} C_{\mathcal{D}} h_n \left\|\varepsilon_n\right\|_2 e^{L \sum_k h_k} \\ &= C_{\mathcal{D}} e^{L \sum_k h_k} \sum_{n=1}^{T} h_n \left\|\varepsilon_n\right\|_2, \end{aligned}$$

which proves equation 87.

For the final error decomposition, let x_0 denote the endpoint of the continuous reference dynamics and let $E_{\mathrm{grid}} = \left\|\pi X_0^\star - x_0\right\|_2$. The triangle inequality and nonexpansiveness of π give

$$\begin{aligned} \left\|\pi \widetilde{X}_0 - x_0\right\|_2 &\le \left\|\pi \widetilde{X}_0 - \pi X_0^\star\right\|_2 + \left\|\pi X_0^\star - x_0\right\|_2 \\ &\le \left\|\widetilde{X}_0 - X_0^\star\right\|_2 + E_{\mathrm{grid}}. \end{aligned}$$

Applying equation 87 to the first term proves equation 88.

The augmented-state formulation is required when the numerical update depends on stored fields or past latent states. An update written only as a function of the current latent can otherwise omit a source of perturbation. The special constants $1 + L h_n$ and $C_{\mathcal{D}} h_n$ are assumptions to verify for the chosen method and state norm; they do not follow automatically from introducing an augmented state.

When the special constants are deterministic and the defects have finite second moments, weighted Cauchy-Schwarz also gives, with $S = \sum_n h_n$,

$$\mathbb{E}\|\widetilde{X}_0 - X_0^\star\|^2 \le C_{\mathcal{D}}^2 e^{2LS} S \sum_n h_n \, \mathbb{E}\left\|\varepsilon_n\right\|_2^2. \tag{90}$$

Indeed, weighted Cauchy-Schwarz gives, pathwise,

$$\begin{aligned}\left(\sum_n h_n \left\|\varepsilon_n\right\|_2\right)^2 &= \left(\sum_n \sqrt{h_n}\,\sqrt{h_n}\left\|\varepsilon_n\right\|_2\right)^2 \\ &\le \left(\sum_n h_n\right)\left(\sum_n h_n \left\|\varepsilon_n\right\|_2^2\right) \\ &= S\sum_n h_n \left\|\varepsilon_n\right\|_2^2 .\end{aligned}$$

Squaring equation 87, applying this bound, and then taking expectations gives

$$\begin{aligned}\mathbb{E}\left\|\widetilde{X}_0 - X_0^\star\right\|_2^2 &\le C_{\mathcal{D}}^2 e^{2LS}\mathbb{E}\left(\sum_n h_n \left\|\varepsilon_n\right\|_2\right)^2 \\ &\le C_{\mathcal{D}}^2 e^{2LS} S\sum_n h_n \mathbb{E}\left\|\varepsilon_n\right\|_2^2 ,\end{aligned}$$

where the finite sum is exchanged with expectation by linearity. This is equation 90. This connects local field MSE to an upper bound on deviation from the full-grid trajectory under the stated stability conditions. It is not a global optimality result for the retention policy: different policies change later states, histories, and target innovations.