When Is a Conditional Path Already Optimal? A Criterion for Flow Matching on Lie Groups

Anonymous authors
Paper under double-blind review

Abstract

Under kinetic optimality, flow matching does not leave the conditional path free: that path is the geodesic of the metric the loss induces. On $\mathrm{SE}(3)$, where models now generate protein backbones, molecular poses and robot actions, that geodesic generally requires a boundary-value solve at a cost comparable to a training step, so implementations substitute one of three closed-form constructions. Yet how far the substitute sits from the curve it replaces has not been reported. We give that deviation in closed form for group interpolation: second order in the endpoint amplitude, with leading term $\frac{\tau(1-\tau)}{2}\Lambda(u)$. A deviation alone does not identify which constructions need changing, so we also prove a criterion for when it vanishes. A factorized construction solves the geodesic equation of a block metric $\mathrm{diag}(J, \beta I)$ exactly when the endpoint rotation axis is an eigenvector of J. The condition is readable off the loss, and every body-frame-invariant loss with a block-diagonal Gram operator satisfies it by Schur's lemma. *For a given axis, what decides the matter is alignment rather than anisotropy*: rotating a metric's eigenbasis at fixed eigenvalues takes the construction from solving the geodesic equation exactly to departing from it. Reading six released loss configurations at source level, in the coordinates where each loss is computed, we find four that already satisfy the criterion, without having been built for it. Where it holds, the departure stays at the solver floor and does not grow with amplitude. Where it fails, the construction sits at second order from the geodesic, and each level of a closed-form correction removes one further order. Whether the departure is zero is therefore decidable before any training. On a robot action head the criterion accepts, training instead on the geodesic of a frozen approximation to its metric changes the error on training episodes by less than 0.3 percent, although the two paths differ by 3.7 percent of that error in translation and 5.9 percent in rotation.

1 Introduction

Under kinetic optimality, flow matching (Lipman et al., 2023; Albergo & Vanden-Eijnden, 2023; Liu et al., 2023b) does not leave its conditional path free: that path is the curve minimising the conditional action (Shaul et al., 2023; Chen & Lipman, 2024). On $\mathrm{SE}(3)$ that curve generally has no closed form. The group admits no bi-invariant metric, and no left-invariant metric on it has all one-parameter subgroups as geodesics (Žefran et al., 1999), so the geodesic of whichever metric is in force can in general be obtained only by solving a boundary-value problem numerically. Implementations therefore substitute one of three closed-form constructions, used by protein backbone models (Yim et al., 2023a;b; Bose et al., 2024), robot action heads (Black et al., 2024; Shukor et al., 2025) and generators of molecular and crystal structure (Cao et al., 2025; Miller et al., 2024). Yet how far any of them sits from the curve it replaces has not been reported.

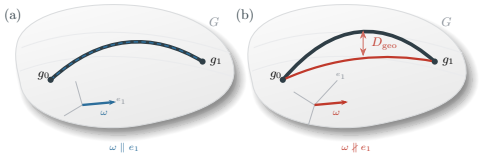


Figure 1: Schematic on $\mathrm{SO}(3)$. A conditional path (coloured) and the minimising curve (dark), for the same endpoints and eigenvalues. **(a)** The eigenbasis (grey) is aligned with the endpoint axis ω, and the two coincide. **(b)** It has turned, and they separate by D_{geo}.

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.

We report that deviation. For every left-invariant metric M, the group interpolation departs from the minimising curve by $\frac{\tau(1-\tau)}{2}\Lambda(u)$ at second order in the endpoint amplitude, where $\Lambda(x) = M^{-1}\mathrm{ad}_x^*(Mx)$ is the generator of the geodesic equation and M is read off the training loss (Appendix C.8). That is one order larger than the difference between the group and algebra interpolations, which meet at third order (Theorem 2): choosing between those two matters less than choosing whether to use either. But a deviation alone does not identify which constructions need changing, so we derive a condition on the metric that settles the question without measuring anything.

The condition is that a factorized construction solves the geodesic equation of a block metric $\mathrm{diag}(J, \beta I)$ exactly when the endpoint rotation axis is an eigenvector of J. Isotropy of J is the special case in which every axis is an eigenvector, so the construction is optimal for every endpoint pair. For a given axis, what separates the cases is therefore alignment rather than anisotropy: rotating the eigenbasis at fixed eigenvalues takes the construction from solving the geodesic equation exactly to departing from it (Figure 1). A loss invariant under a change of body frame makes J a multiple of the identity and satisfies the condition whenever its Gram operator is block-diagonal.

Where the condition is satisfied, the released construction is already the geodesic and requires no change. Where it is not, we give the leading term of the gap in closed form (Theorem 1, Proposition 7). Subtracting it removes one order, and for group interpolation, subtracting the next removes a second. Each subtraction costs a fixed number of Lie-algebra operations and one exponential, independent of the endpoints and the number of integration steps, and requires no boundary-value solve.

released implementation	induced metric	geodesic
EquiGraspFlow	isotropic	✓
DiffDock	per-block scaling	✓
FrameFlow, $t < \frac{1}{4}$	chart-induced	✓
FrameFlow, $t > \frac{1}{4}$	coupled	✗
SmolVLA	flat pullback	✓
3D FlowMatch	not quadratic	—

Table 1: Whether each released construction already solves the geodesic equation of the metric its loss induces (Lim et al., 2025; Corso et al., 2024; Yim et al., 2023a; Shukor et al., 2025; Gkanatsios et al., 2025). FrameFlow appears twice because its checkpoint adds an auxiliary loss above $t = \frac{1}{4}$. DiffDock is score-based and is read from its loss weights.

Evaluating six released configurations at source level, in the coordinates where each loss is computed, we find four already satisfying the condition, one violating it, and one inducing no quadratic metric because its loss is a weighted sum of absolute values (Table 1). Measuring them, we find the four keep their gap at the solver floor across sweeps spanning $32\times$ and $64\times$ in amplitude, so the criterion is exact rather than asymptotic. The one that violates it departs by about 9×10^{-2} Å on pairs where the same model below its gate sits below the floor, and under a constant metric each correction removes one order more. Because both the criterion and the departure are closed form, the chain from a loss to a predicted deviation is algebraic throughout: on the action head, against a frozen approximation to its metric, it returns a per-increment gap that composes over fifty actions to 4.5 mm (Appendix D.7), with no model trained at any point.

Our contributions are as follows. **(i)** We show that group interpolation departs from the curve it replaces by $\frac{\tau(1-\tau)}{2}\Lambda(u) + O(\|u\|^3)$ under any left-invariant metric, and give a correction that removes one order per level at fixed cost. **(ii)** We prove a criterion, readable off the loss and satisfied by every block-diagonal body-frame-invariant one, that decides when the departure vanishes, and show that for a given axis alignment rather than anisotropy separates the cases. **(iii)** We prove that under a Haar or anchored source, a model's operating amplitude is set by its source distribution and not by its data, so that the scale of the departure is fixed before any data is seen. **(iv)** We test all three against six released configurations and four trained policies. Four of the configurations are already optimal, and the criterion holds exactly rather than asymptotically. On the action head, training on a nearby path instead moves the policy's error on training episodes by less than 0.3 percent.

2 Preliminaries

Lie groups and left-invariant metrics. Let G, the domain over which the generative model is learned, be a finite-dimensional Lie group with Lie algebra $\mathfrak{g}$. A curve $g(\tau)$ in G has body velocity $\xi = (g^{-1}\dot{g})^\vee \in \mathfrak{g}$, an element of the algebra rather than of the group, and the exponential $\exp : \mathfrak{g} \to G$ carries it back to the group while its local inverse $\log$ recovers it. The adjoint $\mathrm{ad}_x y = [x, y]$ acts on $\mathfrak{g}$ and its dual ad_x^* on $\mathfrak{g}^*$. A left-invariant metric is a symmetric positive-definite Gram operator M on $\mathfrak{g}$, constant in body coordinates, under which a curve has conditional action $E(g) =$

$\frac{1}{2}\int_0^1 \|\xi(\tau)\|_M^2\, d\tau$. The group of interest throughout is $\mathrm{SE}(3) = \mathrm{SO}(3) \ltimes \mathbb{R}^3$, whose elements are pairs (R, p) and whose algebra elements are pairs (ω, v). Marsden & Ratiu (1999) give a fuller treatment of the geometry.

The conditional path, and the constructions in use. Flow matching supervises a velocity field along a conditional path joining a source sample to a data sample. Under kinetic optimality that path is the curve minimising E, the constant-speed minimising geodesic g_{riem}, which in body coordinates solves the Euler–Poincaré equation

$$\dot{\xi} = \Lambda(\xi), \qquad \Lambda(x) := M^{-1}\mathrm{ad}_x^*(Mx). \tag{1}$$

This is an initial-value problem while the path is specified by its two endpoints, so obtaining g_{riem} in general means solving a boundary-value problem. Implementations therefore substitute one of three constructions available in closed form,

$$g_{\mathrm{geo}}(\tau) = g_0 \exp(\tau u), \qquad u = \log(g_0^{-1} g_1), \tag{2}$$

$$g_{\mathrm{lin}}(\tau) = \exp\big((1-\tau)\xi_0 + \tau\xi_1\big), \qquad g_i = \exp(\xi_i), \tag{3}$$

$$g_{\mathrm{fact}}(\tau) = \big(R_0 \exp(\tau\hat{\omega}),\, (1-\tau)p_0 + \tau p_1\big), \qquad \omega = \log(R_0^\top R_1), \tag{4}$$

adopted respectively by screw-motion flows, by pose-vector action heads, and by losses that treat rotation and translation separately. All three join g_0 to g_1, and in general none solves equation 1.

3 Method

We determine how far each of equation 2–equation 4 sits from g_{riem}, under what condition that deviation is zero, and what closes it when it is not. Every statement below holds for endpoint pairs away from the cut locus, at amplitudes small enough for the expansions to hold. Appendix C.1 states the assumptions in full and Appendix C.13 quantifies both conditions.

3.1 The cost of the substitution

Two quantities separate the constructions of equation 2–equation 4 from the curve they replace and from each other, and this section gives the leading term of each in closed form. They are objects of different kinds. The first is the logarithm of a group element and is invariant under left translation of both endpoints, so it measures a departure from optimality. The second is a difference of two algebra elements and is the quantity a coordinate-based implementation regresses against, so it measures a disagreement between two constructions, both of which may be suboptimal. Explicitly,

$$D_{\mathrm{geo}}(\tau) = \log\big(g_{\mathrm{riem}}(\tau)^{-1} g_{\mathrm{geo}}(\tau)\big), \qquad D(\tau) = \log g_{\mathrm{geo}}(\tau) - \big((1-\tau)\xi_0 + \tau\xi_1\big). \tag{5}$$

The first is measured against the relative amplitude $\|u\|$, the second against the endpoint amplitude $\epsilon = \max(\|\xi_0\|, \|\xi_1\|)$, since each is the natural parameter of the object it describes.

Theorem 1 (Second-order optimality gap)**.** *Under the assumptions of Appendix C.1, uniformly for* $\tau \in [0,1]$*,*

$$\boxed{D_{\mathrm{geo}}(\tau) = \frac{\tau(1-\tau)}{2}\,\Lambda(u) + O(\|u\|^3).} \tag{6}$$

Proof. Let $\mathcal{E}(a) = \log\big(h_a(1)\big)$, where h_a solves equation 1 from the identity with initial body velocity a. Since $D\mathcal{E}(0) = \mathrm{Id}$, the inverse function theorem makes $\mathcal{E}$ a local diffeomorphism, so the boundary-value solution expands in u. As Λ is quadratic, substituting into equation 1 gives $\xi(\tau) = u + a_2 + \tau\Lambda(u) + O(\|u\|^3)$, and the endpoint condition forces $a_2 = -\frac{1}{2}\Lambda(u)$. A Magnus expansion, whose first bracket is $O(\|u\|^3)$, gives $\log(g_0^{-1} g_{\mathrm{riem}}(\tau)) = \tau u - \frac{\tau(1-\tau)}{2}\Lambda(u) + O(\|u\|^3)$, and one Baker–Campbell–Hausdorff step against τu yields equation 6 (Appendix C.2). □

Theorem 2 (Third-order deviation)**.** *Let* ξ_0 *and* ξ_1 *lie where the Baker–Campbell–Hausdorff series converges for the products in equation 2 and equation 3, and set*

$$A = [\xi_0, [\xi_0, \xi_1]], \qquad B = [\xi_1, [\xi_1, \xi_0]]. \tag{7}$$

Then, uniformly for $\tau \in [0,1]$*,*

$$\boxed{D(\tau) = -\frac{\tau(1-\tau)}{12}\,(A + B) + O(\epsilon^5).} \tag{8}$$

Proof. Expanding both constructions to degree five in the free Lie algebra leaves nothing at degree two, since those terms cancel by antisymmetry alone, without the Jacobi identity, so the cancellation belongs to the series rather than to any particular algebra. At degree three the terms collect into $-\frac{\tau(1-\tau)}{12}(A+B)$. Degree four vanishes identically, because inverting each endpoint inverts both constructions, so D is odd in (ξ_0, ξ_1). Appendix C.3 verifies the expansion in exact arithmetic. □

Theorem 2 covers the two logarithm-based constructions, which differ from each other one order below their common distance from the geodesic, so choosing between equation 2 and equation 3 matters less than choosing whether to use either. The factorized construction is not of that family.

Proposition 3 (The factorized construction sits at second order)**.** *On* $\mathrm{SE}(3)$*, uniformly for* $\tau \in [0,1]$*,*

$$\log\big(g_{\mathrm{geo}}(\tau)^{-1} g_{\mathrm{fact}}(\tau)\big) = \Big(0,\ \tfrac{\tau(1-\tau)}{2}\,\omega \times R_0^\top (p_1 - p_0)\Big) + O(\epsilon^3). \tag{9}$$

The rotational component vanishes identically, not merely to leading order.

Proof. Both constructions have rotational part $R_0 \exp(\tau\hat{\omega})$, so the difference is purely translational. On the translational block equation 2 advances by $V(\tau\omega)\,\tau\,V(\omega)^{-1}$ and equation 4 by τI, where V is the translational Jacobian of Appendix C.7. Expanding $V = I + \frac{1}{2}\hat{\omega} + O(\theta^2)$ gives $\tau I - V(\tau\omega)\,\tau\,V(\omega)^{-1} = \frac{\tau(1-\tau)}{2}\hat{\omega} + O(\theta^2)$, and applying this to $R_0^\top(p_1 - p_0)$ yields equation 9. □

The profile $\tau(1-\tau)/2$ is the same as in Theorem 1. Here it comes from the linear term of V, which is what mixes rotation into translation. Figure 2 shows the resulting arrangement, with equation 2 and equation 3 within $O(\epsilon^3)$ of each other and equation 4 apart from both at $O(\epsilon^2)$. On $\mathrm{SE}(3)$ with M block isotropic, $\Lambda(\omega, v) = (0, v \times \omega)$ is a pure translation, so the gap of equation 2 has no rotational component either. Anisotropy within the rotational block is what makes Λ turn the rotation, and §3.2 shows that not every anisotropy does. The gap of equation 2 vanishes identically under a bi-invariant metric, and both gaps vanish on an abelian algebra, neither of which holds on $\mathrm{SE}(3)$ (Žefran et al., 1999), so no choice of M removes the second-order gap for every endpoint pair (Appendix C.4).

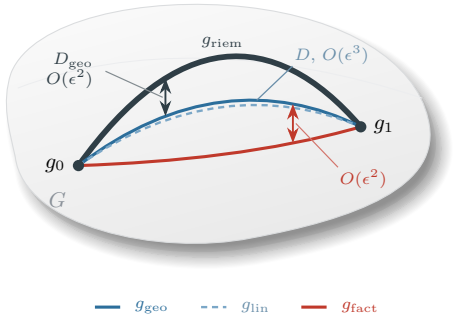


Figure 2: Schematic. Group and algebra interpolation differ at third order (Theorem 2) and sit at second order from the geodesic (Theorem 1). The factorized construction differs from both at second order (Proposition 3).

3.2 When the substitution is free

Theorem 1 measures a cost. This section identifies when that cost is absent. The gap of equation 2 is generically second order under every left-invariant metric on $\mathrm{SE}(3)$, whereas the gap of equation 4 is exactly zero on a class of metrics that includes most of those in use, and membership of that class is decidable from the loss. Under $M = \mathrm{diag}(J, \beta I)$ the translational term in body coordinates is $\beta\|\dot{p}\|^2$, so the metric splits, $(\mathrm{SE}(3), M)$ is a Riemannian product, and equation 4 is the product geodesic provided each factor path minimises its factor metric (Proposition 15, Appendix C.5). The translational factor satisfies this identically, which reduces the question to the rotational one.

Proposition 4 (Alignment criterion)**.** *Let* $M = \mathrm{diag}(J, \beta I_3)$ *with* J *a* constant *symmetric positive-definite matrix and* $\beta > 0$*. The factorized path equation 4 satisfies equation 1 if and only if* $\omega = 0$ *or* ω *is an eigenvector of* J*, and it does so for* every *endpoint pair if and only if* J *is isotropic.*

Proof. The body velocity of equation 4 is $\xi = (\omega, \nu)$ with ω constant and $\nu(\tau) = R(\tau)^\top w$ with $w = p_1 - p_0$, so $\dot{\omega} = 0$ and $\dot{\nu} = -\omega \times \nu$. Constancy of J removes the $\dot{J}$ term from equation 1, leaving $J\dot{\omega} = (J\omega) \times \omega$ and $\beta\dot{\nu} = \beta\,\nu \times \omega$. The second holds identically because $-\omega \times \nu = \nu \times \omega$, and the first reduces to $(J\omega) \times \omega = 0$, which is $J\omega \parallel \omega$. □

This is a condition for stationarity. Minimality requires in addition that the endpoints lie where the exponential map is a diffeomorphism, which under anisotropic J need not extend to $\|\omega\| = \pi$. The proposition also assumes J constant, and metrics in practical use often are not: a loss measured in a chart induces a block that varies along the path. When the chart is fixed, such a metric is a pullback

of the Euclidean one, for which the conclusion is exact (Collas et al., 2025), and when it moves with the path the metric is frozen at the point of evaluation (Appendix C.8).

Corollary 5 (Losses measured in a logarithm chart)**.** *Let a rotational loss measure its residual in the chart* $\Phi(R) = \log(R_0^\top R)$*, so that the metric it induces on body velocities is* $J(\tau) = \mathcal{J}_r(\phi)^{-\top}\mathcal{J}_r(\phi)^{-1}$ *with* $\phi = \Phi(R(\tau))$*. Then for every pair with* $\|\omega\| < \pi$ *the rotational factor of equation 4 is the unique minimising geodesic of that metric, and the same holds on* $\mathrm{SE}(3)$ *for* $(R, p) \mapsto (\Phi(R), p - p_0)$*.*

The proof is in Appendix C.6, together with the statement it rests on, that a pullback of the Euclidean metric along a chart onto a convex set is flat and its geodesics are the chart's straight lines. The criterion depends on the eigenvalues of J and on the alignment of ω with them, the corollary on the chart alone, so rotating the eigenframe at fixed eigenvalues separates the two (§5.2).

Proposition 6 (Body-frame invariance forces isotropic blocks)**.** *If the loss is invariant under a rotation* Q *of the body frame, which acts on body velocities by* $\xi \mapsto \mathrm{Ad}_Q\xi$*, then its rotational block satisfies* $Q^\top JQ = J$ *for every* $Q \in \mathrm{SO}(3)$*, hence* $J = \lambda I$*. A loss with this invariance and a block-diagonal Gram operator therefore satisfies the criterion of Proposition 4.*

Proof. Invariance of $\xi^\top M\xi$ under Ad_Q gives $\mathrm{Ad}_Q^\top M\,\mathrm{Ad}_Q = M$, whose rotational block is $Q^\top JQ = J$. A symmetric matrix commuting with every rotation is a multiple of the identity. A world-frame rotation $g \mapsto Qg$ leaves body velocities unchanged and so does not constrain M. □

Three classes of metric fall outside both conditions. **(i)** A coupling block between rotation and translation leaves a second-order departure, and rotational invariance alone does not exclude one. **(ii)** A metric $\mathrm{diag}(J, B)$ with B anisotropic adds $(B\nu) \times \nu$ to the rotational equation, and since $\nu(\tau)$ sweeps a cone about ω, this vanishes for every endpoint pair only when $B = \beta I_3$. **(iii)** A loss whose Hessian in body coordinates varies with the group element induces no left-invariant metric, and is judged by its pullback when its chart is fixed and by its frozen metric otherwise. Appendix C.8 sorts the released losses among these.

3.3 Closing the gap when it is not

Section 3.2 identifies the metrics under which equation 4 has no gap. For the rest, Theorem 1 supplies the leading term in closed form, and subtracting it removes one order. The construction is a post-multiplication, $g_{\mathrm{base}}(\tau)\exp(F_\ell(\tau))$, which keeps the endpoints fixed whenever $F_\ell(0) = F_\ell(1) = 0$ and leaves the base otherwise untouched. The profile of F_ℓ is then forced: the deviation to be cancelled carries $\tau(1-\tau)/2$ by Theorem 1, so F_1 carries the same factor with the opposite sign, and its generator is the difference between the generator of the geodesic and that of the base. The base must therefore satisfy $\dot\xi = \Lambda_0(\xi)$ for a fixed quadratic generator Λ_0. That admits equation 2 with $\Lambda_0 = 0$ and equation 4 with $\Lambda_0 = \Lambda_{\mathrm{iso}}$, and excludes equation 3, which can be reached only by replacing its base. About equation 2, the second level cancels the residual the first leaves, and its closed form requires the polar $D\Lambda$ of the generator,

Algorithm 1 Corrected conditional path

Require: g_0, g_1, M, τ, level $\ell \in \{1, 2\}$, base with generator Λ_0 and velocity ξ_{base} (group base if $\ell = 2$)

▷ Increment and generator

$u \leftarrow \log(g_0^{-1}g_1)$

$a \leftarrow M^{-1}\mathrm{ad}_u^*(Mu)$

▷ Corrector

$c \leftarrow -\tau(1-\tau)/2$

if $\ell = 1$ **then**

  $F \leftarrow c\,(a - \Lambda_0(u))$

  $\dot F \leftarrow -\frac{1-2\tau}{2}(a - \Lambda_0(u))$

else

  $N_2 \leftarrow D\Lambda(u)[a] - [u, a]$

  $d \leftarrow \tau(1-\tau)(1-2\tau)/12$

  $F \leftarrow c\,a + dN_2$

  $\dot F \leftarrow -\frac{1-2\tau}{2}a + \frac{1-6\tau+6\tau^2}{12}N_2$

end if

▷ Path and velocity

$g \leftarrow g_{\mathrm{base}}(\tau)\exp(F)$

$\xi_{\mathrm{corr}} \leftarrow \mathrm{Ad}_{\exp(-F)}(\xi_{\mathrm{base}}(\tau)) + \mathrm{dexp}_{-F}(\dot F)$

return g, ξ_{corr}

$$D\Lambda(u)[h] = M^{-1}\big(\mathrm{ad}_h^*(Mu) + \mathrm{ad}_u^*(Mh)\big), \quad N_2(u) = D\Lambda(u)\big[\Lambda(u)\big] - \big[u, \Lambda(u)\big]. \tag{10}$$

The corrected path and the conditional velocity field a regression consumes are

$$g_{\mathrm{corr}}^{(\ell)}(\tau) = g_{\mathrm{base}}(\tau)\exp\big(F_\ell(\tau)\big), \quad F_1 = c(\tau)\big(\Lambda(u) - \Lambda_0(u)\big), \quad F_2 = c(\tau)\Lambda(u) + d(\tau)N_2(u), \tag{11}$$

$$u_{\mathrm{corr}}(\tau) = \mathrm{Ad}_{\exp(-F_\ell)}(u) + \mathrm{dexp}_{-F_\ell}\big(\dot F_\ell\big), \qquad c = -\tfrac{\tau(1-\tau)}{2}, \qquad d = \tfrac{\tau(1-\tau)(1-2\tau)}{12}. \tag{12}$$

Equation equation 12 is the group-interpolation case, where $\xi_{\text{base}} \equiv u$; about a general base it reads $\mathrm{Ad}_{\exp(-F_\ell)}(\xi_{\text{base}}(\tau)) + \mathrm{dexp}_{-F_\ell}(\dot{F}_\ell)$ (Appendix C.11). Algorithm 1 states the procedure about a general base, with F_2 written about equation 2, where $\Lambda_0 = 0$. Differentiating a product whose exponent depends on τ produces dexp_{-F_ℓ} rather than the identity, so equation 12 is not $\mathrm{Ad}_{\exp(-F_\ell)}(u) + \dot{F}_\ell$, the difference being $-\frac{1}{2}[F_\ell, \dot{F}_\ell] + O(\|u\|^5)$. That bracket vanishes identically at level one, where F_1 and $\dot{F}_1$ are parallel. At level two it is $(c\dot{d} - d\dot{c})[\Lambda(u), N_2(u)]$, of fifth order, so the simpler expression departs from the field equation 11 generates only at that order.

Proposition 7 (Two corrections)**.** *Under the standing assumptions, with $\|u\|$ small enough:* ***(i)*** *about any admissible base, F_1 leaves a residual $O(\|u\|^3)$.* ***(ii)*** *About equation 2 the third-order residual has the closed form*

$$\log\big(g_{\text{corr}}^{(1)}(\tau)^{-1} g_{\text{riem}}(\tau)\big) = \frac{\tau(1-\tau)(1-2\tau)}{12}\, N_2(u) + O(\|u\|^4), \tag{13}$$

so F_2 leaves $O(\|u\|^4)$ and the corrected path agrees with the geodesic through degree three. ***(iii)*** *About equation 2, Equation equation 12 is the exact body velocity of equation 11.*

Proof of (i). Write $g_{\text{riem}} = g_{\text{base}} \exp(F)$, whose body velocity is $\xi_{\text{base}} + \dot{F} + \mathrm{ad}_{\xi_{\text{base}}} F + O(\|F\|^2)$, so $\delta\xi := \xi_{\text{riem}} - \xi_{\text{base}} = \dot{F} + O(\|u\|^3)$ since $F = O(\|u\|^2)$. Subtracting the two Euler–Poincaré equations and linearising, $\delta\dot{\xi} = \Lambda(u) - \Lambda_0(u) + O(\|u\|^3)$, the term $D\Lambda(u)[\delta\xi]$ being $O(\|u\|) \cdot O(\|u\|^2)$. Integrating twice with $F(0) = F(1) = 0$ returns F_1 (Appendix C.11). □

Where the loss can be changed, making the criterion hold is preferable to correcting the path: with both blocks isotropised equation 4 is the geodesic exactly, whereas each level of equation 11 leaves a residual that composition accumulates (Appendix C.14). Two levels are derived and no recurrence is claimed beyond them (Appendix C.11). Each level costs a fixed number of Lie-algebra operations and one exponential, independent of the endpoints and of the number of integration steps, and calls no boundary-value solver (Appendix C.12).

4 Related Work

Conditional paths on Lie groups. The conditional path is the central modelling choice in flow matching, and a large Euclidean literature chooses it by straightening, optimal-transport coupling or schedule reweighting (Liu et al., 2023b; Pooladian et al., 2023; Tong et al., 2024; Bansal et al., 2026). On manifolds the path is built from a premetric (Chen & Lipman, 2024), and on Lie groups from exponential curves (Sherry & Smets, 2025). On these spaces flow matching now generates protein backbones (Yim et al., 2023a; Bose et al., 2024), torsion angles (Jing et al., 2022; Lee & Kim, 2025), molecular conformers, crystals and Stiefel-valued structures (Cao et al., 2025; Miller et al., 2024; Cheng et al., 2025), and robot action chunks (Black et al., 2024; Shukor et al., 2025). In each, the path is set by convention, and equation 2–equation 4 coexist without a reported comparison. Most such models are equivariant under world-frame rotations, which leave body velocities unchanged and so do not constrain the metric the loss induces. Invariance of the loss under a change of body frame forces its rotational block to be isotropic (Proposition 6).

Straightening, and what it acts on. Sampling on $\mathrm{SE}(3)$ often takes hundreds of network evaluations (Chen et al., 2025), and a line of work straightens trajectories to reduce that count, through Riemannian optimal transport (Bose et al., 2024), quaternion rectification (Yue et al., 2025), and schedules chosen for fewer integration steps (Yim et al., 2023a). These act on the *marginal* field, curved because conditional paths cross in state space, and they act *after* averaging. Theorems 1 and 2 concern the conditional path itself, before averaging and independently of the coupling, so the two are complementary. Collas et al. (2025) observe that a metric pulled back from Euclidean space by a global diffeomorphism reduces the Riemannian case to the Euclidean one. Lemma 16 uses that fact to decide whether a construction already in use is optimal.

Where the geodesic is unavailable, and where paths break. On $\mathrm{SE}(3)$ the geodesic of a left-invariant metric generally has no closed form: the group admits no bi-invariant metric, no left-invariant metric has all one-parameter subgroups as geodesics (Žefran et al., 1999), and the kinematics literature generates optimal rigid-body motions only variationally (Žefran et al., 1998). A second difficulty is topological. Flow matching has been reported to underperform Riemannian diffusion on compact manifolds, with a conditional path discontinuous at the antipode (Lou et al.,

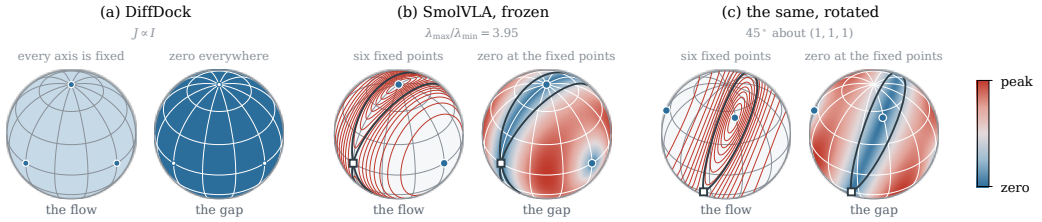


Figure 3: **The criterion is a fixed-point condition on a flow, and the gap follows.** Closed form, on the sphere of rotation axes. Left of each pair, orbits of $J\dot{\omega} = (J\omega) \times \omega$, with centres as dots, saddles as squares and separatrices heavy. Right, the leading-order gap of the factorized construction, proportional to $\|J^{-1}((J\omega) \times \omega)\|$ and normalised to its peak, which vanishes exactly at the fixed points. **(a)** An isotropic block fixes every axis. **(b)** SmolVLA's weights, frozen. **(c)** The same eigenvalues with the eigenbasis rotated by $45°$, which moves the fixed points with it. Anisotropy sets how fast the axis moves, not where it stops.

2023), the failure reproducing under different parameterisations (Zhu et al., 2025), and underfitting attributed to cut-locus discontinuity and shrinking support (Cheng et al., 2025; Stark et al., 2024). A spectral premetric on the sphere is degenerate at the antipode, so the flow is defined only *almost everywhere* (Chen & Lipman, 2024, Fig. 11). These works replace the probability path rather than measure what the discontinuity costs. Appendix A expands the comparison.

5 Experiments

Our experiments are organised around the criterion of §3.2. §5.1 reads the metric each released loss induces and records what the criterion decides. §5.2 tests whether the gap sits at the apparatus floor wherever the criterion holds, and whether alignment rather than anisotropy decides it. §5.3 measures how the gap scales with amplitude on either side of the criterion and what closing it costs. §5.4 follows a deviation on the action head through composition and into four trained policies. Full details are provided in Appendix D.

5.1 RQ1: which metric each loss induces, and what the criterion decides

Table 1 states what the criterion decides for six released configurations from five implementations, each acting on $\mathrm{SE}(3)$ or a factor of it. Four satisfy it, through three mechanisms. Two are isotropic: EquiGraspFlow's unweighted loss induces $M = cI$, and DiffDock's per-block schedules scale each block without changing the optimal path (Appendix D.3). FrameFlow below its gate is anisotropic but aligned, since its logarithm chart is anchored at the current point and its Jacobian has the endpoint axis as an eigenvector (Proposition 4). SmolVLA is anisotropic but flat, since its loss and path share one fixed chart and the path is the geodesic of the pullback metric (Lemma 16). Only FrameFlow above its gate violates the criterion, because the hat map of its backbone template couples the rotational and translational blocks. Centring the template removes that block, but it also changes the frame coordinates, so the factorized interpolation taken in the new coordinates is a different path. On the same pairs, centring lowers the gap by a factor of 32 under the released metric (Appendix D.1). 3D FlowMatch induces no quadratic form, so the criterion does not apply.

5.2 RQ2: the gap is at the floor wherever the criterion is satisfied

By Proposition 4 the gap vanishes exactly when the endpoint rotation axis is an eigenvector of the rotational block, and Figure 3 shows why. Along the geodesic the body angular velocity obeys $\dot{\xi} = \Lambda(\xi)$ and its direction moves, while the factorized construction holds it fixed, so the two agree at the fixed points of the flow. What decides whether an axis is fixed is alignment rather than anisotropy, and a released metric shows it directly. FrameFlow's frozen chart metric below its gate has an eigenvalue ratio of only 1.22 and is aligned by construction, and on $30\,000$ real protein pairs its gap is 1.8×10^{-13} Å, below the solver floor. Rotating its eigenbasis by $45°$ at fixed eigenvalues raises the gap on the same pairs to 3.3×10^{-2} Å, over three seeds. The same holds across forty-nine metrics that cross seven eigenvalue ratios with seven eigenbasis rotations. At a ratio of 10 the gap is below the floor when the eigenbasis is aligned and 5.2×10^{-2} in dimensionless units when it is rotated by $45°$, and over the whole grid the gap tracks the misalignment rather than the ratio, at rank correlation 0.93 against $\|\omega \times J\omega\|$ and 0.54 against $\lambda_{\max}/\lambda_{\min}$ (Appendix D.4).

Table 2: **Residual against the solved geodesic**, normalised by $\|u\|$, at $\tau = \frac{1}{4}$ over 300 pairs per rung, where $\|u\|$ is the $\mathfrak{se}(3)$ twist norm in degrees. *Top:* released configurations satisfying the criterion, with drift the largest over the smallest residual. *Bottom:* two constant metrics on random pairs, SmolVLA's frozen weights and FrameFlow's template, with algebra interpolation (alg.) and group interpolation corrected once and twice. Slopes 1, 2, 3 (bootstrap errors) are orders 2, 3, 4.

criterion satisfied	**sweep, $\|u\|$ (deg)**	**residual range**	**drift**
EquiGraspFlow	0.7–46	5.70–5.87e−12	1.03
DiffDock	3.6–115	1.47–6.55e−14	4.5
FrameFlow, $t < \frac{1}{4}$	3.6–115	0.71–1.91e−14	2.7

constant metric	**lvl**	**$\|u\|$ (deg)**							**slope**
SmolVLA weights, frozen		0.7	1.4	2.9	5.7	11	23	46	
	alg.	7.9e–4	1.6e–3	3.2e–3	6.3e–3	1.3e–2	2.5e–2	5.1e–2	**1.000**(.007)
	one	9.1e–7	3.6e–6	1.5e–5	5.9e–5	2.4e–4	9.7e–4	4.2e–3	**2.023**(.007)
	two	6.7e–9	5.4e–8	4.3e–7	3.4e–6	2.8e–5	2.2e–4	1.8e–3	**3.001**(.012)
FrameFlow template		3.6	7.2	14	29	57	86	115	
	alg.	2.0e–3	4.1e–3	8.2e–3	1.6e–2	3.2e–2	4.8e–2	6.4e–2	**0.992**(.007)
	one	1.4e–5	5.6e–5	2.2e–4	9.0e–4	3.6e–3	8.0e–3	1.4e–2	**1.997**(.013)
	two	1.3e–7	1.0e–6	8.2e–6	6.6e–5	5.3e–4	1.8e–3	4.2e–3	**2.998**(.022)

5.3 RQ3: WHERE THE CRITERION IS VIOLATED, WHAT THE GAP IS AND WHAT IT COSTS

The source distribution fixes the amplitude. Under a Haar source drawn independently of the data, or a source anchored to the data sample, the law of the relative rotation does not depend on the data (Proposition 20). For the five configurations with such a source, the medians match the analytic Haar or $\mathrm{IGSO3}(\sigma)$ value to within the sampling error of 3 000 draws, about one degree. The exception is the action head, whose source is a unit Gaussian in normalised increment coordinates rather than a draw on $\mathrm{SO}(3)$, and whose median is 3.46° over three seeds. The six fall into two clusters, the upper five spanning 117.1° to 132.4° and the nearest sitting 113.6° above the action head.

The residual is flat where the criterion holds and second order where it fails, and each correction removes one order. Against a numerically solved geodesic, the three residuals in the top panel of Table 2 stay between 7×10^{-15} and 6×10^{-12}, at or below the resolution of the apparatus, and none drifts by more than a factor of 4.5 across sweeps spanning $32\times$ and $64\times$ in amplitude (Figure 4a). EquiGraspFlow has the smallest drift of the three despite the largest residual: its metric is cI exactly, so the criterion holds by algebra rather than for the pairs drawn, and its 5.7×10^{-12} is the convergence of the solver. Under the two constant metrics of the bottom panel the fitted slopes are 1, 2 and 3 to within 0.03 (Figure 4b). Released configurations show the same separation when measured on shared pairs. FrameFlow, on 3 000 pairs over three seeds, departs by 8.7 to 9.3×10^{-2} Å above its gate and lies below the solver floor below it. SmolVLA, whose path is the geodesic of its pullback metric by Lemma 16, sits at 0.09 to 0.18 of the floor under that metric at every rung of its ladder and at least ten orders above it under the freeze (Appendix D.1). The slopes constrain the orders but not the coefficients, and the coefficients can be checked independently: writing C_1 and C_2 for the residuals after one and two levels, their ratio falls as $1/\epsilon$ if the orders are three and four, and under the template-coupled metric $C_1/C_2 = 6.79/\epsilon$ to within 0.56 percent across all seven rungs. The factorized path is the geodesic of the product metric these models declare (Yim et al., 2023a; Bose et al., 2024), but not of the coupled metric FrameFlow's loss induces above its gate.

The criterion determines the route, and both routes are affordable. The in-place correction requires a base satisfying $\dot{\xi} = \Lambda_0(\xi)$, which the algebra construction does not (Appendix C.11), so correcting it requires replacing the base first. Above FrameFlow's gate, on 1 200 pairs per seed under its released metric, replacing the factorized base by group interpolation and correcting once raises the gap by a factor of 14.6, while correcting the factorized base in place lowers it by a factor of 2.8. On the action head, under the frozen metric, replacing the algebra base raises the gap by a factor of 1.001, and one level then lowers it by a factor of 125. One closed-form level costs 38.0 μs per pair, and one solved geodesic costs 21.9 to 158.1 ms depending on the domain, a fifth of a training step at 3.4° and one and a half steps at 116° on forty cores. The exact route is usable in both regimes, and cheapest where the deviation it removes is smallest.

5.4 FROM THE GEOMETRY TO A TRAINED POLICY

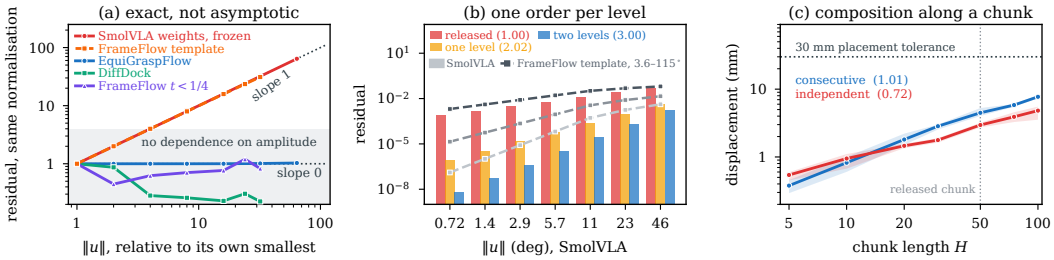


Figure 4: **The criterion is exact, each correction removes one order, and the deviation composes along a chunk.** **(a)** Each series is normalised to its value at the smallest $\|u\|$. Under the constant metrics of Table 2 algebra interpolation follows slope one, while the released configurations that satisfy the criterion stay flat. **(b)** Each correction level carries its fitted slope in brackets. Bars use SmolVLA's frozen weights, and dashed lines use FrameFlow's template, drawn at the same rung positions over its range of $3.6°$ to $115°$. **(c)** Curves show the positional separation at $\tau = \frac{1}{4}$ between the action head's released path and its frozen-metric geodesic, composed along H consecutive or independent increments, with exponents in brackets and bands over three seeds. Dotted lines mark the 30 mm tolerance and the released chunk length. 3D FLOWMATCH is omitted as non-quadratic.

From a deviation to a chunk. A policy emits a chunk of increments rather than one, and the per-increment deviation composes along the chunk through the adjoint (Appendix C.14). We compare the released path with the geodesic of the frozen metric at $\tau = \frac{1}{4}$ over fifty consecutive recorded increments, with 120 chunks under each of three noise realisations. Their separation composes to 4.5 ± 0.7 mm of position and a fifth of a degree of rotation, against a 30 mm placement tolerance (Appendix D.7). The exponent in the chunk length is 1.01 for consecutive increments and 0.72 for independent ones (Figure 4c).

metric	effect	95% CI	MDE
per-step trans.	-0.010	$[-0.088, 0.066]$	.109
per-step rot.	-0.020	$[-0.093, 0.046]$	.101
composed trans.	-0.099	$[-0.257, 0.056]$	.223
composed rot.	-0.046	$[-0.234, 0.115]$	.256

Table 3: Released path against the exact geodesic of the frozen metric, in percent of the released-path policy's error, with paired 95% intervals over episodes. MDE is the smallest effect detectable at $\alpha = 0.05$ and 80% power.

From a chunk to a trained policy. The criterion of §3.2 is decided from the loss, before any training. What a deviation does to a trained model is a separate question, which we examine on the action head. At $H = 50$ the composed separation is 3.7 percent of the translational error and 5.9 percent of the rotational error of the policy trained on the released path, measured on 1 000 chunks from its training episodes. We trained four policies that are identical except for the conditional path: the released construction, one and two levels of correction, and the exact geodesic of the frozen metric. Table 3 compares the released path with that geodesic. Every interval covers zero, the design detects effects of 0.10 to 0.26 percent at $\alpha = 0.05$ and 80 percent power, and no layer of the five hundred moves further under the change of path than under a change of random seed. The trained models therefore differ far less than the paths they were trained on. Under the loss's own metric the released path is already the geodesic, so there is no gain to recover, and the geometric size of a departure overstates its effect on the model by more than an order of magnitude.

6 Conclusion

We give a criterion that decides, from the training loss alone, whether a conditional path on $\mathrm{SE}(3)$ is already the geodesic it is meant to approximate. The criterion is exact rather than asymptotic, and where it fails the departure and its correction are both closed form. For a given axis, what separates the two cases is alignment rather than anisotropy: at fixed eigenvalues, rotating the eigenbasis alone moves a construction from optimal to second order away. Applied to six released configurations, the criterion finds four already optimal, none of them designed that way. A conditional path's optimality has until now been judged by training two models and comparing them, and the criterion decides it from the loss, before any training. On one action head, training on a path that differs from the released one by several percent of the policy's error changed the policy by under 0.3 percent.

THE USE OF LARGE LANGUAGE MODELS (LLMS)

In this work, we used generative AI tools to edit the manuscript for readability. None of the tasks requiring disclosure are applicable to this work. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Riad Ahmed, Sujosh Nag, Moniruzzaman Akash, Mostafa Hussein, and Momotaz Begum. Trajectory-consistent flow matching for robust visuomotor policy learning. *arXiv preprint arXiv:2605.08511*, 2026.

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023. arXiv:2209.15571.

Vansh Bansal, Saptarshi Roy, Alessandro Rinaldo, and Purnamrita Sarkar. On the convergence and straightness of rectified flow. In *International Conference on Artificial Intelligence and Statistics*, volume 300, pp. 4636–4644, 2026. arXiv:2410.14949.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. π_0: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

S. Blanes, F. Casas, J. A. Oteo, and J. Ros. The Magnus expansion and some of its applications. *Physics Reports*, 470(5–6):151–238, 2009. arXiv:0810.5488.

Avishek Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian Fatras, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. SE(3)-stochastic flow matching for protein backbone generation. In *International Conference on Learning Representations*, 2024. arXiv:2310.02391.

Zhonglin Cao, Mario Geiger, Allan Dos Santos Costa, Danny Reidenbach, Karsten Kreis, Tomas Geffner, Franco Pellegrini, Guoqing Zhou, and Emine Kucukbenli. Efficient molecular conformer generation with SO(3)-averaged flow matching and reflow. In *International Conference on Machine Learning*, volume 267, pp. 6522–6542, 2025. arXiv:2507.09785.

John-Marc Chandonia, Lindsey Guan, Shiangyi Lin, Changhua Yu, Naomi K. Fox, and Steven E. Brenner. SCOPe: improvements to the structural classification of proteins – extended database to facilitate variant interpretation and machine learning. *Nucleic Acids Research*, 50(D1):D553–D559, 2022. doi: 10.1093/nar/gkab1054.

Junhua Chen, Simon V. Mathis, Charles Harris, Kieran Didi, and Pietro Lio. Flows, straight but not so fast: Exploring the design space of rectified flows in protein design. *arXiv preprint arXiv:2510.24732*, 2025.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, 2024. arXiv:2302.03660.

Austin H. Cheng, Alston Lo, Kin Long Kelvin Lee, Santiago Miret, and Alán Aspuru-Guzik. Stiefel flow matching for moment-constrained structure elucidation. In *International Conference on Learning Representations*, 2025. arXiv:2412.12540.

Antoine Collas, Ce Ju, Nicolas Salvy, and Bertrand Thirion. Riemannian flow matching for brain connectivity matrices via pullback geometry. In *Advances in Neural Information Processing Systems*, 2025. arXiv:2505.18193.

Gabriele Corso, Arthur Deng, Nicholas Polizzi, Regina Barzilay, and Tommi Jaakkola. Deep confident steps to new pockets: Strategies for docking generalization. In *International Conference on Learning Representations*, 2024. arXiv:2402.18396.

Clemens Eppner, Arsalan Mousavian, and Dieter Fox. ACRONYM: A large-scale grasp dataset based on simulation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6222–6227, 2021. doi: 10.1109/ICRA48506.2021.9560844.

Andreas René Geist, Jonas Frey, Mikel Zhobro, Anna Levina, and Georg Martius. Learning with 3D rotations, a hitchhiker's guide to SO(3). In *International Conference on Machine Learning*, volume 235, pp. 15331–15350, 2024. arXiv:2404.11735.

Nikolaos Gkanatsios, Jiahe Xu, Matthew Bronars, Arsalan Mousavian, Tsung-Wei Ke, and Katerina Fragkiadaki. 3D FlowMatch Actor: Unified 3D policy for single- and dual-arm manipulation. *arXiv preprint arXiv:2508.11002*, 2025.

Arieh Iserles, Hans Z. Munthe-Kaas, Syvert P. Nørsett, and Antonella Zanna. Lie-group methods. *Acta Numerica*, 9:215–365, 2000. doi: 10.1017/S0962492900002154.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24240–24253, 2022. arXiv:2206.01729.

Nayoung Kim, Seongsu Kim, Minsu Kim, Jinkyoo Park, and Sungsoo Ahn. MOFFlow: Flow matching for structure prediction of metal-organic frameworks. In *International Conference on Learning Representations*, 2025. arXiv:2410.17270.

Jin Sub Lee and Philip M Kim. FlowPacker: protein side-chain packing with torsional flow matching. *Bioinformatics*, 41(3):btaf010, 2025. doi: 10.1093/bioinformatics/btaf010.

Byeongdo Lim, Jongmin Kim, Jihwan Kim, Yonghyeon Lee, and Frank C. Park. EquiGraspFlow: SE(3)-equivariant 6-DoF grasp pose generative flows. In *Conference on Robot Learning*, volume 270, pp. 5067–5086, 2025.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. arXiv:2210.02747.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. LIBERO: Benchmarking knowledge transfer for lifelong robot learning. In *Advances in Neural Information Processing Systems 36, Datasets and Benchmarks Track*, 2023a. arXiv:2306.03310.

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*, 2023b. arXiv:2209.03003.

Yunqing Liu, Yi Zhou, and Wenqi Fan. Geometric flow matching for molecular conformation generation via manifold decomposition. *arXiv preprint arXiv:2605.25577*, 2026.

Aaron Lou, Minkai Xu, Adam Farris, and Stefano Ermon. Scaling Riemannian diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 80291–80305, 2023. arXiv:2310.20030.

Jerrold E. Marsden and Tudor S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, volume 17 of *Texts in Applied Mathematics*. Springer, 2 edition, 1999. doi: 10.1007/978-0-387-21792-5.

Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M. Wood. FlowMM: Generating materials with Riemannian flow matching. In *International Conference on Machine Learning*, volume 235, pp. 35664–35686, 2024. arXiv:2406.04713.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In *International Conference on Machine Learning*, volume 202, pp. 28100–28127, 2023. arXiv:2304.14772.

Congzhou M. Sha. On the continuity of flows. *arXiv preprint arXiv:2512.12821*, 2025.

Neta Shaul, Ricky T. Q. Chen, Maximilian Nickel, Matthew Le, and Yaron Lipman. On kinetic optimal probability paths for generative models. In *International Conference on Machine Learning*, volume 202, pp. 30883–30907, 2023. arXiv:2306.06626.

Finn M. Sherry and Bart M. N. Smets. Flow matching on Lie groups. In *International Conference on Geometric Science of Information*, volume 16033 of *Lecture Notes in Computer Science*, pp. 54–62. Springer, 2025. doi: 10.1007/978-3-032-03918-7_6.

Mustafa Shukor, Dana Aubakirova, Francesco Capuano, Pepijn Kooijmans, Steven Palma, Adil Zouitine, Michel Aractingi, Caroline Pascal, Martino Russi, Andres Marafioti, et al. SmolVLA: A vision-language-action model for affordable and efficient robotics. *arXiv preprint arXiv:2506.01844*, 2025.

Hannes Stark, Bowen Jing, Chenyu Wang, Gabriele Corso, Bonnie Berger, Regina Barzilay, and Tommi Jaakkola. Dirichlet flow matching with applications to DNA sequence design. In *International Conference on Machine Learning*, volume 235, pp. 46495–46513, 2024. arXiv:2402.05841.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. arXiv:2302.00482.

Zhitao Wang, Yanke Wang, Jiangtao Wen, Roberto Horowitz, and Yuxing Han. ReSeFlow: Rectifying SE(3)-equivariant policy learning flows. *arXiv preprint arXiv:2509.22695*, 2025.

Jason Yim, Andrew Campbell, Andrew Y. K. Foong, Michael Gastegger, José Jiménez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S. Veeling, Regina Barzilay, Tommi Jaakkola, and Frank Noé. Fast protein backbone generation with SE(3) flow matching. *arXiv preprint arXiv:2310.05297*, 2023a.

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. In *International Conference on Machine Learning*, volume 202, pp. 40001–40039, 2023b. arXiv:2302.02277.

Angxiao Yue, Zichong Wang, and Hongteng Xu. ReQFlow: Rectified quaternion flow for efficient and high-quality protein backbone generation. In *International Conference on Machine Learning*, volume 267, pp. 73828–73850, 2025. arXiv:2502.14637.

Olga Zaghen, Floor Eijkelboom, Alison Pouplin, Cong Liu, Max Welling, Jan-Willem van de Meent, and Erik J. Bekkers. Riemannian variational flow matching for material and protein design. In *International Conference on Learning Representations*, 2026. arXiv:2502.12981.

Miloš Žefran, Vijay Kumar, and Christopher B. Croke. On the generation of smooth three-dimensional rigid body motions. *IEEE Transactions on Robotics and Automation*, 14(4):576–589, 1998.

Miloš Žefran, Vijay Kumar, and Christopher Croke. Metrics and connections for rigid-body kinematics. *The International Journal of Robotics Research*, 18(2):243–258, 1999.

Pengze Zhang, Hubery Yin, Chen Li, and Xiaohua Xie. Tackling the singularities at the endpoints of time intervals in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6945–6954, 2024. arXiv:2403.08381.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5745–5753, 2019. arXiv:1812.07035.

Yuchen Zhu, Tianrong Chen, Lingkai Kong, Evangelos A. Theodorou, and Molei Tao. Trivialized momentum facilitates diffusion generative modeling on Lie groups. In *International Conference on Learning Representations*, 2025. arXiv:2405.16381.

Appendix

Table of Contents

A Extended related work **14**
- A.1 Conditional path construction on Lie groups . . . 14
- A.2 Equivariant architectures and the metric behind them . . . 15
- A.3 Straightening the marginal field . . . 15
- A.4 Optimal rigid-body motion in kinematics . . . 15
- A.5 The cut locus and reported failures on compact manifolds . . . 16
- A.6 Rotation representations and error accumulation . . . 16

B Notations, datasets, theorems and implementation details **17**

C Derivations and proofs **21**
- C.1 Standing assumptions and scope . . . 21
- C.2 Theorem 1, in full . . . 22
- C.3 Theorem 2, in full . . . 24
- C.4 The two degenerate cases . . . 27
- C.5 Factorized paths on a product . . . 28
- C.6 Chart-induced metrics are flat . . . 29
- C.7 Scaling of the algebra-versus-group deviation . . . 30
- C.8 Which losses give a left-invariant metric . . . 33
- C.9 Proposition 7, in full . . . 34
- C.10 The corrections as a coupling homotopy . . . 36
- C.11 The general base, and the obstruction at level two . . . 38
- C.12 Operation counts on $\mathrm{SE}(3)$ . . . 39
- C.13 Where the statements hold . . . 40
- C.14 Composition of a sequence of increments . . . 42

D Experimental detail **43**
- D.1 Scope and inclusion . . . 43
- D.2 The choice of path parameter . . . 44
- D.3 Block weights do not change the optimal path . . . 44
- D.4 Alignment and anisotropy varied independently . . . 44
- D.5 The ladders, and what the apparatus can resolve . . . 45
- D.6 Wall-clock cost . . . 46
- D.7 Composition, and why the protocol matters . . . 46
- D.8 Displacement, not the mixed norm . . . 47
- D.9 The four policies, and what the null rests on . . . 47
- D.10 Two things the harness had to get right . . . 48
- D.11 The correction is a training-time object . . . 48

A EXTENDED RELATED WORK

This section expands the comparison in Section 4. Prior work relevant to this paper spans two broad objectives: constructing a conditional path that a generative model can train against, and characterising the geometry of the group on which that path is defined. Within these objectives we distinguish methods for conditional path construction on Lie groups (§A.1), equivariant architectures and the metric they induce (§A.2), straightening of the marginal field (§A.3), optimal rigid-body motion in kinematics (§A.4), the cut locus and reported failures on compact manifolds (§A.5), and rotation representations together with error accumulation over sequences (§A.6). Table 4 summarises their relationship to the components of this paper.

Table 4: **Relationship between this work and the principal lines of prior work.**

Line of work	Primary objective	Connection to this work	Key distinction
Conditional path construction	Supply a closed-form curve joining source to data	Provides the three constructions measured in §3.1	The construction is fixed by convention and its distance from the action-minimising curve is not reported, whereas we measure that distance in closed form
Equivariant architectures	Impose a symmetry on the learned field	Is sometimes taken to fix the metric a loss induces, which is the input to the criterion of §3.2	World-frame equivariance leaves body velocities unchanged and fixes no metric. Only invariance of the loss under a change of body frame decides the criterion
Straightening the marginal field	Reduce the integration cost of the learned flow	Addresses the same practical bottleneck as §3.3	Acts on the marginal field after conditional averaging and changes the coupling, whereas the deviation here is a property of a single conditional path before averaging
Optimal motion in kinematics	Characterise optimal curves between rigid-body poses	Establishes the qualitative facts behind Theorem 1 and the block-isotropic family	Determines which metrics admit which geodesics without quantifying the gap or connecting it to generative models
Cut locus and compact-manifold failure	Explain degraded performance on compact manifolds	Shares the branch locus with the degeneracy of §2	Replaces the probability path in response, and does not report what fraction of endpoint pairs is affected

A.1 CONDITIONAL PATH CONSTRUCTION ON LIE GROUPS

Flow matching requires a curve joining a source sample to a data sample along which the velocity field is regressed. In Euclidean space the straight line is the standard choice and a substantial literature concerns modifying it, through straightening, optimal-transport coupling, or schedule reweighting (Liu et al., 2023b; Pooladian et al., 2023; Tong et al., 2024; Bansal et al., 2026). On general manifolds the curve is derived from a premetric, and when that premetric is the geodesic distance the resulting path is the minimising geodesic (Chen & Lipman, 2024). On Lie groups three closed-form constructions are in use. Screw-motion flows take the one-parameter subgroup path equation 2 (Sherry & Smets, 2025). Pose-vector action heads interpolate in the algebra and exponentiate once (Black et al., 2024; Shukor et al., 2025). Losses that treat rotation and translation separately induce a factorized path (Lee & Kim, 2025; Cao et al., 2025; Miller et al., 2024). A related line works on a single factor: torsional diffusion operates on the hypertorus of torsion angles alone (Jing et al., 2022), where Proposition 14 makes the deviation structurally zero for pairs within half a period on every circle. Further deployments cover crystals, matrix manifolds, and molecular docking (Kim et al., 2025; Cheng et al., 2025; Collas et al., 2025; Liu et al., 2026).

The group interpolation, the algebra interpolation, and the factorized construction are not interchangeable. They agree at the endpoints and differ in between, and which one an implementation uses is determined by how its loss is written rather than by a comparison among them. A loss on the relative logarithm produces the group interpolation, a loss on the endpoint logarithms produces the algebra interpolation, and separate rotational and translational terms produce the factorized construction. **No published comparison states how far apart they are, or how far any of them lies from the curve that minimises the conditional action, and on $\mathrm{SE}(3)$ none of the three is that curve in general.** Section 3.1 gives both distances in closed form, Section 3.2 states the condition under which the factorized construction is the minimiser exactly, and Section 5 evaluates both at the amplitudes these implementations operate at.

A.2 Equivariant architectures and the metric behind them

Equivariance under the action of $\mathrm{SE}(3)$ is a design principle in most generative models on the group, and is generally treated as a constraint on the architecture and its representations (Bose et al., 2024; Miller et al., 2024; Cheng et al., 2025). Whether it also determines the metric the training loss induces, the quantity the criterion of §3.2 takes as input, depends on which symmetry is meant.

A loss invariant under a rotation Q of the body frame is invariant under $\xi \mapsto \mathrm{Ad}_Q \xi$. Invariance of the quadratic form $\xi^\top M \xi$ then requires $\mathrm{Ad}_Q^\top M \,\mathrm{Ad}_Q = M$ for every $Q \in \mathrm{SO}(3)$, whose rotational block is $Q^\top J Q = J$, and a symmetric matrix commuting with every rotation is a multiple of the identity. **Body-frame invariance therefore implies an isotropic rotational block. Schur's lemma leaves the cross block a multiple of the identity rather than forcing it to vanish, so the factorized construction is already a geodesic by Proposition 4 whenever that cross block is zero.** Equivariance as these architectures build it is a different symmetry. A rotation of the world frame, $g \mapsto Qg$, leaves body velocities unchanged, so it constrains the network without constraining the metric its loss induces. Of the released configurations in §5, the two with isotropic blocks satisfy the criterion through their loss weights, not through the equivariance of their architectures.

A.3 Straightening the marginal field

A substantial line of work reduces the curvature of the learned marginal velocity field, which arises from averaging over conditional paths that intersect in state space and which inflates the truncation error of coarse solvers. In Euclidean space, rectified flow retrains a model on its own samples so that the learned trajectories become straighter (Liu et al., 2023b), minibatch optimal transport couples endpoints so that conditional paths intersect less often (Pooladian et al., 2023; Tong et al., 2024), and convergence analyses bound the discretisation error of a coarse solver by a curvature functional (Bansal et al., 2026). The corresponding constructions on $\mathrm{SE}(3)$ follow the same objective. FoldFlow builds its coupling from Riemannian optimal transport (Bose et al., 2024), ReQFlow rectifies in a quaternion parameterisation and reports substantial reductions in the number of sampling steps (Yue et al., 2025), FrameFlow selects a schedule under which trajectories can be simulated in fewer integration steps (Yim et al., 2023a), and Chen et al. (2025) examine the design space of rectification for protein backbones and record that current models require hundreds to thousands of network evaluations, which constrains design campaigns generating 10^4 to 10^6 candidates.

Each of these operates on the marginal field, after the conditional average has been taken, and none modifies the shape of the curve joining a fixed pair. Rectification modifies the coupling and retrains, and optimal transport modifies which endpoints are paired. The deviation analysed here is that shape. It is defined for a single endpoint pair, is present before any averaging, and is independent of the coupling, so the two quantities are complementary and their contributions have not previously been separated. A related analysis concerns curvature in the objective rather than in the path. Zaghen et al. (2026) show that the flow matching loss and an endpoint-prediction loss coincide in Euclidean space but differ on curved manifolds, and identify the difference as a Taylor truncation of a Jacobi field along the geodesic, whose linear term retains no curvature information while the full field does. Their comparison is between two parameterisations of the loss evaluated along a single geodesic, and ours is between two paths.

A.4 Optimal rigid-body motion in kinematics

The question of which curve between two rigid-body poses is optimal predates generative modelling. Žefran et al. (1998) formulated smooth motion generation on $\mathrm{SE}(3)$ variationally. Žefran et al. (1999) proved that no Riemannian metric on $\mathrm{SE}(3)$ admits screw motions, the one-parameter subgroup paths equation 2, as its geodesics, the metrics with that property forming a two-parameter family of indefinite ones. The same work identified the products of the bi-invariant metric on $\mathrm{SO}(3)$ with a Euclidean metric on translations as the kinematically distinguished family, which is the block-isotropic family of §3.2.

Those results establish which metrics admit which optimal curves. They do not measure how far a curve that is not optimal lies from one that is. That distance, its leading term in closed form, and its connection to the conditional paths of generative models are what §3.1 and §3.3 supply. A construction of similar form appears in geometric integration. The constant-time bracket corrections of Lie-group integrators, the Magnus and Runge–Kutta–Munthe-Kaas methods surveyed by

Iserles et al. (2000), share the algebraic shape of equation 11 while correcting the flow of a known differential equation, whereas equation 11 corrects a boundary-value object without solving it.

A.5 The cut locus and reported failures on compact manifolds

The cut locus of a point p is the closure of the set of points joined to p by two or more minimising geodesics, and the injectivity radius is the distance from p to it, within which the exponential map is a diffeomorphism. That radius is the minimum of two quantities, the distance to the first conjugate point, which is determined by curvature, and half the length of the shortest non-trivial closed geodesic, which is not. A flat manifold can therefore possess a cut locus, and a flat torus does, at half-period on each circle. Flow matching is reported to underperform Riemannian diffusion on such manifolds. Performance degrades on complex torus densities and the geodesic path is discontinuous at the antipode (Lou et al., 2023), the failure reproduces under different parameterisations (Zhu et al., 2025), and underfitting is attributed jointly to cut-locus discontinuity and to shrinking support of the probability path (Cheng et al., 2025; Stark et al., 2024). The construction records the degeneracy: for a spectral premetric on the sphere, non-degeneracy fails at the antipode and the conditional flow is defined only almost everywhere (Chen & Lipman, 2024, Fig. 11).

Lou et al. (2023) and Cheng et al. (2025) address the degradation by replacing the probability path rather than by measuring what fraction of endpoint pairs is affected. A separate mechanism is sometimes confused with it. Instability near the endpoints of the time interval is generally attributed to the score or to the loss weighting (Zhang et al., 2024), while Sha (2025) shows that when source and target have mismatched topology the optimal marginal field is spatially discontinuous across the decision boundary between modes, with a jump growing without bound as $t \to 1$. The exposure described here has the opposite time dependence, concentrating as $t \to 0$, where the relative amplitude $(1-t)\theta_{01}$ is largest, and that sign separates the two. Both are supported on an equidistance set, theirs between two data modes and a cut locus between two minimising geodesics, which is a structural parallel rather than a shared cause.

A.6 Rotation representations and error accumulation

That no representation of $\mathrm{SO}(3)$ in four or fewer Euclidean dimensions is continuous is a standard result (Zhou et al., 2019), and the resulting difficulties for regression, together with the quaternion double cover, are surveyed by Geist et al. (2024). In visuomotor policies, the mismatch between a pointwise velocity objective at training time and numerical integration at inference is reported to produce compounding trajectory error, addressed with trajectory-level supervision and higher-order solvers (Ahmed et al., 2026), and long-horizon action sequences are noted to be difficult on the non-Euclidean geometry of $\mathrm{SE}(3)$ (Wang et al., 2025).

Both lines concern a target that is hard to represent or to integrate. The target of a conditional path is undefined where the minimising geodesic is not unique. The root cause, the topology of the group, is shared, and the surface on which it appears is not. Appendix C.14 derives the composition of per-increment errors through the adjoint, which §5 uses to bound the geometric contribution to a sequence rather than to propose a competing remedy.

B NOTATIONS, DATASETS, THEOREMS AND IMPLEMENTATION DETAILS

Table 5: **Mathematical notation used throughout.**

Symbol	Definition
Group, algebra, and metric	
G, $\mathfrak{g}$	Lie group and its Lie algebra
$\mathrm{SO}(3), \mathrm{SE}(3)$	Rotation group and rigid-motion group
$g_0, g_1 \in G$	Endpoints of a conditional path
$\xi_0, \xi_1 \in \mathfrak{g}$	Endpoint logarithms, $g_i = \exp(\xi_i)$
ξ	Body velocity, $\dot{g} = g\hat{\xi}$
$u = \log(g_0^{-1} g_1)$	Relative increment
p, q	Translational components of ξ_0, ξ_1
$\Delta = \xi_1 - \xi_0$	Endpoint difference in the algebra
$\epsilon = \max(\|\xi_0\|, \|\xi_1\|)$	Endpoint amplitude
$\|u\|$	Relative amplitude
$\tau \in [0, 1]$	Path parameter, $\tau = 0$ at the source
M	Gram operator of a left-invariant metric
$J, \beta I_3$	Rotational and translational blocks of $M = \mathrm{diag}(J, \beta I_3)$
$\mathrm{ad}_x y = [x, y]$	Adjoint action
ad_x^*	Coadjoint, $\langle \mathrm{ad}_x^* \mu, y \rangle = \langle \mu, \mathrm{ad}_x y \rangle$
$\mathrm{Ad}_g, \mathrm{dexp}_{-F}$	Adjoint of a group element, and the differential of exp
$\mathcal{J}_r(\phi)$	Right Jacobian of $\mathrm{SO}(3)$, $\xi = \mathcal{J}_r(\phi)\dot{\phi}$
$[\cdot]_k$	Component of total degree k in (ξ_0, ξ_1)
$A = [\xi_0, [\xi_0, \xi_1]]$, $B = [\xi_1, [\xi_1, \xi_0]]$	Cubic brackets
Paths and deviations	
g_{riem}	Geodesic of M
$g_{\mathrm{geo}} = g_0 \exp(\tau u)$	Group interpolation equation 2
$g_{\mathrm{lin}} = \exp((1-\tau)\xi_0 + \tau\xi_1)$	Algebra interpolation equation 3
$(R_0 e^{\tau\hat{\omega}}, (1-\tau)p_0 + \tau p_1)$	Factorized construction equation 4
g_{base}	Base path with $\dot{\xi} = \Lambda_0(\xi)$
$D_{\mathrm{geo}}(\tau) = \log(g_{\mathrm{riem}}^{-1} g_{\mathrm{geo}})$	Deviation from the geodesic
$D(\tau) = \log g_{\mathrm{geo}} - \xi_{\mathrm{lin}}$	Deviation between the two logarithm-based constructions
d_R, d_T	Rotational and translational components of a deviation
$\theta_{\mathrm{rel}}, \theta_{\mathrm{end}}$	Relative and endpoint rotation angles
Correction and generators	
$\Lambda(x) = M^{-1}\mathrm{ad}_x^*(Mx)$	Generator of the Euler–Poincaré equation, equation 1
$\Lambda_M, \Lambda_{\mathrm{iso}}$	Λ under a stated metric, and under a block-isotropic one
Λ_0	Generator of the base path
$D\Lambda(u)[h]$	Polar of Λ, see equation 10
$N_2(u) = D\Lambda(u)[\Lambda(u)] - [u, \Lambda(u)]$	Second-level generator
$F_\ell(\tau)$	Correction at level ℓ, see equation 11
$c(\tau) = -\frac{\tau(1-\tau)}{2}$	First-level profile
$d(\tau) = \frac{\tau(1-\tau)(1-2\tau)}{12}$	Second-level profile
u_{corr}	Conditional velocity field of the corrected path, see equation 12
Criterion, charts, and constants	
$\omega = \log(R_0^\top R_1)$	Endpoint rotation vector
$w = R_0^\top (p_1 - p_0)$	Body-frame endpoint translation
$\nu(\tau) = R(\tau)^\top w$	Body-frame translation velocity
Φ	Chart, $\Phi(R) = \log(R_0^\top R)$
$J(\tau) = \mathcal{J}_r(\phi)^{-\top} \mathcal{J}_r(\phi)^{-1}$	Chart-induced rotational metric
Q	Fixed rotation of an eigenframe
$\kappa_R^{(1)}, \kappa_R^{(2)}, \kappa_T^{(2)}$	Criterion constants, $16/15$, $2/\sqrt{3}$, 2
$s(\theta) = d\log d / d\log\theta$	Deviation exponent
θ_{01}	Endpoint angle, with $\theta(\tau) = (1-\tau)\theta_{01}$
$d_\pm, \theta_M$	Distances to the two lifts under the antipodal double cover, and the quotient distance $2\min(d_+, d_-)$
Experimental quantities	
gap	Deviation of a construction from the shot geodesic
floor	Apparatus reproducibility on a case whose answer is known
misalignment	$\|\omega \times J\omega\| / (\|J\|_2 \|\omega\|^2)$
σ	Regression residual of a trained network
ρ	Ratio of the per-increment gap to that residual, §5.4

Table 6: **Source of datasets used in this research.**

Dataset	Details
Public datasets	
Protein Data Bank chains	Protein backbones, 115 X-ray chains at resolution below 2.5 Å, clustered at 30% sequence identity. Source IGSO3(σ=1.5), anchored to a data frame, as FrameFlow (Yim et al., 2023a) draws it. Factorized construction, chart-induced rotational metric. Reported in Ångström. Used for the audit (§5.1), the criterion (§5.2) and the released-metric comparisons of §5.3. The FrameFlow weights read throughout are those trained on SCOPe (Chandonia et al., 2022).
LIBERO (Liu et al., 2023a)	Action chunks, 1693 episodes and 273 465 frames, as used by SmolVLA (Shukor et al., 2025). Each action is an increment between consecutive recorded poses, and the conditional path joins a unit Gaussian in normalised increment coordinates to that action. Algebra construction in one fixed chart, flat pullback metric. Reported in metres. Its normalisation statistics set the frozen metric of the bottom ladder in §5.3, and its trajectories are used for composition and transfer (§5.4).
DockGen (Corso et al., 2024)	903 ligand poses from 330 protein-ligand complexes. Source IGSO3(σ=1.55) in rotation with a Gaussian of 19 Å in translation, from the released checkpoint's configuration. Factorized construction, rotational block $\sigma_{\mathrm{rot}}(t)^{-2} I_3$. Reported in Ångström. Used for the audit and the criterion.
ACRONYM (Eppner et al., 2021)	Simulated grasp poses, as used by EquiGraspFlow (Lim et al., 2025). The subset in hand covers 93 object–scale files over 52 meshes in two categories, Bowl and Laptop. Factorized construction, $M = I_6$. Used for the audit and the source-law comparison.
Synthetic datasets	
Amplitude grid	SE(3) endpoint pairs with $\|\omega\| = \|v\| = \epsilon$ at both endpoints and $M = I$, the amplitude set rather than observed. Five seeds up to $\epsilon = 1.2$, and 500 pairs per seed at two seeds from $\epsilon = 1.4$. Used for the relative size of the two deviations (Table 10) and the large-amplitude check.
Fixed-direction scaling	Two endpoint directions held fixed while the amplitude is halved. Two metrics: $M = I$ and $\mathrm{diag}(1, 2, 4; 1, 1, 1)$. Used for the residual slopes of Theorem 1 and Proposition 7.
Eigenframe-rotation control	A metric with its eigenframe rotated through a fixed Q, $\tau \mapsto QJ(\tau)Q^{\top}$, every eigenvalue held fixed. Used to separate anisotropy from alignment in §5.2.
Free algebra	$\mathbb{Q}\langle X, Y\rangle$ on two non-commuting generators, truncated at total degree five, exact rational coefficients. Used for the degree-two and degree-four vanishing of Theorem 2.
Floor calibration	SE(3) with $M = cI$, where the factorized construction is the geodesic exactly. Used to measure the apparatus against a known answer before any signal is read.

Table 7: **Complete list of the formal statements, grouped by topic.**

No.	Type	Statement
The two deviations		
1	Theorem 1	Second-order optimality gap: $D_{\mathrm{geo}}(\tau) = \frac{\tau(1-\tau)}{2}\Lambda(u) + O(\|u\|^3)$, uniformly in τ, for any fixed left-invariant metric
2	Theorem 2	Third-order deviation: $D(\tau) = -\frac{\tau(1-\tau)}{12}(A+B) + O(\epsilon^5)$, the degree-two and degree-four terms vanishing on every Lie algebra
3	Cor. 12	The relative error of the cubic form is $O(\epsilon^2)$ on directions where $\|A+B\| \geq c\,\epsilon^3$
4	Prop. 3	The factorized construction differs from group interpolation at second order, with the rotational component vanishing identically
When the gap vanishes		
5	Prop. 15	A factorized path on a Riemannian product is the minimising geodesic exactly when every schedule agrees and equals the identity
6	Prop. 4	Alignment criterion: under a constant $M = \mathrm{diag}(J, \beta I_3)$ the factorized path satisfies the geodesic equation iff $\omega = 0$ or ω is an eigenvector of J, and for every endpoint pair iff J is isotropic
7	Appendix C.6, after Collas et al. (2025)	A metric pulled back from Euclidean space by a chart onto a convex set is flat, and the minimising geodesic is the preimage of the segment, exactly and at every amplitude
8	Cor. 5	The metric a rotational loss written in a fixed logarithm chart induces is of that kind, so the factorized path is its unique minimising geodesic whenever $\|\omega\| < \pi$
9	Prop. 6	Body-frame invariance forces isotropy: a loss invariant under a rotation of the body frame has $Q^\top J Q = J$ for every Q, hence $J = \lambda I$, and satisfies the criterion when its Gram operator is block-diagonal
Closing the gap		
10	Prop. 7(i)	About any base with $\dot\xi = \Lambda_0(\xi)$ for a fixed generator, the correction $c(\tau)\big(\Lambda(u) - \Lambda_0(u)\big)$ leaves a residual $O(\|u\|^3)$. The hypothesis admits group interpolation ($\Lambda_0 = 0$) and the factorized construction ($\Lambda_0 = \Lambda_{\mathrm{iso}}$), and excludes algebra interpolation, which solves no such equation
11	Prop. 7(ii)	About group interpolation the third-order residual is $\frac{\tau(1-\tau)(1-2\tau)}{12}N_2(u)$, and the second level leaves $O(\|u\|^4)$
12	Prop. 7(iii)	equation 12 is the exact body velocity of equation 11, using $\mathrm{dexp}_{-F_\ell}(\dot F_\ell)$ rather than $\dot F_\ell$
13	Prop. 18	A homotopy in the metric reproduces the first two levels by a route independent of the Magnus computation
Boundary		
14	Prop. 19	Exposure to the branch locus is confined to the first integration step, is not removed by the double cover, and is largely removed by independent pairing
15	Prop. 14	On an abelian algebra every bracket vanishes, so $D(\tau) \equiv 0$ exactly at every order within the injectivity domain, separating curvature from topology
16	Prop. 20	The operating amplitude is set by the source law, not by the data. A rotationally invariant source makes $R_0^\top R_1$ Haar whatever the data is, giving the angle density $(1-\cos\theta)/\pi$ whose median solves $m - \sin m = \pi/2$ and equals 132.35°. Under IGSO3(σ) the median is displaced below that by an amount fixed by σ

Table 8: **Implementation details.**

Setting	Value	Scope
Geodesic solve, shared by every measurement		
Integrator	DOP853, tolerance 10^{-13}	All shot geodesics
Boundary solve	Shooting with Newton polishing, residual below 10^{-12} in $\log(g(1)^{-1}g_1)$	All shot geodesics
Apparatus floor	5.7×10^{-12} on $\mathrm{SE}(3)$ with $M = cI$, where the answer is algebraically zero. The $\mathrm{SO}(3)$ figure of 5.2×10^{-14} calibrates the integrator on a generator that vanishes identically and understates $\mathrm{SE}(3)$	All reported gaps
Seeds	$\{42, 43, 44\}$ unless a row states otherwise	All measurements
The order ladders (§5.3)		
Pairs, parameter	300 per rung at $\tau = \frac{1}{4}$, thirty-five cells, no non-converged solves	Table 2
Endpoint draw	g_0 from the implementation's own endpoint draw (top panel) or a random pose (bottom panel), $u = \epsilon\hat{u}$ with $\hat{u}$ uniform on the unit sphere of $\mathfrak{se}(3)$	Per-domain rows
Synthetic grid	$\|\omega\| = \|v\| = \epsilon$, $M = I$. Five seeds to $\epsilon = 1.2$; 500 pairs per seed at two seeds from $\epsilon = 1.4$	Table 10, large amplitude
Profile check	Five values of τ at $\epsilon \in \{0.2, 0.1, 0.05\}$, against $\tau(1-\tau)/2$	Table 9
Amplitude ladders	Five amplitudes from 0.4 to 0.025 over six generic pairs (Table 11); four from 0.2 to 0.025 over five pairs (Table 12)	Correction orders
The criterion (§5.2)		
Pairs, metric	30 000 per metric and seed, under the metric each loss induces, not $M = I$	Criterion rows
Eigenframe control	The same metric rotated through a fixed Q, every eigenvalue held	Anisotropy against alignment
Metric grid	Seven eigenvalue ratios by seven eigenbasis rotations, three seeds per cell	Appendix D.4
τ sweep	Thirteen values from 0.05 to 0.90, clustered at $\{0.225, 0.25, 0.275\}$, 5 000 pairs per cell, both metrics at every τ	Gated metric
The correction (§5.3)		
Pairs per cell	1 200 per arm and seed	Arm comparisons
Fixed-direction scaling	Two directions held while the amplitude is halved. $M = I$ and $\mathrm{diag}(1, 2, 4; 1, 1, 1)$, three seeds	Residual slopes
Block scaling	Four seeds at three scalings $M = \mathrm{diag}(\alpha I, \beta I)$	Λ invariant to the block weight
Within-block anisotropy	$M = \mathrm{diag}(J, I)$, $J = \mathrm{diag}(1, 2, 4)$, two seeds	First level modular in the metric
Large amplitude	$\epsilon \in [1.4, 2.4]$, residual as a ratio to the uncorrected deviation rather than by order	Beyond the expansion
Composition and transfer (§5.4)		
Chunk composition	$H = 50$ consecutive increments from single recorded trajectories, thirty chunks, seed 42	Composition
Exponent sweep	$H \in \{5, 10, 20, 30, 50, 75, 100\}$, 120 chunks per seed, both sampling schemes	Exponent in H
Four policies	25 000 steps each, identical but for the conditional path, evaluated on 1 000 paired chunks from 166 training episodes with the flow-matching noise shared across arms	Table 3
Weight comparison	One additional policy at a different seed, per-layer and per-output-row contrasts over five hundred tensors	Table 15
Symbolic and Monte Carlo checks (Appendix C)		
Bracket identity	The semidirect bracket against a 4×4 matrix commutator, 2 000 random draws	The $\mathrm{SE}(3)$ instance of the split
Scaling exponents	Log-log slopes over nine decades in each of λ and μ	The exponents of the split
Closed-form identities	$\|\tilde{R}\|$ at 3 000 draws; M, N and $\|M\|_F^2 + \|N\|_F^2$ at 2 000	The normalisation constants
Constants	2×10^7 uniform draws for $\kappa_R^{(1)}$, $\kappa_R^{(2)}$, $\kappa_T^{(1)}$ and $\kappa_T^{(2)}$	The normalisation constants
Bi-invariance	$\|\mathrm{ad}_u^*(Mu)\|$ on 4 000 uniform unit u, at four metrics on $\mathrm{SE}(3)$ and one on $\mathfrak{so}(3)$	The two degenerate cases
Velocity field	Central differences against $\mathrm{Ad}_{\exp(-F)}(u) + \mathrm{dexp}_{-F}(\dot{F})$ at three amplitudes	Prop. 7(iii)
Homotopy check	Five generic endpoint pairs at $M = I$, comparing the two derivations of the third-order term	Prop. 18
Sequence composition	40 draws at each of $H \in \{3, 6, 10\}$	Composition of increments
Exact computation (Theorem 2)		
Free algebra	$\mathbb{Q}\langle X, Y\rangle$, truncated at total degree five, exact rational arithmetic	Degree-two and degree-four vanishing
Interpolation	Degree-four word coefficients at seven rational τ	Degree-four vanishing
Magnus cross-check	Coefficients through degree four against the recursion of Blanes et al. (2009)	The BCH input

C Derivations and proofs

These sections follow the order of §3: the assumptions carried throughout, the two deviations of §3.1, the criterion of §3.2, the corrections of §3.3, the conditions those results share, and what the implementation requires in practice. Every symbolic and numerical check reported below is listed with its sample size and tolerance in Table 8.

C.1 Standing assumptions and scope

Everything in §3 concerns the pairwise curve joining a fixed endpoint pair. We do not optimise the coupling between source and data, and we make no claim about the kinetic energy of the induced marginal path, which is a different functional obtained after conditional averaging (Shaul et al., 2023).

G is a finite-dimensional Lie group and M a fixed symmetric positive-definite Gram operator. The endpoints lie in a common normal neighbourhood on which every logarithm is taken on one smooth branch and on which equation 1 has a unique solution. Every $O(\cdot)$ estimate is uniform in $\tau \in [0, 1]$, with constants depending only on G, M, the norm and the neighbourhood.

Two radii appear and they are not interchangeable. Theorem 1 and the corrections of §3.3 are left-translation invariant and expand in the relative increment $\|u\|$, where $u = \log(g_0^{-1} g_1)$, so they do not require g_0 and g_1 to lie individually near the identity. Theorem 2 compares the logarithms of the endpoints themselves and expands in $\epsilon = \max(\|\xi_0\|, \|\xi_1\|)$. Inside a fixed normal neighbourhood the two are equivalent up to a constant, $\|u\| \leq C\epsilon$, but that constant is not tracked and the two symbols are kept distinct.

C.2 THEOREM 1, IN FULL

For a left-invariant metric with Gram matrix M on $\mathfrak{g}$, the energy of a curve $g(\tau)$ with body velocity $\xi = (g^{-1}\dot{g})^{\vee}$ is $E = \frac{1}{2}\int_0^1 \xi^{\top} M\xi \, d\tau$. Taking variations $\delta\xi = \dot{\eta} + \mathrm{ad}_{\xi}\eta$ with η vanishing at the endpoints gives

$$0 = \delta E = \int_0^1 \big\langle M\xi,\ \dot{\eta} + \mathrm{ad}_{\xi}\eta \big\rangle d\tau = \int_0^1 \big\langle -\tfrac{d}{d\tau}(M\xi) + \mathrm{ad}^{*}_{\xi}(M\xi),\ \eta \big\rangle d\tau, \tag{14}$$

so a curve is a geodesic if and only if $M\dot{\xi} = \mathrm{ad}^{*}_{\xi}(M\xi)$, which is equation 1. Writing $\Lambda(x) = M^{-1}\mathrm{ad}^{*}_{x}(Mx)$, note that Λ is quadratic and homogeneous of degree two, so $\Lambda(\lambda x) = \lambda^2\Lambda(x)$. This is what makes the gap second order.

The shooting map and the reduction to the identity The curve g_{riem} begins at g_0, not at the identity, so the Magnus expansion is applied not to it but to its left translate

$$h(\tau) = g_0^{-1} g_{\mathrm{riem}}(\tau), \qquad h(0) = e, \qquad h(1) = g_0^{-1} g_1 = e^{u}. \tag{15}$$

Left translation preserves the body velocity, $h^{-1}\dot{h} = \xi$, so the Euler–Poincaré equation is unchanged while the expansion now has the correct base point.

The boundary-value problem is solved by shooting. For $a \in \mathfrak{g}$ let h_a be the solution of equation 1 with $h_a(0) = e$ and $\xi_a(0) = a$, and define the local endpoint map

$$\mathcal{E}(a) = \log h_a(1). \tag{16}$$

At $a = 0$ the solution is constant and $D\mathcal{E}(0) = \mathrm{Id}$, so by the implicit function theorem there is a neighbourhood of the origin on which $\mathcal{E}$ is a diffeomorphism onto its image and a unique smooth $a = a(u)$ with $\mathcal{E}(a(u)) = u$. This is the step that licenses expanding the boundary-value solution in powers of u, and it is where the uniqueness hypothesis of §3 is used.

From the logarithm to the group-level deviation Expanding $a = u + a_2 + O(\|u\|^3)$ with $a_2 = O(\|u\|^2)$ and using $\dot{\xi} = \Lambda(\xi)$ with Λ quadratic homogeneous gives $\xi(\tau) = u + a_2 + \tau\Lambda(u) + O(\|u\|^3)$ uniformly in τ, whence, with $\Omega(\tau) = \log h(\tau)$ and the reconstruction estimate below,

$$\Omega(\tau) = \tau u + \tau a_2 + \tfrac{\tau^2}{2}\Lambda(u) + O(\|u\|^3). \tag{17}$$

The endpoint condition $\Omega(1) = u$ forces $a_2 = -\frac{1}{2}\Lambda(u)$ and therefore $\Omega(\tau) = \tau u - \frac{\tau(1-\tau)}{2}\Lambda(u) + O(\|u\|^3)$.

The quantity of interest is a relative group element, not a difference of logarithms, so one further step is needed. Since $g_{\mathrm{geo}}(\tau) = g_0 e^{\tau u}$,

$$D_{\mathrm{geo}}(\tau) = \log\big(g_{\mathrm{riem}}(\tau)^{-1} g_{\mathrm{geo}}(\tau)\big) = \log\big(e^{-\Omega(\tau)} e^{\tau u}\big). \tag{18}$$

The two exponents differ by $\tau u - \Omega(\tau) = O(\|u\|^2)$, so the first Baker–Campbell–Hausdorff commutator, being a bracket of that difference with $\tau u = O(\|u\|)$, is $O(\|u\|^3)$ and the logarithm of the product is the sum of the exponents to the order retained. This gives $D_{\mathrm{geo}}(\tau) = \frac{\tau(1-\tau)}{2}\Lambda(u) + O(\|u\|^3)$, uniformly in $\tau \in [0, 1]$, which is Theorem 1.

The reconstruction estimate The proof of Theorem 1 uses

$$\log g(\tau) = \int_0^{\tau} \xi(s) ds + \tfrac{1}{2}\int_0^{\tau}\int_0^{s} [\xi(s'), \xi(s)]\, ds' ds + O(\|\xi\|^3), \tag{19}$$

the first two terms of the Magnus expansion. With $\xi(s) = u + O(\|u\|^2)$ the bracket integrand is $[u + O(\|u\|^2),\ u + O(\|u\|^2)] = [u, u] + O(\|u\|^3) = O(\|u\|^3)$, so the double integral does not contribute at second order. This is the only place the estimate is used, and it is why the leading behaviour of D_{geo} is governed entirely by the boundary condition on $\xi(0)$.

Numerical verification of the profile and the correction The apparatus, its tolerances, and the floor against which every number below is reported are given in Table 8. Three checks follow. Table 9 divides $\|D_{\mathrm{geo}}(\tau)\|$ by $\|\Lambda(u)\|$ at three amplitudes, a ratio that should equal $\tau(1-\tau)/2$ independently of amplitude. Table 11 compares g_{geo} and g_{corr} against the numerically integrated geodesic at the midpoint. Its uncorrected column has local log-log slopes 1.90, 1.99, 1.99, 2.00, consistent with Theorem 1, while the corrected column has slopes 3.95, 3.98, 3.99, 3.99, one order

better than Proposition 7(i) gives, because the third-order term equation 13 carries the factor $(1-2\tau)$ and vanishes at $\tau = \frac{1}{2}$. The order is three at any other parameter, so measuring only at the midpoint would suggest a correction one order stronger than it is. Table 10 then compares the two deviations directly, with both endpoints generic rather than one at the identity.

Table 9: **The profile of the optimality gap.** Ratio $\|D_{\mathrm{geo}}(\tau)\|/\|\Lambda(u)\|$ at three amplitudes, against the analytic $\tau(1-\tau)/2$. The columns collapse onto the profile and approach it as the amplitude falls, which is the $O(\|u\|^3)$ remainder of Theorem 1.

τ	$\epsilon = 0.2$	$\epsilon = 0.1$	$\epsilon = 0.05$	$\tau(1-\tau)/2$
0.1	0.044869	0.044967	0.044992	0.045000
0.3	0.104735	0.104934	0.104983	0.105000
0.5	0.124700	0.124925	0.124981	**0.125000**
0.7	0.104735	0.104934	0.104983	0.105000
0.9	0.044869	0.044967	0.044992	0.045000

Table 10: **Relative size of the two deviations.** Both endpoints generic with $\|\omega_i\| = \|v_i\| = \epsilon$, $M = I$, $\tau = \frac{1}{4}$, medians over five seeds. The ratio grows as the amplitude falls, which is the difference between ϵ^2 and ϵ^3 made concrete: the smaller the amplitude, the more the departure from optimality dominates.

ϵ (rad)	degrees	$\|D_{\mathrm{geo}}\|$	$\|D\|$	ratio
0.02	1.15	7.52×10^{-5}	2.53×10^{-7}	297
0.05	2.86	4.74×10^{-4}	3.97×10^{-6}	119
0.10	5.73	1.92×10^{-3}	3.21×10^{-5}	60
0.20	11.46	7.91×10^{-3}	2.62×10^{-4}	30
0.40	22.92	3.08×10^{-2}	2.19×10^{-3}	14

Table 11: **The correction at the midpoint.** Deviation from the numerically integrated geodesic at $\tau = \frac{1}{2}$, medians over six generic endpoint pairs. The final column reports the corrected residual as a multiple of the apparatus floor.

ϵ	uncorrected	corrected	ratio	corrected/floor
0.4	3.125×10^{-2}	3.403×10^{-4}	1.1×10^{-2}	6.5×10^{9}
0.2	8.388×10^{-3}	2.202×10^{-5}	2.6×10^{-3}	4.2×10^{8}
0.1	2.117×10^{-3}	1.401×10^{-6}	6.6×10^{-4}	2.7×10^{7}
0.05	5.318×10^{-4}	8.832×10^{-8}	1.7×10^{-4}	1.7×10^{6}
0.025	1.333×10^{-4}	5.544×10^{-9}	4.2×10^{-5}	1.1×10^{5}

C.3 THEOREM 2, IN FULL

Theorem 2 requires ξ_0 and ξ_1 to lie in a neighbourhood of the origin on which two Baker–Campbell–Hausdorff maps are analytic on a common branch: $\mathrm{BCH}(-\xi_0, \xi_1)$, which produces u, and $\mathrm{BCH}(\xi_0, \tau u)$ for every $\tau \in [0, 1]$, which produces $\log g_{\mathrm{geo}}(\tau)$. On a matrix group, in the operator 2-norm, the first converges for $\|\xi_0\| + \|\xi_1\| < \pi$ (Blanes et al., 2009, Thm. 10). A bound for the second follows on a possibly smaller neighbourhood.

The expansion below is carried out in the free associative algebra $\mathbb{Q}\langle X, Y\rangle$ on two non-commuting generators, truncated at total degree five, with exact rational coefficients, rather than in a matrix representation. A three-dimensional algebra such as $\mathfrak{so}(3)$ satisfies identities a general Lie algebra does not, so a wrong coefficient can survive a check performed there. In the free algebra nothing cancels by accident, and an identity that holds there holds in every Lie algebra.

Proposition 8 (Baker–Campbell–Hausdorff series to fourth order)**.** *For P, Q in a Lie algebra, within the radius of convergence,*

$$\log\left(e^P e^Q\right) = P + Q + \tfrac{1}{2}[P,Q] + \tfrac{1}{12}[P,[P,Q]] - \tfrac{1}{12}[Q,[P,Q]] - \tfrac{1}{24}[Q,[P,[P,Q]]] + O(5), \quad (20)$$

graded by total degree in (P, Q) jointly. In the free Lie algebra on two generators the degree-four component is three-dimensional, and the coefficients on $[P,[P,[P,Q]]]$ and $[Q,[Q,[P,Q]]]$ are exactly zero.

Proof. The series is classical. Appendix C.3 verifies the coefficients through degree four against the Magnus recursion of Blanes et al. (2009), which yields the same series when the generator is taken piecewise constant, and independently by exact rational computation in a free associative algebra. Because every expansion below rests on equation 20, an error in it would be invisible to any downstream numerical check, which is why two independent paths are used. □

Lemma 9 (The geodesic tangent)**.** *With u as in §3,*

$$u \;=\; \underbrace{\Delta}_{u_1} \;\underbrace{-\tfrac{1}{2}[\xi_0,\xi_1]}_{u_2} \;+\; \underbrace{\tfrac{1}{12}\big(A - B\big)}_{u_3} \;+ O(\epsilon^4), \quad (21)$$

where u_k denotes the component of total degree k.

Proof. Apply Proposition 8 with $P = -\xi_0$ and $Q = \xi_1$, and use $[\xi_1, [\xi_0, \xi_1]] = -B$. Details in Appendix C.3. □

Note the *minus* sign between A and B in equation 21. The plus sign of equation 8 appears only after the next step.

Lemma 10 (Composition with g_0)**.**

$$\log g_{\mathrm{geo}}(\tau) \;=\; \xi_0 + \tau u + \tfrac{1}{2}[\xi_0, \tau u] + \tfrac{1}{12}[\xi_0,[\xi_0,\tau u]] - \tfrac{1}{12}[\tau u,[\xi_0,\tau u]] + O(\epsilon^4). \quad (22)$$

Proof. Proposition 8 with $P = \xi_0$ and $Q = \tau u$. □

Because τ multiplies u, and u has no component of degree zero, τ appears with the same power as the degree contributed by u. The term $[\tau u, [\xi_0, \tau u]]$ carries τ^2 while $[\xi_0, [\xi_0, \tau u]]$ carries τ^1. That asymmetry produces the $\tau(1 - \tau)$ profile.

Lemma 11 (Collection identity)**.**

$$[\Delta, [\xi_0, \xi_1]] \;=\; -\,(A + B). \quad (23)$$

Proof. $[\Delta, [\xi_0, \xi_1]] = [\xi_1, [\xi_0, \xi_1]] - [\xi_0, [\xi_0, \xi_1]] = -[\xi_1, [\xi_1, \xi_0]] - [\xi_0, [\xi_0, \xi_1]] = -(A + B)$, using only antisymmetry of the bracket. □

Proof of Theorem 2. Collect equation 22 by total degree, substituting equation 21.

Degree one. $\xi_0 + \tau u_1 = \xi_0 + \tau\Delta = \xi_{\mathrm{lin}}(\tau)$, so D has no first-order component by construction rather than by cancellation.

Degree two. Exactly two contributions appear,

$$\tau u_2 = -\frac{\tau}{2}[\xi_0, \xi_1] \qquad \text{and} \qquad \tfrac{1}{2}[\xi_0, \tau u_1] = +\frac{\tau}{2}[\xi_0, \xi_1], \quad (24)$$

using $[\xi_0, \Delta] = [\xi_0, \xi_1]$. They cancel.

Degree three. Four contributions survive. In the basis $\{A, B\}$:

	source	coeff. of A	coeff. of B
(i)	τu_3	$\tau/12$	$-\tau/12$
(ii)	$\frac{1}{2}[\xi_0, \tau u_2]$	$-\tau/4$	0
(iii)	$\frac{1}{12}[\xi_0, [\xi_0, \tau u_1]]$	$\tau/12$	0
(iv)	$-\frac{1}{12}[\tau u_1, [\xi_0, \tau u_1]]$	$\tau^2/12$	$\tau^2/12$
	total	$-\tau(1-\tau)/12$	$-\tau(1-\tau)/12$

Entry (iv) uses Lemma 11, which converts the antisymmetric combination $A - B$ of equation 21 into the symmetric $A + B$. Summing the first column gives $-\tau(1-\tau)/12$, and identically for the second.

Degree four. The degree-four component of $D(\tau)$ vanishes identically, and so does every even-degree component. Replacing each endpoint by its inverse, $(\xi_0, \xi_1) \mapsto (-\xi_0, -\xi_1)$, sends $g_{\mathrm{geo}}(\tau) = g_0 \exp(\tau \log(g_0^{-1} g_1))$ to its inverse, since $g_0^{-1} \exp(\tau \log(g_0 g_1^{-1})) = \big(\exp(-\tau \log(g_0^{-1} g_1))\, g_0^{-1}\big)$ on the common branch, and sends $\xi_{\mathrm{lin}}(\tau)$ to its negative, so $D(\tau; -\xi_0, -\xi_1) = -D(\tau; \xi_0, \xi_1)$. An odd function of (ξ_0, ξ_1) has no even-degree terms. Appendix C.3 verifies the degree-four vanishing independently by computing $D(\tau)$ in a free associative algebra with exact rational arithmetic, using that $D_4(\tau)$ is a polynomial in τ of degree at most four, so that vanishing at five or more distinct τ makes it identically zero. The term-by-term collection above is verified by the same computation. □

Corollary 12 (Order of the relative error)**.** *Fix a direction with $\|A+B\| \geq c\,\epsilon^3$ for a constant $c > 0$. The remainder in equation 8 is $O(\epsilon^5)$, so on that direction the relative error of the third-order form is $O(\epsilon^2)$. The normalisation is vacuous where the leading cubic term degenerates, and the same caveat applies to any quantity divided by $\|\Lambda(u)\|$.*

Two features of equation 8 are worth isolating. The degree-two cancellation equation 24 uses only antisymmetry and bilinearity, with no Jacobi identity, no metric and no dimension entering, so it is a property of the series equation 20 rather than of any particular algebra. And the τ dependence is fixed rather than fitted. D vanishes at both endpoints, and to leading order it is symmetric about $\tau = \frac{1}{2}$ and attains its maximum $\|A + B\|/48$ there, so $d(\tau)/d(\frac{1}{2}) = 4\tau(1-\tau)$ to leading order with an $O(\epsilon^2)$ relative remainder. The full deviation need not be symmetric or attain its maximum at the midpoint, the remainder being unconstrained in shape. The exact identity $D(\tau; \xi_0, \xi_1) = D(1-\tau; \xi_1, \xi_0)$, which follows from reversing the path, relates the two orderings of the endpoints and does not by itself constrain the shape for a fixed ordered pair.

The Baker–Campbell–Hausdorff input Every expansion in §3 is built on equation 20, and so is every numerical check we run. An error in it would therefore be invisible. The prediction and the measurement would agree with each other and both would be wrong. We verify it along two independent paths.

Path one, literature. Blanes et al. (2009) obtain the same series as a special case of the Magnus expansion, by taking the piecewise constant generator equal to X_2 on $[0, 1]$ and X_1 on $(1, 2]$, so that the solution at $t = 2$ is $e^{X_1} e^{X_2}$. Their recursion gives

$$\Omega_1 = X_1 + X_2, \qquad \Omega_2 = \tfrac{1}{2}[X_1, X_2],$$
$$\Omega_3 = \tfrac{1}{12}[X_1, [X_1, X_2]] - \tfrac{1}{12}[X_2, [X_1, X_2]], \qquad \Omega_4 = \tfrac{1}{24}[X_1, [X_2, [X_2, X_1]]]. \tag{25}$$

Their conventions match ours. The product is $e^{X_1} e^{X_2}$ with X_1 on the left, the bracket is $[A, B] = AB - BA$, and the grading is by total degree in (X_1, X_2) jointly rather than per argument or by bracket count. With $X_1 = P$ and $X_2 = Q$, equation 25 is equation 20 term by term through degree three.

Path two, symbolic. Computing $\log(e^X e^Y)$ directly in $\mathbb{Q}\langle X, Y\rangle$ and projecting onto a Lie basis in each degree by exact rational elimination gives, with residual exactly zero in every degree,

degree	Lie basis	coefficients	Witt dim. / rank
1	$X,\ Y$	$1,\ 1$	2 / 2
2	$[X, Y]$	$\frac{1}{2}$	1 / 1
3	$[X, [X, Y]],\ [Y, [X, Y]]$	$\frac{1}{12},\ -\frac{1}{12}$	2 / 2
4	$[X, [X, [X, Y]]],\ [Y, [X, [X, Y]]],\ [Y, [Y, [X, Y]]]$	$0,\ -\frac{1}{24},\ 0$	3 / 3

The rank in each degree equals both the basis size and Witt's dimension $\frac{1}{n}\sum_{d|n}\mu(d)2^{n/d} = 2, 1, 2, 3$, so the coefficients are *unique* and no basis element has been omitted.

Reconciling the two degree-four forms. The literature form $\frac{1}{24}[X,[Y,[Y,X]]]$ and the Lyndon-basis form $-\frac{1}{24}[Y,[X,[X,Y]]]$ look unrelated and coincide only through antisymmetry and Jacobi. Expanding both into words gives

$$\tfrac{1}{24}XXYY - \tfrac{1}{12}XYXY + \tfrac{1}{12}YXYX - \tfrac{1}{24}YYXX \tag{26}$$

in each case, so their difference is exactly zero. **We record this because comparing the two by inspection rather than by expansion would have produced a spurious disagreement.**

Two facts from this table are used later. The degree-four component of the series is a *single* basis element: the coefficients on $[X,[X,[X,Y]]]$ and $[Y,[Y,[X,Y]]]$ are exactly zero, not small. And under $X \leftrightarrow Y$ the degree-three component is symmetric while the degree-four component is antisymmetric.

Stage one of the expansion Applying equation 20 with $P = -\xi_0$, $Q = \xi_1$ and using $[-\xi_0, \xi_1] = -[\xi_0, \xi_1]$,

$$\begin{aligned} u = \log\left(e^{-\xi_0}e^{\xi_1}\right) &= (-\xi_0 + \xi_1) - \tfrac{1}{2}[\xi_0,\xi_1] + \tfrac{1}{12}[\xi_0,[\xi_0,\xi_1]] + \tfrac{1}{12}[\xi_1,[\xi_0,\xi_1]] + O(\epsilon^4) \\ &= \Delta - \tfrac{1}{2}[\xi_0,\xi_1] + \tfrac{1}{12}\big(A - B\big) + O(\epsilon^4), \end{aligned} \tag{27}$$

where the last step uses $[\xi_1,[\xi_0,\xi_1]] = -[\xi_1,[\xi_1,\xi_0]] = -B$. Projecting the free-algebra computation of u onto the basis $\{A, B\}$ in degree three returns coefficients $(\frac{1}{12}, -\frac{1}{12})$ with residual zero, confirming equation 27. **The sign between A and B is a minus at this stage**. The plus of equation 8 appears only after stage two.

Third order, term by term Applying equation 20 with $P = \xi_0$, $Q = \tau u$ and writing $u_k = [u]_k$,

$$\log g_{\mathrm{geo}}(\tau) = \xi_0 + \tau u + \tfrac{1}{2}[\xi_0,\tau u] + \tfrac{1}{12}[\xi_0,[\xi_0,\tau u]] - \tfrac{1}{12}[\tau u,[\xi_0,\tau u]] + O(\epsilon^4). \tag{28}$$

Collecting the terms of total degree three, and noting that τ accompanies each factor of u so that only the last term carries τ^2:

$$\big[\log g_{\mathrm{geo}}(\tau)\big]_3 = \underbrace{\tau u_3}_{\text{(i)}} + \underbrace{\tfrac{1}{2}[\xi_0,\tau u_2]}_{\text{(ii)}} + \underbrace{\tfrac{1}{12}[\xi_0,[\xi_0,\tau u_1]]}_{\text{(iii)}} \underbrace{-\tfrac{1}{12}[\tau u_1,[\xi_0,\tau u_1]]}_{\text{(iv)}}. \tag{29}$$

Since $\xi_{\mathrm{lin}}(\tau)$ is of degree one, $[D(\tau)]_3$ equals this sum. Evaluating each term:

	expression	value	coeff. (A, B)
(i)	τu_3	$\frac{\tau}{12}(A - B)$	$\big(\frac{\tau}{12}, -\frac{\tau}{12}\big)$
(ii)	$\frac{1}{2}[\xi_0, \tau u_2]$	$-\frac{\tau}{4}[\xi_0,[\xi_0,\xi_1]]$	$\big(-\frac{\tau}{4}, 0\big)$
(iii)	$\frac{1}{12}[\xi_0,[\xi_0,\tau\Delta]]$	$\frac{\tau}{12}[\xi_0,[\xi_0,\xi_1]]$	$\big(\frac{\tau}{12}, 0\big)$
(iv)	$-\frac{\tau^2}{12}[\Delta,[\xi_0,\Delta]]$	$-\frac{\tau^2}{12}[\Delta,[\xi_0,\xi_1]] = \frac{\tau^2}{12}(A + B)$	$\big(\frac{\tau^2}{12}, \frac{\tau^2}{12}\big)$
	total	$-\frac{\tau(1-\tau)}{12}(A + B)$	$\big(-\frac{\tau(1-\tau)}{12}, -\frac{\tau(1-\tau)}{12}\big)$

Three elementary identities are used. In (ii) and (iii), $[\xi_0, \Delta] = [\xi_0, \xi_1 - \xi_0] = [\xi_0, \xi_1]$. In (iv), the same identity followed by equation 23, whose proof is

$$[\Delta,[\xi_0,\xi_1]] = [\xi_1,[\xi_0,\xi_1]] - [\xi_0,[\xi_0,\xi_1]] = -[\xi_1,[\xi_1,\xi_0]] - [\xi_0,[\xi_0,\xi_1]] = -(A + B), \tag{30}$$

using only antisymmetry of the bracket. Summing the A column, $\frac{\tau}{12} - \frac{\tau}{4} + \frac{\tau}{12} + \frac{\tau^2}{12} = -\frac{\tau}{12} + \frac{\tau^2}{12} = -\frac{\tau(1-\tau)}{12}$, and identically for B.

The mechanism in one line. Terms (i)–(iii) are linear in τ and contribute $-\tau/12$ to each of A and B. Term (iv) is the only quadratic one and contributes $+\tau^2/12$. The profile $\tau(1-\tau)$ is therefore the difference between the linear and quadratic contributions, and it is fixed with no quantity left to choose.

Independent confirmation. Computing $D(\tau)$ in $\mathbb{Q}\langle X, Y\rangle$ at $\tau \in \{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{2}{5}, \frac{3}{7}, \frac{9}{11}, \frac{1}{6}\}$ and subtracting $-\frac{\tau(1-\tau)}{12}(A + B)$ gives exactly zero in every case.

Second order, and why the cancellation is universal The degree-two component receives exactly two contributions,

$$\tau u_2 = -\frac{\tau}{2}[\xi_0, \xi_1], \qquad \tfrac{1}{2}[\xi_0, \tau u_1] = \frac{\tau}{2}[\xi_0, \Delta] = +\frac{\tau}{2}[\xi_0, \xi_1], \tag{31}$$

which sum to zero. Only antisymmetry and bilinearity of the bracket are used. No Jacobi identity, no metric, no dimension. The cancellation is therefore a property of the BCH series rather than of any particular algebra, and it holds in the free algebra, which is how we verify it: $[D(\tau)]_2 = 0$ exactly for every τ tested.

The fourth order vanishes

Proposition 13. $[D(\tau)]_4 = 0$ *identically, for every Lie algebra and every* τ.

Proof. Work in $\mathbb{Q}\langle X, Y\rangle$ truncated at degree five with exact rational coefficients, and compute

$$D(\tau) = \log\left(e^X \exp\left(\tau \log(e^{-X} e^Y)\right)\right) - \left((1-\tau)X + \tau Y\right) \tag{32}$$

by series, for rational τ.

First, a degree bound in τ. In the expression above, τ appears only multiplying u, and u has no component of degree zero. A term of total degree four in (X, Y) therefore contains at most four factors of u and carries τ^k with $k \leq 4$. Hence $[D(\tau)]_4$ is, coefficient by coefficient in the word basis, a polynomial in τ of degree at most four.

Second, evaluation. At the seven rational values $\tau \in \{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{2}{5}, \frac{3}{7}, \frac{9}{11}, \frac{1}{6}\}$ every word coefficient of $[D(\tau)]_4$ is exactly zero. A polynomial of degree at most four vanishing at five or more distinct points is the zero polynomial, and seven values were used.

Since an identity in the free algebra holds under every homomorphism onto a Lie algebra, $[D(\tau)]_4 = 0$ in general. □

The same computation gives $[D(\tau)]_5 \neq 0$, so degree five is the first non-vanishing order and the remainder in equation 8 is $O(\epsilon^5)$.

What the computation certifies, and how to check it. The argument is an interpolation argument: each degree-four word coefficient is a polynomial in τ of degree at most four, so exact vanishing at five distinct rational values of τ implies identical vanishing. The evaluation is carried out in exact rational arithmetic in the free associative algebra, so no floating-point cancellation is involved.

C.4 THE TWO DEGENERATE CASES

The gap D_{geo} of §3.1 vanishes identically under a bi-invariant metric, and both deviations vanish on an abelian algebra within the domain of Proposition 14. Neither case arises on $\mathrm{SE}(3)$.

Bi-invariant metrics. If $\mathrm{ad}_x^*(Mx) = 0$ for every $x \in \mathfrak{g}$ then $\Lambda \equiv 0$, so equation 1 reduces to $\dot{\xi} = 0$ and equation 2 is the geodesic at every amplitude with $D_{\text{geo}} \equiv 0$. On a connected group the condition $\mathrm{ad}_x^*(Mx) = 0$ for all x is equivalent to Ad-invariance of M, hence to bi-invariance of the metric. $\mathrm{SE}(3)$ admits no bi-invariant Riemannian metric (Žefran et al., 1999), so no choice of M removes the second-order gap there for every endpoint pair.

Proposition 14 (Abelian algebras)**.** *If every bracket vanishes then* $A = B = 0$ *and Theorem 2 gives* $D(\tau) \equiv 0$ *exactly, at every order, so with* ξ_0, ξ_1 *the principal logarithms of the endpoints, equation 2 and equation 3 coincide for every pair whose difference* $\xi_1 - \xi_0$ *lies in the domain of the principal logarithm.*

The proof is one line. On an abelian algebra exp is a homomorphism, so $g_0 \exp(\tau u) = \exp(\xi_0 + \tau u)$, and this is equation 3 exactly when $u = \log\exp(\xi_1 - \xi_0)$ equals $\xi_1 - \xi_0$. This recovers the Euclidean case, where that domain is the whole algebra, and it also covers the flat torus, on which torsion-angle models operate: within half a period on every circle the deviation there is structurally zero rather than small. The restriction is set by the branch of the logarithm. Beyond half a period on some circle the principal logarithm changes branch and u differs from $\xi_1 - \xi_0$ by a period, so equation 2 remains the minimising geodesic while equation 3 takes the longer arc, and D is of order one, set by the branch rather than by a bracket. A flat compact group therefore keeps its cut locus, and with it the branches of its logarithm, where it loses its deviation, which separates the two mechanisms attributed jointly to compact manifolds.

C.5 FACTORIZED PATHS ON A PRODUCT

This appendix supports the opening claim of §3.2, that a factorized path on a Riemannian product is the product geodesic, and states the schedule condition under which that holds.

Let (N, g) be the Riemannian product of (N_i, g_i), $i = 1, \dots, n$, let γ_i be the minimising constant-speed geodesic of g_i between the i-th components of the endpoints, and let $\gamma(\tau) = \big(\gamma_1(a_1(\tau)), \dots, \gamma_n(a_n(\tau))\big)$ for schedules $a_i : [0,1] \to [0,1]$ with $a_i(0) = 0$ and $a_i(1) = 1$.

Proposition 15 (Factorized paths on a product)**.** *γ is the minimising constant-speed geodesic of g between its endpoints if and only if $a_1 = \cdots = a_n$ and that common schedule is the identity. Writing ℓ_i for the length of γ_i in g_i and $m_i = \ell_i^2 / \sum_j \ell_j^2$, the two ways of failing are separately quantified. The image of γ departs from the geodesic segment by*

$$\frac{\operatorname{dist}\big(\gamma(\tau),\, \textit{chord}\big)}{\textit{chord length}} = \Big(\textstyle\sum_i m_i \big(a_i(\tau) - \bar{a}(\tau)\big)^2 \Big)^{1/2}, \qquad \bar{a} = \textstyle\sum_i m_i a_i, \tag{33}$$

which vanishes exactly when the schedules agree and is at most $\frac{1}{2}$ whatever they are, and the kinetic energy exceeds its minimum by the factor $\int_0^1 \sum_i m_i \dot{a}_i^2 \, d\tau \geq 1$, with equality exactly when every factor of positive length has $a_i(\tau) = \tau$ almost everywhere, schedules on zero-length factors being irrelevant. When all schedules coincide this reduces to $\int_0^1 \dot{a}^2 \, d\tau$, which is the form used in the special case below, with equality only for the identity schedule.

Proof. Geodesics of a product metric are products of geodesics of the factors sharing one affine parameter, so the stated condition is sufficient. For the converse and for equation 33, note that only the reduced data $\big(a_i(\tau)\big)_i$ and $(\ell_i)_i$ enter: the displacement of $\gamma(\tau)$ along the i-th factor has g_i-length $a_i(\tau)\ell_i$ and the factors are g-orthogonal, so distances to the chord are those of the point $d(\tau) = (a_1\ell_1, \dots, a_n\ell_n)$ from the line through 0 and $c = (\ell_1, \dots, \ell_n)$ in Euclidean $\mathbb{R}^n$. Writing $\hat{c} = c/\|c\|$ gives $\langle d, \hat{c}\rangle = \|c\|\bar{a}$ and $d - \langle d, \hat{c}\rangle \hat{c} = \big((a_i - \bar{a})\ell_i\big)_i$, whose norm divided by $\|c\|$ is equation 33. It vanishes for all τ exactly when every a_i equals $\bar{a}$, that is when the schedules agree. The bound follows because equation 33 is the standard deviation of the values $a_i(\tau) \in [0,1]$ under the probability weights m, and a distribution supported in an interval of length one has variance at most $\frac{1}{4}$. When the schedules agree at the common value a, the image is the chord and the speed is $\|c\|\dot{a}$, so the energy is $\|c\|^2 \int_0^1 \dot{a}^2$, minimised at $\int_0^1 \dot{a}^2 = 1$ by Cauchy–Schwarz with $\int_0^1 \dot{a} = 1$, with equality only when $\dot{a}$ is constant. □

Two consequences are worth separating. Different constant weights on the factors do not matter, since they enter only through m and leave the condition itself untouched, which is the product-metric form of the block-scaling invariance of §3.3. And a schedule shared by every factor but not affine leaves the image correct while raising the energy, so a path can trace the geodesic segment and still not be kinetic-optimal. The two failures are independent and are measured by different quantities.

Proposition 15 assumes a metric product, and on $\mathrm{SE}(3)$ the two propositions nest rather than compete. Under a block-diagonal $M = \operatorname{diag}(J, \beta I_3)$ the translational term in body coordinates is $\beta\|\dot{p}\|^2$, since R is orthogonal, so the metric splits and $(\mathrm{SE}(3), M)$ is a Riemannian product of $(\mathrm{SO}(3), J)$ with Euclidean $\mathbb{R}^3$ whatever J is. Proposition 15 therefore applies, and its schedule condition is automatic because the two factors of equation 4 share τ by construction. Its remaining hypothesis, that each factor path be a minimising geodesic of its factor metric, holds on the translational factor because a straight line minimises the Euclidean length. On the rotational factor it is the question of whether $\exp(\tau\hat{\omega})$ is a geodesic of $(\mathrm{SO}(3), J)$, which the criterion below answers for stationarity, minimality requiring in addition that $\|\omega\|$ lie inside the injectivity radius of $(\mathrm{SO}(3), J)$. The criterion is thus the prerequisite and Proposition 15 takes over once it holds. If instead M couples the two blocks, $(\mathrm{SE}(3), M)$ is not a Riemannian product, so neither statement applies.

C.6 CHART-INDUCED METRICS ARE FLAT

This appendix states and proves the flatness lemma behind Corollary 5 of §3.2. The statement is exact rather than asymptotic: where it applies, no expansion is involved and there is no remainder to bound.

Lemma 16 (Geodesics of a pullback metric)**.** *Let $U \subseteq G$ be open, let Φ map U diffeomorphically onto a convex open subset of $\mathbb{R}^n$, and give U the pullback $\Phi^*\langle\cdot,\cdot\rangle$. Then Φ is an isometry, the metric is flat, and for every $g_0, g_1 \in U$ the unique minimising geodesic is*

$$\tau \mapsto \Phi^{-1}\big((1-\tau)\Phi(g_0) + \tau\Phi(g_1)\big). \tag{34}$$

Attribution. Lemma 16 builds on an observation due to Collas et al. (2025), who use pullback metrics induced by global diffeomorphisms to reduce Riemannian conditional flow matching on matrix manifolds to the Euclidean case. Their Proposition 1 is the statement at the level of the flow matching loss, and ours is its analogue at the level of a single pair of endpoints, which is the level the criterion of §3.2 operates at. Corollary 5 then identifies the metric a rotational loss written in a fixed logarithm chart induces on $\mathrm{SO}(3)$ as one of this kind, and a fixed rotation of its eigenframe as one that is not.

Proof of Lemma 16. By construction Φ is an isometry from U onto its image. An isometry carries geodesics to geodesics and minimisers to minimisers, and flatness is preserved because the curvature tensor is an isometry invariant. The geodesics of the Euclidean metric on a convex set are its straight segments in affine parametrisation, and between two points the segment is unique, so equation 34 is the unique minimiser. □

Proof of Corollary 5. The map $\Phi(R) = \log(R_0^\top R)$ is a diffeomorphism from $\mathrm{SO}(3)$ minus the cut locus of R_0 onto the open ball of radius π in $\mathbb{R}^3$, which is convex, and it carries $R(\tau) = R_0 \exp(\tau\hat{\omega})$ to the segment $\tau\omega$. Writing $\phi = \Phi(R)$, the right Jacobian relates the chart velocity to the body velocity by $\xi = \mathcal{J}_r(\phi)\dot{\phi}$, so

$$\xi^\top J(\tau)\xi = \xi^\top \mathcal{J}_r(\phi)^{-\top}\mathcal{J}_r(\phi)^{-1}\xi = \|\mathcal{J}_r(\phi)^{-1}\xi\|^2 = \|\dot{\phi}\|^2, \tag{35}$$

which is exactly the pullback $\Phi^*\langle\cdot,\cdot\rangle$ of Lemma 16. The lemma then gives the conclusion, with the endpoints $\Phi(R_0) = 0$ and $\Phi(R_1) = \omega$ both in the ball when $\|\omega\| < \pi$.

On $\mathrm{SE}(3)$ the chart is $\Psi(R, p) = \big(\Phi(R),\, p - p_0\big)$, a diffeomorphism onto the product of that ball with $\mathbb{R}^3$, again convex. The induced metric is the orthogonal sum of the rotational pullback above and the Euclidean metric on translations, and Ψ carries the factorized path equation 4 to the segment $\tau(\omega, w)$. □

Two remarks bound the statement. The anchor may be either endpoint, since $\log(R_1^\top R)$ carries the same path to the segment $(\tau - 1)\omega$, equally straight. And the conclusion depends on the chart rather than on the eigenvalues: a fixed rotation of the eigenframe, $\tau \mapsto QJ(\tau)Q^\top$, leaves every eigenvalue unchanged but takes the metric outside the representation Corollary 5 uses, since $\mathcal{J}_r(\phi)^{-1}Q^\top$, the square root of $QJ(\tau)Q^\top$, need not be the differential of any map. Whether some other chart realises it as a pullback is an integrability question we leave open, and §5.2 measures the two metrics against each other.

C.7 SCALING OF THE ALGEBRA-VERSUS-GROUP DEVIATION

This appendix quantifies D, the smaller of the two deviations, whose structure admits closed-form constants and a dominance ratio between its two components. The larger gap needs neither, since its leading term is the single generator $\Lambda(u)$, which §5 evaluates directly. On a semidirect product $\mathfrak{g} \ltimes V$, with bracket

$$\big[(X_1, v_1), (X_2, v_2)\big] = \big([X_1, X_2]_{\mathfrak{g}},\ X_1 \cdot v_2 - X_2 \cdot v_1\big), \tag{36}$$

the deviation splits into two contributions that scale differently:

$$A + B = (R,\, C), \qquad \|R\| \propto \lambda^3, \qquad \|C\| \propto \lambda^2 \mu \quad \text{under} \quad X \mapsto \lambda X,\ v \mapsto \mu v, \tag{37}$$

where R collects the double brackets of the $\mathfrak{g}$ components and C the action terms (Appendix C.7 for the explicit forms). The scaling holds for any dimension, representation, and inner product, because the $\mathfrak{g}$ component of a semidirect bracket never involves V and every term of C is bilinear in the $\mathfrak{g}$ arguments. If $\mathfrak{g}$ is abelian only the coupling survives and the deviation is second order in θ. If the action is trivial the factors decouple and $C = 0$.

Aggregation over endpoints requires a distribution. Under the canonical law in which the four directions are independent and uniform and the magnitudes are equal, $\|X_i\| = \theta$, $\|p\| = \|q\| = t$, the magnitudes factor out, $R = \theta^3 \hat{R}$, $C = \theta^2 t\, \hat{C}$, and the directional constants have closed forms on $\mathfrak{so}(3) \ltimes \mathbb{R}^3$ (Appendix C.7):

$$\kappa_R^{(1)} = \mathbb{E}\|\hat{R}\| = \tfrac{16}{15}, \qquad \kappa_R^{(2)} = \big(\mathbb{E}\|\hat{R}\|^2\big)^{1/2} = \tfrac{2}{\sqrt{3}}, \qquad \kappa_T^{(2)} = \big(\mathbb{E}\|\hat{C}\|^2\big)^{1/2} = 2 \text{ exactly}, \tag{38}$$

with $\kappa_T^{(1)}$ reported as a Monte Carlo estimate. Assuming further that magnitudes are independent of directions,

$$\mathbb{E}\big[d_R(\tfrac{1}{2})\big] = \frac{\kappa_{\mathrm{amp}}\, \theta_{50}^3\, \kappa_R^{(1)}}{48} = \frac{\kappa_{\mathrm{amp}}\, \theta_{50}^3}{45}, \qquad \kappa_{\mathrm{amp}} = \frac{\mathbb{E}[\theta^3]}{\theta_{50}^3}, \tag{39}$$

where d_R is the *rotational* component of the deviation, measured as an angle in radians. The constants 45 and 48 are one statement, with $\kappa_R^{(1)}$ already absorbed into 45, so applying both scales the correction twice. The ratio $r_R = 45\, \mathbb{E}[d_R]/\mathbb{E}[\theta^3]$ equals one under the canonical law. A value $r \neq 1$ therefore certifies a violation. A value $r = 1$ does not certify the law, since distinct violations can in principle cancel, and the law decomposes into three claims that are testable independently: directional isotropy, equal endpoint magnitudes, and independence of the $\mathfrak{g}$ and V directions.

These constants describe the rotational block alone. The ratio $r_{RT} = d_R/d_T$ carries the lever arm L through $\theta L/t$, and where it falls below one the translational component dominates and the ratio sits in the $r_{RT} \ll 1$ branch of equation 40. The six-dimensional norms $\|D_{\mathrm{geo}}\|$ and $\|D\|$ are a different quantity from d_R.

Two quantities are external and must be stated by the application: the length scale L converting a $\mathfrak{g}$-valued deviation into a V-valued one, for a rigid motion the distance from the frame origin to the point whose accuracy matters, and the tolerance ε in the units of V. Combining the two components in quadrature, $d^2 = (L\, d_R)^2 + d_T^2$ with $d_R = \kappa_R^{(2)} \theta^3/48$ and $d_T = \kappa_T^{(2)} \theta^2 t/48$, the second-moment constants being the appropriate ones for a quadrature. The relation is exact pointwise when the two blocks are orthogonal in the reporting metric and is a second-moment approximation otherwise:

Proposition 17 (Dominance ratio and thresholds)**.** *Let* $r_{RT} = d_R/d_T = (\kappa_R/\kappa_T)(\theta L/t)$ *and* $s(\theta) = d \log d / d \log \theta$*. Under the quadrature assumption,*

$$s(\theta) = 2 + \frac{r_{RT}^2}{1 + r_{RT}^2}, \qquad \theta^* = \Big(\frac{48\varepsilon}{\kappa_R L}\Big)^{1/3}\ (r_{RT} \gg 1), \qquad \theta^* = \sqrt{\frac{48\varepsilon}{\kappa_T t}}\ (r_{RT} \ll 1), \tag{40}$$

where θ^* *is the amplitude at which* d *reaches* ε*.*

The exponent increases monotonically from 2 to 3 and is real-analytic throughout: there is no critical point and no transition, and dominance in the sense $s > 2.9$ requires $r_{RT} \gtrsim 3$, not $r_{RT} > 1$. Since the quadrature is a second-moment operation, the self-consistent constants are the second-moment pair, $\kappa_R^{(2)}/\kappa_T^{(2)} = 1/\sqrt{3}$, which at the reference point $\theta = t/L$ gives $s = 9/4$ exactly. Of every symbol in equation 40, only ε is external, and r_{RT} does not contain it, so we report r_{RT} as the primary quantity and leave the tolerance to the reader.

Proofs of the split and of the exponents. The split $A + B = (R, C)$ of §C.7 has $R = [X_0, [X_0, X_1]] + [X_1, [X_1, X_0]]$ and $C = X_0(X_0 q - X_1 p) - [X_0, X_1]p + X_1(X_1 p - X_0 q) - [X_1, X_0]q$, the translational components of ξ_0, ξ_1 being written p and q so that u stays reserved for the relative logarithm.

Proof. Substituting equation 36 into equation 7 gives $[\xi_0, \xi_1] = \big([X_0, X_1],\ X_0 \cdot q - X_1 \cdot p\big)$ and hence

$$A = \Big([X_0, [X_0, X_1]],\ X_0 \cdot (X_0 \cdot q - X_1 \cdot p) - [X_0, X_1] \cdot p\Big), \tag{41}$$

with B obtained by exchanging the indices, and adding gives equation 37. The exponents follow from the shape of the two components. The $\mathfrak{g}$-component of a semidirect bracket never involves V, so R is trilinear in X_0, X_1 alone. Every term of C has the form $X \cdot (X \cdot v)$ or $[X, X] \cdot v$ and is therefore bilinear in the $\mathfrak{g}$ arguments and linear in the V argument. The effective exponent of equation 40 is then

$$s = \frac{6d_R^2 + 4d_T^2}{2(d_R^2 + d_T^2)} = 2 + \frac{d_R^2}{d_R^2 + d_T^2} = 2 + \frac{r_{RT}^2}{1 + r_{RT}^2}. \tag{42}$$

□

The $\mathrm{SE}(3)$ instance. Proposition C.7 is general. On $\mathrm{SE}(3) = \mathrm{SO}(3) \ltimes \mathbb{R}^3$ the action is $X \cdot v = \omega \times v$ and the bracket equation 36 becomes

$$[(\omega_1, v_1), (\omega_2, v_2)] = \big(\omega_1 \times \omega_2,\ \omega_1 \times v_2 - \omega_2 \times v_1\big). \tag{43}$$

With $\xi_0 = (a, p)$ and $\xi_1 = (b, q)$,

$$[\xi_0, \xi_1] = \big(a \times b,\ a \times q - b \times p\big), \tag{44}$$

$$A = [\xi_0, [\xi_0, \xi_1]] = \big(a \times (a \times b),\ a \times (a \times q - b \times p) - (a \times b) \times p\big), \tag{45}$$

$$B = [\xi_1, [\xi_1, \xi_0]] = \big(b \times (b \times a),\ b \times (b \times p - a \times q) - (b \times a) \times q\big). \tag{46}$$

Summing, the rotational component is $a \times (a \times b) + b \times (b \times a) = (a - b) \times (a \times b)$ and the translational component reduces to $(a - b) \times (a \times q - b \times p) + (a \times b) \times (q - p)$, which is equation 37. Evaluating the bracket instead as a 4×4 matrix commutator agrees to 10^{-10} over 2000 random draws. Scaling $a, b \mapsto \lambda a, \lambda b$ and $p, q \mapsto \mu p, \mu q$ multiplies the two components by λ^3 and $\lambda^2 \mu$, and the log-log slopes over nine decades read 3.000000 and $(2.000000, 1.000000)$, so the exponents follow from equation 43 rather than from a fit.

The translational block. On $\mathrm{SE}(3)$ the exponential acts on the translational part through

$$V(\omega) = I + \frac{1 - \cos\theta}{\theta^2}\hat{\omega} + \frac{\theta - \sin\theta}{\theta^3}\hat{\omega}^2 = I + \tfrac{1}{2}\hat{\omega} + O(\theta^2), \tag{47}$$

whose linear term $\frac{1}{2}\hat{\omega}$ mixes rotation into translation. As $\theta \to 0$ the semidirect product degenerates to a direct sum and the coupling vanishes, which is Remark C.7 in the instance. The determinant is $\det V(\omega) = \mathrm{sinc}^2(\theta/2)$, verified to 5.6×10^{-16} over $\theta \in (0, \pi)$, and since the first zero of $\mathrm{sinc}(\theta/2)$ is at $\theta = 2\pi$, V is invertible on the whole principal branch with $\det V \geq (2/\pi)^2 \approx 0.405$ on $(0, \pi]$. The residual $\|V(\epsilon\hat{n}) - (I + \frac{1}{2}\widehat{\epsilon\hat{n}})\|$ is $O(\epsilon^2)$, measured at 10^{-8} for $\epsilon = 10^{-4}$.

The constants: reduction to one dimension. For $\hat{a}, \hat{b}$ independent and uniform on S^2, the inner product $c = \hat{a} \cdot \hat{b}$ is uniform on $[-1, 1]$, since fixing $\hat{a}$ and integrating over the polar angle of $\hat{b}$ gives density $\frac{1}{2}$ in c. This collapses the four-dimensional integral to a one-dimensional one.

$\kappa_R^{(1)}$ and $\kappa_R^{(2)}$. With $\hat{R} = (\hat{b} - \hat{a}) \times (\hat{a} \times \hat{b})$ and γ the angle between $\hat{a}$ and $\hat{b}$,

$$\|\hat{R}\| = 4s\sqrt{1 - s}, \qquad s = \sin^2(\gamma/2) = \tfrac{1-c}{2}, \tag{48}$$

verified to 4.7×10^{-15} over 3000 draws. Since c is uniform on $[-1, 1]$, s is uniform on $[0, 1]$, and

$$\kappa_R^{(1)} = \int_0^1 4s\sqrt{1 - s}\, ds = \frac{16}{15}, \qquad \big(\kappa_R^{(2)}\big)^2 = \int_0^1 16s^2(1 - s)\, ds = \frac{4}{3}, \tag{49}$$

so $\kappa_R^{(2)} = 2/\sqrt{3}$. Monte Carlo over 2×10^7 draws at seed 20270101 gives 1.066651 ± 0.000099 and 1.154701 ± 0.000079 against the exact 1.066667 and 1.154701.

M, N, and $\kappa_T^{(2)}$. The coupling component is linear in $\hat{p}$ and $\hat{q}$, so $\hat{C} = M\hat{p} + N\hat{q}$ with

$$M = \hat{a}\hat{b}^\top - 2\hat{b}\hat{a}^\top + \hat{b}\hat{b}^\top - (1-c)I, \tag{50}$$

$$N = \hat{a}\hat{a}^\top - 2\hat{a}\hat{b}^\top + \hat{b}\hat{a}^\top - (1-c)I, \tag{51}$$

verified to 1.3×10^{-15}. Taking the expectation over $\hat{p}, \hat{q}$ independent and uniform, and using $\mathbb{E}[\hat{p}\hat{p}^\top] = \frac{1}{3}I$ and $\mathbb{E}[\hat{p}] = 0$,

$$\mathbb{E}_{\hat{p},\hat{q}}\|\hat{C}\|^2 = \tfrac{1}{3}\operatorname{tr}(M^\top M) + \tfrac{1}{3}\operatorname{tr}(N^\top N) = \tfrac{1}{3}\big(\|M\|_F^2 + \|N\|_F^2\big). \tag{52}$$

Expanding the traces with $\|\hat{a}\| = \|\hat{b}\| = 1$ gives $\|M\|_F^2 + \|N\|_F^2 = 14 - 8c - 6c^2$, verified to 1.1×10^{-14}. This depends on the endpoints only through c, which is what leaves the remaining integral one-dimensional. With $\mathbb{E}[c] = 0$ and $\mathbb{E}[c^2] = \frac{1}{3}$,

$$\mathbb{E}\|\hat{C}\|^2 = \tfrac{1}{3}\big(14 - 0 - 6 \cdot \tfrac{1}{3}\big) = \tfrac{1}{3} \cdot 12 = 4, \tag{53}$$

so $\kappa_T^{(2)} = 2$ exactly. The same run gives $\mathbb{E}\|\hat{C}\|^2 = 4.001086 \pm 0.000844$, a z-score of 1.29.

$\kappa_T^{(1)}$. The first moment requires $\mathbb{E}\|M\hat{p} + N\hat{q}\|$, and the norm does not reduce to a function of c alone, since the direction of $\hat{C}$ depends on $\hat{p}, \hat{q}$ in a way that survives the expectation. No closed form is available, and the measured value is

$$\kappa_T^{(1)} = 1.769655 \pm 0.000208 \qquad (2 \times 10^7 \text{ draws, seed } 20270101). \tag{54}$$

The same run gives $\kappa_T^{(2)} = 2.000271$ against the exact 2, and the ratio of moments is 0.88471, which would be 1 if $\|\hat{C}\|$ were deterministic. The dominance ratio normalises by a second moment, so $\kappa_T^{(1)}$ does not enter it.

C.8 WHICH LOSSES GIVE A LEFT-INVARIANT METRIC

Theorem 1 and Proposition 4 concern a fixed left-invariant metric, that is, a Gram operator constant in body coordinates. In Riemannian flow matching the metric a loss induces is the one its norm is measured in. If the loss is quadratic with a constant form W in coordinates q, and $\dot{q} = B(g)\xi$, the form it induces on body velocities is $M(g) = B(g)^\top W B(g)$, which is constant in body coordinates only when B is. Four cases arise among the losses examined in §5, and different statements reach them.

(i) **A fixed chart.** If the loss and the path both live in one chart that does not move along the path, the induced metric is the pullback of a constant form, it is flat, and the path is its geodesic exactly when it is a straight line in that chart (Lemma 16). SmolVLA is of this kind. Its path is linear and its loss an unweighted mean-squared error in one chart of normalised increments, so $B(g) = \mathrm{dexp}_{-\xi}^{-1}$, and the constant $W = \mathrm{diag}(1/(\sigma_i s_i)^2)$ is the $\|\xi\| \to 0$ limit of $M(g)$ rather than $M(g)$ itself.

(ii) **A chart that moves with the path.** Measuring the rotational residual in a logarithm chart anchored at the current point gives $J(\tau) = \mathcal{J}_r(\phi)^{-\top}\mathcal{J}_r(\phi)^{-1}$ with ϕ changing along the path. No single chart carries the loss, so Lemma 16 does not apply. We freeze the metric at the point of evaluation and apply Proposition 4 to the frozen form. FrameFlow below its gate is of this kind, and since $\mathcal{J}_r(s\omega)\omega = \omega$ for every s, the endpoint axis is an eigenvector of the frozen block at every τ.

(iii) A weighted sum of terms that are not quadratic, absolute values for instance, gives no quadratic form at all. Nothing in §3 applies, and the domain is recorded as such rather than assigned a construction.

(iv) Gating a term on the flow parameter gives a family indexed by τ, chart induced below the gate and carrying an extra block above it. Every statement here concerns a fixed metric, so a quantity computed against such a family is local to the τ at which it is evaluated, not a global claim about the path.

Freezing is an approximation, and it can manufacture a departure. Frozen to its left-invariant limit, SmolVLA's metric makes the same released path depart at second order, while under the pullback it sits below the floor (Appendix D.1). A fixed-chart loss is therefore judged under its pullback, and the frozen form is used only where no fixed chart exists.

C.9 Proposition 7, in full

The proof extends that of Theorem 1 by one order, working in body coordinates from g_0 throughout. Write the geodesic body velocity as $\xi(\tau) = u + a_2(\tau) + a_3(\tau) + O(\epsilon^4)$ with $a_k = O(\epsilon^k)$, and let $D\Lambda(u)[h]$ denote the polar of Λ, so that $\Lambda(u+h) = \Lambda(u) + D\Lambda(u)[h] + \Lambda(h)$.

Velocity, order by order. Substituting into $\dot{\xi} = \Lambda(\xi)$ gives $\dot{a}_2 = \Lambda(u)$ at order ϵ^2, and the endpoint condition below fixes $a_2(\tau) = (\tau - \frac{1}{2})\Lambda(u)$. At order ϵ^3, $\dot{a}_3 = D\Lambda(u)[a_2(\tau)] = (\tau - \frac{1}{2})Q$ with $Q := D\Lambda(u)[\Lambda(u)]$, so $a_3(\tau) = \alpha_3 - \frac{\tau(1-\tau)}{2}Q$ with α_3 constant.

Reconstruction to third order. The Magnus expansion gives

$$\log g_{\mathrm{riem}}(\tau) = \int_0^\tau \xi + \tfrac{1}{2}\int_0^\tau \int_0^s [\xi(s'), \xi(s)]\, ds' ds + O(\epsilon^4), \tag{55}$$

its third term being $O(\epsilon^4)$ because the integrand is a double bracket of $\xi = u + O(\epsilon^2)$ and $[u,[u,u]] = 0$. In the double integral $[\xi(s'), \xi(s)] = [u, a_2(s) - a_2(s')] + O(\epsilon^4) = (s - s')[u, \Lambda(u)] + O(\epsilon^4)$, and $\int_0^\tau \int_0^s (s - s')\, ds' ds = \tau^3/6$, so the bracket contributes $\frac{\tau^3}{12}[u, \Lambda(u)]$.

Endpoint condition. Imposing $\log g_{\mathrm{riem}}(1) = u$ recovers $a_2(0) = -\frac{1}{2}\Lambda(u)$ at order two, as in Theorem 1, and forces $\alpha_3 = \frac{1}{12}\big(Q - [u, \Lambda(u)]\big)$ at order three. Collecting,

$$\log g_{\mathrm{riem}}(\tau) = \tau u - \tfrac{\tau(1-\tau)}{2}\Lambda(u) + \tfrac{\tau(1-\tau)}{12}\Big[(1-2\tau)\, Q - (1+\tau)\, [u, \Lambda(u)]\Big] + O(\epsilon^4). \tag{56}$$

For the corrected path, the Baker–Campbell–Hausdorff formula gives $\log g_{\mathrm{corr}}(\tau) = \tau u + c(\tau)\Lambda(u) + \frac{\tau c(\tau)}{2}[u, \Lambda(u)] + O(\epsilon^4)$ with $c = -\frac{\tau(1-\tau)}{2}$, whose third-order part is $-\frac{\tau^2(1-\tau)}{4}[u, \Lambda(u)]$. Subtracting, the coefficient of Q is $\frac{\tau(1-\tau)(1-2\tau)}{12}$ and that of $[u, \Lambda(u)]$ is $-\frac{\tau(1-\tau)(1+\tau)}{12} + \frac{\tau^2(1-\tau)}{4} = -\frac{\tau(1-\tau)(1-2\tau)}{12}$, which gives equation 13 with $N_2 = Q - [u, \Lambda(u)]$. The two logarithms differ at $O(\epsilon^3)$ while each is $O(\epsilon)$, so their difference equals $\log(g_{\mathrm{corr}}^{-1} g_{\mathrm{riem}})$ to the order retained. Adding $d(\tau)N_2$ inside the correction exponential cancels that term at the cost of cross-brackets that are $O(\epsilon^5)$, which is the second level of Proposition 7.

The third-order formula, measured. Comparing $\log(g_{\mathrm{corr}}^{-1} g_{\mathrm{riem}})$ against $d(\tau)N_2(u)$ at $\tau \in \{0.15, 0.25, 0.75, 0.85\}$, the relative error of the prediction is 7.5%, 3.7% and 1.9% at $\epsilon = 0.2, 0.1, 0.05$. It halves with ϵ, which is the $O(\epsilon^4)$ remainder against an $O(\epsilon^3)$ term. At $\tau = \frac{1}{2}$ the prediction is identically zero and the measured residual is the next order, as the profile requires.

The two levels at a generic parameter. Table 12 reports the residuals at $\tau = \frac{1}{4}$, away from the midpoint where the third-order term vanishes.

Table 12: **Each level removes one order.** Deviation from the shot geodesic at $\tau = \frac{1}{4}$, medians over five generic endpoint pairs at $M = I$. All twelve cells come from one run on the same pairs.

ϵ	uncorrected	corrected	twice-corrected
0.2	7.908×10^{-3}	4.339×10^{-4}	3.119×10^{-5}
0.1	1.924×10^{-3}	5.273×10^{-5}	1.896×10^{-6}
0.05	4.742×10^{-4}	6.498×10^{-6}	1.168×10^{-7}
0.025	1.177×10^{-4}	8.064×10^{-7}	7.302×10^{-9}

The local slopes are 2.04, 2.02, 2.01 uncorrected, 3.04, 3.02, 3.01 corrected, and 4.04, 4.02, 4.00 twice-corrected, which is the progression $2 \to 3 \to 4$ of §3.3.

The velocity field. On the group base with $M = I$, integrating the exact field equation 12 reproduces g_{corr} to 2.83×10^{-15} at $\epsilon = 0.2$, and to 4.3×10^{-17} and 7.3×10^{-17} at $\epsilon = 0.1$ and 0.05. Integrating its third-order form instead gives endpoint errors with slopes 2.95 and 2.98 against a predicted 3.

Block-scaling invariance. Across $M = \mathrm{diag}(\alpha I, \beta I)$ with $(\alpha, \beta) \in \{(1,1), (1, 7.3), (12.9, 1), (0.03, 41)\}$, the generator $\Lambda_M(u)$ agrees with its reference value to at most 2.7×10^{-15}, and the shot geodesics coincide as curves to at most 3.1×10^{-15}. Both are at the apparatus floor, so scaling one block against the other leaves the optimal path unchanged.

The factorized path. Its pointwise Euler–Poincaré residual is at most 2.2×10^{-17} across the block-isotropic metrics tested, and it agrees with the shot geodesic to 1.7×10^{-13} even at $\epsilon = 0.4$, that is $23°$. This is floor-level agreement at an amplitude where the one-parameter subgroup path deviates by 3×10^{-2}. Under $M = \mathrm{diag}(J, I)$ with $J = \mathrm{diag}(1, 2, 4)$ its deviation from the M-geodesic at $\tau = \frac{1}{2}$ is 7.50×10^{-3}, 1.89×10^{-3} and 4.75×10^{-4} at $\epsilon = 0.2, 0.1, 0.05$, with local slopes 1.99 and 2.00.

Modularity under anisotropy. With the same anisotropic M at $\tau = \frac{1}{4}$, the subgroup path corrected by the full Λ_M and twice-corrected by N_2 computed from M shows slopes 2.07, 2.03 then 3.08, 3.04 then 4.00, 4.09 for the three arms. The factorized path corrected by the gap generator $\Lambda_M - \Lambda_{\mathrm{iso}}$ reaches 1.08×10^{-4}, 6.65×10^{-6} and 4.01×10^{-7} at $\tau = \frac{1}{2}$, with slopes 4.02 and 4.05, the midpoint again gaining one order. The construction therefore applies unchanged under an anisotropic metric.

C.10 THE CORRECTIONS AS A COUPLING HOMOTOPY

§3.3 derives the two corrections level by level. This appendix exhibits the object both levels are coefficients of and rederives the third-order residual equation 13 from it by an independent route. Fix endpoints g_0, g_1 with $u = \log(g_0^{-1} g_1)$ and a left-invariant metric M with operator Λ, and insert a coupling parameter s into the geodesic equation,

$$\dot{\xi}_s(\tau) = s\,\Lambda(\xi_s(\tau))\,, \qquad s \in [0,1], \tag{57}$$

with g_s defined by $g_s^{-1}\dot{g}_s = \xi_s$, $g_s(0) = g_0$, and $g_s(1) = g_1$ imposed for every s. At $s = 0$ the velocity is constant, the endpoint forces $\xi_0(0) = u$, and the path is $g_0 \exp(\tau u)$. At $s = 1$ the family is the geodesic boundary-value problem. The two levels are the leading amplitude terms of the Taylor coefficients of this family at $s = 0$.

Proposition 18 (Coupling homotopy)**.** *Let $\Omega_s(\tau) = \log\big(g_0^{-1} g_s(\tau)\big)$ and expand $\Omega_s = \tau u + s\,\Omega_1 + s^2\Omega_2 + O(s^3)$. Writing $P = D\Lambda(u)[\Lambda(u)]$ and grading by homogeneous degree in u,*

$$\Omega_1(\tau) = -\,\frac{\tau(1-\tau)}{2}\,\Lambda(u) \;-\; \frac{\tau(1-\tau)(1+\tau)}{12}\,[u, \Lambda(u)] \;+\; O(\epsilon^4), \tag{58}$$

$$\Omega_2(\tau) = \frac{\tau(1-\tau)(1-2\tau)}{12}\,P \;+\; O(\epsilon^4), \tag{59}$$

and in the multiplicative coordinates of equation 11 the degree-three terms of $s\,\Omega_1 + s^2\Omega_2$ at $s = 1$ combine into

$$\log\Big(g_{\mathrm{corr}}^{(1)}(\tau)^{-1} g_{\mathrm{riem}}(\tau)\Big) = \frac{\tau(1-\tau)(1-2\tau)}{12}\,N_2(u) + O(\epsilon^4), \tag{60}$$

which is equation 13.

Proof. Substituting the series for ξ_s into equation 57 gives linear equations with polynomial forcing. First $\dot{\xi}^0 = 0$ so $\xi^0 \equiv u$, then $\dot{\xi}^1 = \Lambda(u)$ so $\xi^1(\tau) = a_1 + \tau\Lambda(u)$, and $\dot{\xi}^2 = D\Lambda(u)[\xi^1]$ so $\xi^2(\tau) = a_2 + \tau\, D\Lambda(u)[a_1] + \frac{\tau^2}{2}P$.

The endpoint condition lives on the group and is transported to the algebra through $\dot{\Omega} = \xi + \frac{1}{2}[\Omega, \xi] + \frac{1}{12}[\Omega, [\Omega, \xi]] + \cdots$ with $\Omega_s(1) = u$ at every order in s. At order s^1 and degree two, $\dot{\Omega}_1 = \xi^1$ gives $\Omega_1 = a_1\tau + \Lambda(u)\tau^2/2$ and $a_1 = -\Lambda(u)/2$, the first term of equation 58, and the departure $-\int_0^\tau (\xi - u)\,d\sigma = \Lambda(u)\tau(1-\tau)/2$ recovers Theorem 1. At degree three the brackets enter through $\frac{1}{2}[\tau u, \xi^1] + \frac{1}{2}[\Omega_1, u]$, whose degree-three part is $\frac{\tau^2}{4}[u, \Lambda(u)]$. Integrating and imposing $\Omega_1(1) = 0$ fixes the degree-three part of a_1 to $-[u, \Lambda(u)]/12$ and gives the second term, with profile $\tau^3 - \tau = -\tau(1-\tau)(1+\tau)$. At order s^2 every bracket has degree at least four, so $\dot{\Omega}_2 = \xi^2$ at degree three. Inserting a_1 gives $\dot{\xi}^2 = (\tau - \frac{1}{2})P$, integration gives $\Omega_2 = a_2\tau + P(\tau^3/6 - \tau^2/4)$, and $\Omega_2(1) = 0$ gives $a_2 = P/12$, which is equation 59.

For equation 60, pass to the right-multiplied corrector of equation 11 through $\log\big(e^{-F_1} e^{-\tau u} e^{\Omega}\big) = \Omega - \tau u - F_1 - \frac{1}{2}[\tau u, \Omega - \tau u] + O(\epsilon^4)$ with $F_1 = -\Lambda(u)\tau(1-\tau)/2$. The three degree-three pieces are the tail of equation 58, the whole of equation 59, and the cross bracket $\frac{\tau^2(1-\tau)}{4}[u, \Lambda(u)]$. The two bracket profiles combine as $-\frac{\tau(1-\tau)(1+\tau)}{12} + \frac{\tau^2(1-\tau)}{4} = -\frac{\tau(1-\tau)(1-2\tau)}{12}$, so the total is $\frac{\tau(1-\tau)(1-2\tau)}{12}\,(P - [u, \Lambda(u)]) = N_2(u)$ by equation 10. □

Why the coupling and the amplitude are one expansion. Λ is quadratic, so $\Lambda(\epsilon u) = \epsilon^2\Lambda(u)$ and each application of the coupling raises the amplitude degree by one: the s^k coefficient has leading degree $k + 1$ in u. Switching on the coupling and shrinking the amplitude are therefore the same expansion at leading order. The correspondence is not exact, because the group endpoint condition is not homogeneous, and its Magnus corrections produce the degree-three tail in equation 58. That tail is the commutator half of N_2, reabsorbed at the next level by equation 60, so the homotopy rederives equation 13 including the coefficient $1/12$ and the term $-[u, \Lambda(u)]$, by a route independent of the Magnus computation of Appendix C.9.

What the homotopy controls, and what it does not. The implicit-function construction requires only that the shooting Jacobian be nonsingular, which is local. The radius in s of the Taylor series is a separate matter, since a series in an artificial coupling can be limited by complex singularities in s, by a logarithm branch, or by loss of analyticity, none of which is a conjugate point, and the amplitude expansion can fail for reasons unrelated to one. For $s \neq 1$ the family is not the geodesic equation of any metric we have identified. In this vocabulary the construction is a predictor-only continuation with closed-form predictors, the ratio $\|F_\ell\|/\|u\|$ measures how far along it a pair extrapolates, and appending a Newton corrector would reinstate the solver the construction exists to avoid. The first conjugate point of the metric in use bounds that trust region from above.

C.11 THE GENERAL BASE, AND THE OBSTRUCTION AT LEVEL TWO

Proposition 7(i) is proved for any base path whose body velocity satisfies $\dot{\xi} = \Lambda_0(\xi)$ for a fixed generator. Write $g_{\mathrm{riem}} = g_{\mathrm{base}} \exp(F)$. The body velocity of the right-hand side is

$$\xi_{\mathrm{riem}} = \mathrm{Ad}_{e^{-F}} \xi_{\mathrm{base}} + \mathrm{dexp}_{-F}(\dot{F}) = \xi_{\mathrm{base}} + \dot{F} + \mathrm{ad}_{\xi_{\mathrm{base}}} F + O(\|F\|^2 \|\xi_{\mathrm{base}}\| + \|F\|\,\|\dot{F}\|), \quad (61)$$

so with $F = O(\|u\|^2)$, $\dot{F} = O(\|u\|^2)$ and $\xi_{\mathrm{base}} = u + O(\|u\|^2)$ every displayed remainder is $O(\|u\|^3)$ and $\delta\xi := \xi_{\mathrm{riem}} - \xi_{\mathrm{base}} = \dot{F} + O(\|u\|^3)$. Subtracting the two Euler–Poincaré equations,

$$\delta\dot{\xi} = \Lambda(\xi_{\mathrm{riem}}) - \Lambda_0(\xi_{\mathrm{base}}) = \Lambda(u) - \Lambda_0(u) + D\Lambda(u)[\delta\xi] + O(\|u\|^3), \quad (62)$$

and since Λ is quadratic, $D\Lambda(u)[\delta\xi] = O(\|u\|) \cdot O(\|u\|^2) = O(\|u\|^3)$. Hence $\ddot{F} = \Lambda(u) - \Lambda_0(u) + O(\|u\|^3)$ and integrating twice with $F(0) = F(1) = 0$ gives $F = -\frac{\tau(1-\tau)}{2}(\Lambda(u) - \Lambda_0(u)) + O(\|u\|^3)$, which is F_1. The residual is therefore $O(\|u\|^3)$, uniformly in τ.

Two checks. With $\Lambda_0 = 0$ this reduces to the group-interpolation case of Theorem 1. On $\mathrm{SE}(3)$ with the factorized base, $\Lambda_0 = \Lambda_{\mathrm{iso}}$ and the base body velocity $\xi = (\omega, R^\top w)$ satisfies $\dot{\xi} = (0, \nu \times \omega) = \Lambda_{\mathrm{iso}}(\xi)$ exactly, so the hypothesis holds. Numerically, under $M = \mathrm{diag}(1, 2, 4; I)$ with fixed directions and decreasing amplitude, the uncorrected factorized path has residual slopes 2.00 to 2.07 and the corrected path 2.97 to 3.07 over three seeds, consistent with the orders claimed.

Which constructions the hypothesis admits. The condition $\dot{\xi} = \Lambda_0(\xi)$ for a fixed generator is not vacuous, and of the three constructions of §3 it admits exactly two. Group interpolation has constant body velocity, so $\Lambda_0 = 0$. The factorized path on $\mathrm{SE}(3)$ has $\xi = (\omega, R^\top w)$ with $\dot{\xi} = (0, \nu \times \omega) = \Lambda_{\mathrm{iso}}(\xi)$, an exact solution. Algebra interpolation is neither. Its body velocity is $\xi = \mathrm{dexp}_{-\zeta(\tau)}(\xi_1 - \xi_0)$ with $\zeta = (1-\tau)\xi_0 + \tau\xi_1$, and since $\dot{\zeta} = \xi_1 - \xi_0$, $\dot{\xi} = \frac{1}{6}[\xi_1 - \xi_0, [\zeta, \xi_1 - \xi_0]] + O(\epsilon^4)$, of degree three in the endpoints, whereas $\Lambda_0(\xi)$ for a quadratic Λ_0 is $\Lambda_0(\xi_1 - \xi_0) + O(\epsilon^3)$, of degree two. Matching the two forces $\Lambda_0 \equiv 0$, and then $\dot{\xi} \neq 0$ at degree three for generic pairs. There is therefore no Λ_0 to subtract for equation 3, and the in-place correction does not reach it.

The route that remains for such a construction is to replace the base with one the hypothesis admits and then correct. That is always available and its cost is decided by the criterion rather than by the correction. Where the released construction is already an Euler–Poincaré solution of the metric in force, replacing it discards that optimality and the correction does not recover it. Where it is not, replacing it costs nothing. §5 measures both cases.

The second level does not extend this way. About the factorized base the level-two generator is not N_2 evaluated at the difference of generators. That base solves $\dot{\xi} = \Lambda_{\mathrm{iso}}(\xi)$ rather than $\dot{\xi} = 0$, so the third-order term in the expansion above carries cross terms between Λ_{iso} and $\Lambda - \Lambda_{\mathrm{iso}}$, and N_2 was derived under $\dot{\xi}_{\mathrm{base}} = 0$. We do not claim a level-two formula for a general base, and the second level is stated for group interpolation only.

C.12 Operation counts on $\mathrm{SE}(3)$

Section 3.3 states the cost as a fixed number of Lie-algebra operations. On $\mathrm{SE}(3)$, with $\Lambda(x) = M^{-1}\mathrm{ad}^*_x(Mx)$ and a diagonal or block-diagonal M, one evaluation of Λ costs two 6×6 applications of M and M^{-1}, one ad^*, which is two cross products, and one addition. Level one needs one such evaluation and one $\exp$, the latter in closed form through Rodrigues and the translational Jacobian. Level two needs, in addition, one $D\Lambda(u)[\Lambda(u)]$, which is two further ad^* applications and two metric applications, and one bracket $[u, \Lambda(u)]$. Both generators are then summed into a single F_ℓ, so the number of exponentials does not grow with the level. The field equation 12 adds one Ad and one dexp, both closed form on $\mathrm{SE}(3)$. Algorithm 1 states the procedure.

The count is constant in the number of integration steps, which is the comparison that matters against a shooting solve. It is not constant in the dimension of the group: each metric application is quadratic in $\dim \mathfrak{g}$ for a dense M, and ad^* is cubic in general, though on $\mathrm{SE}(3)$ both reduce to cross products.

Generality. The formulas of §3.3 are algebraic in brackets, $\exp$, $\log$, Ad and dexp, so they hold on any Lie group as local statements. Whether they are also *closed form* in the computational sense depends on the group: the solver-free implementation requires those five operations to be available in closed form, which they are on $\mathrm{SE}(3)$ and are not on an arbitrary group presented abstractly. We distinguish the two throughout.

C.13 WHERE THE STATEMENTS HOLD

The results of §3 hold for endpoint pairs away from the cut locus, at amplitudes that, under a Haar or anchored source, are fixed by the source distribution rather than by the data. Both conditions are quantified here, and the choices this work does not address are recorded at the end.

The cut locus. The cut locus of a point p is the closure of the set of points joined to p by two or more minimising geodesics, and the injectivity radius is the distance to it. That radius is the minimum of two quantities. The first is the distance to the first conjugate point, which is curvature-driven. The second is half the length of the shortest closed geodesic, which is not.

Three conditions bound the analysis, and one differs in kind. The Baker–Campbell–Hausdorff series equation 20 converges only within a radius, and the logarithm is single-valued only within the injectivity radius. Both are matters of accuracy, and both improve when more terms are added. The cut locus is not of this kind. There the minimising geodesic is not unique, so no expansion of any order has a single object to approximate.

Proposition 19 (Where the exposure lives, and why the cover does not remove it)**.** *(i) Integrate the conditional path equation 2 in N uniform Euler steps, so that $\theta(\tau) = (1-\tau)\theta_{01}$ and one step advances by θ_{01}/N. The cut locus, at distance π, lies within one step's reach of the point at parameter τ only if*

$$\theta_{01} > \frac{\pi}{1/N + 1 - \tau}, \tag{63}$$

and the right-hand side is at least π at every grid point $\tau \geq 1/N$, with equality at $\tau = 1/N$ and strict inequality beyond it. Since $\theta_{01} \leq \pi$, the condition fails at every such point and the exposure is exactly zero after the first step. (ii) Let $S^n \to \mathbb{RP}^n$ be the antipodal double cover with the round metric of radius one, of which the case $n = 3$ is $S^3 \to \mathrm{SO}(3)$, and let $d_\pm$ be the distances from a fixed lift to the two lifts $q, -q$ of a point. Then

$$d_+ + d_- = \pi, \qquad \theta_M = 2\min(d_+, d_-), \tag{64}$$

so the cut locus of the base is exactly the set on which $d_+ = d_-$.

Part (i) is the displayed inequality. Part (ii) is $\arccos(x) + \arccos(-x) = \pi$ together with the definition of the quotient distance. Part (i) uses $\theta(\tau) = (1-\tau)\theta_{01}$, which holds for the conditional path equation 2 and not for every curve between the same endpoints. The corrected path of §3.3 is one such other curve, and it stays as far from the cut locus as the uncorrected one. The statement concerns the conditional path, which is the object a training target is computed on and the idealised trajectory of sampling. The trajectory of a learned marginal field is not constrained to obey $\theta(\tau) = (1-\tau)\theta_{01}$. The two lifts are equidistant precisely on the cut locus, so the choice between them is undetermined there: passing to the cover relocates the discontinuity from the geometry into the sign of the data lift, for which there is no continuous global section.

In the homotopy of Appendix C.10 the first conjugate point is a singularity of the shooting map, hence a turning point of the continuation that carries the uncorrected path to the geodesic, and the distance to it bounds from above how far any predictor-only expansion can be trusted.

The amplitude follows from the source.

Proposition 20 (Source law and operating amplitude)**.** *Let the source rotation R_0 be drawn from a distribution on $\mathrm{SO}(3)$ invariant under left multiplication, and let R_1 be drawn from any data distribution independently. Then $R_0^\top R_1$ is Haar-distributed, the density of $\theta = \|\log(R_0^\top R_1)\|$ is*

$$p(\theta) = \frac{1 - \cos\theta}{\pi}, \qquad \theta \in [0, \pi], \tag{65}$$

and its median solves $m - \sin m = \pi/2$, giving $m = 2.3099$ radians or $132.35°$. Under an anchored source $R_0 = R_1 Z$ with $Z \sim \mathrm{IGSO3}(\sigma)$, the relative rotation is Z^{-1} and its median is that of the $\mathrm{IGSO3}(\sigma)$ angle density, below the Haar value by an amount fixed by σ. This is a coupling rather than an independent draw: an $\mathrm{IGSO3}(\sigma)$ source drawn independently of the data does not make the relative rotation $\mathrm{IGSO3}(\sigma)$-distributed.

Proof. Haar measure on the compact group $\mathrm{SO}(3)$ is bi-invariant and invariant under inversion, so for R_0 Haar and R_1 fixed, $R_0^\top R_1$ is Haar. Conditioning on R_1 and averaging over the data distribution leaves the law unchanged, so the relative element carries no information about the data.

The density equation 65 is the standard angle marginal of Haar measure on $\mathrm{SO}(3)$: writing a rotation in axis–angle form, the measure factors as the uniform measure on the axis times $(1-\cos\theta)\,d\theta/\pi$. Integrating, $\int_0^m p = (m-\sin m)/\pi$, and setting this to $\frac{1}{2}$ gives the stated equation, whose root is 2.30988.

For $\mathrm{IGSO3}(\sigma)$ the angle density is $p_\sigma(\theta) = \frac{1-\cos\theta}{\pi}\sum_{\ell\geq 0}(2\ell+1)e^{-\ell(\ell+1)\sigma^2/2}\,\frac{\sin((\ell+\frac{1}{2})\theta)}{\sin(\theta/2)}$, which reduces to equation 65 as $\sigma\to\infty$ and concentrates as $\sigma\to 0$. Its median is therefore a decreasing function of σ^{-1} and lies below the Haar value for every finite σ. □

The amplitude is therefore fixed by the source distribution and carries no information about the data. Under a uniform source the relative rotation is Haar-distributed whatever the target is, so two domains sharing a source law operate at the same amplitude however different their data. §5.3 compares five configurations against these values.

The deviation is a function of θ, the relative magnitude between the endpoints, which invites reading θ off the data. On a compact group that fails, since $g_0^{-1}g_1$ is Haar distributed whatever the law of g_1 when g_0 is Haar and independent of it. A model with a uniform source therefore operates at the same amplitude in every domain, so calling a domain small- or large-amplitude describes its data-to-data structure rather than the path its model traverses. The source is thus a design variable, setting the operating amplitude and with it whether path geometry matters at all.

Choices not addressed here. A conditional flow matching model is specified by more than its path. The source distribution, the coupling between source and data samples, the metric the loss induces, and the curvature of the marginal field are separate choices. §3 fixes the first two, takes the third as given by the released loss, and addresses none of the fourth. Two limits follow from taking the metric and the endpoints as given: the classification identifies the optimal construction under a metric without bearing on which metric a loss ought to induce, and it constrains the deviation at a given amplitude without determining that amplitude.

C.14 Composition of a sequence of increments

A policy or a kinematic chain emits a sequence of increments rather than a single one, and errors compose. This appendix derives how, on any Lie group, and is used in the sequence-level analysis of §5.4.

The endpoint sensitivity. Let $T = \prod_{i=1}^{H} \exp(\xi_i)$ and perturb on the right, $\exp(\xi_i) \mapsto \exp(\xi_i)\exp(\epsilon_i)$. Writing $P_i = \prod_{j\le i}\exp(\xi_j)$ and using $\exp(\xi)\exp(\epsilon) = \exp(\mathrm{Ad}_{\exp(\xi)}\epsilon)\exp(\xi)$ repeatedly to move each perturbation to the left, the perturbed product is, to first order,

$$T' = \exp\Big(\underbrace{\sum_i \mathrm{Ad}(P_i)\epsilon_i}_{E_H}\Big)\, T + O(\|\epsilon\|^2), \tag{66}$$

so that $\log(T'T^{-1}) = \sum_i \mathrm{Ad}(P_i)\epsilon_i$: the error *in the frame of the origin*. Moving it to the frame of the endpoint conjugates by T, and since $T = P_i B_i$ with $B_i = \prod_{j>i}\exp(\xi_j)$,

$$\log(T^{-1}T') = \mathrm{Ad}(T^{-1})\sum_i \mathrm{Ad}(P_i)\epsilon_i = \sum_i \mathrm{Ad}(B_i^{-1})\epsilon_i. \tag{67}$$

The two forms are the same quantity in two frames and are not interchangeable. Substituting one for the other gives a relative error of order unity. We verify both against ground truth obtained by composing the perturbed product numerically and taking the appropriate logarithm: over 40 draws at each of $H \in \{3, 6, 10\}$ the maximum relative deviation is 5×10^{-7}, consistent with the $O(\|\epsilon\|^2)$ neglected term at the perturbation size used.

The component split. With $P_i = (R_i, p_i)$ and $\epsilon_i = (\epsilon_i^\omega, \epsilon_i^v)$, the adjoint of $\mathrm{SE}(3)$ is

$$\mathrm{Ad}(R, p) = \begin{pmatrix} R & 0 \\ \hat{p}R & R \end{pmatrix}, \tag{68}$$

so the translational component of equation 66 is

$$E_H^v = \sum_i \big(\hat{p}_i R_i \epsilon_i^\omega + R_i \epsilon_i^v\big) = \underbrace{\sum_i R_i \epsilon_i^v}_{\text{direct}} + \underbrace{\sum_i p_i \times (R_i \epsilon_i^\omega)}_{\text{lever arm}}, \tag{69}$$

verified against ground truth to 6.7×10^{-7}. The first sum is the *direct* term and the second the *lever-arm* term.

The scaling exponent. Rotations preserve norms, so if the ϵ_i are independent with common scale then $\|\sum_i R_i \epsilon_i^v\|$ is the displacement of an H-step random walk and grows as $H^{1/2}$, while $\|\sum_i p_i \times (R_i \epsilon_i^\omega)\|$ is the same walk with step scale $\|p_i\|$. If $\|p_i\| \sim i^\alpha$ then $\sqrt{\sum_i \|p_i\|^2} \sim H^{\alpha+1/2}$, so that

$$\text{direct} \sim H^{1/2}, \qquad \text{lever} \sim H^{\alpha+1/2}, \qquad \Phi \;:=\; \frac{\text{lever}}{\text{direct}} \;\sim\; H^{\alpha}. \tag{70}$$

The exponent of Φ is therefore α, a property of the trajectory, $\frac{1}{2}$ for isotropic increments and approaching 1 for directionally persistent ones. **Any reference value for this exponent must therefore be taken from the same data whose exponent is being tested.**

D EXPERIMENTAL DETAIL

D.1 SCOPE AND INCLUSION

We include every implementation we had identified by 22 September 2026 that trains a simulation-free generative model on $\mathrm{SE}(3)$ or one of its factors, distributes trained weights, and distributes the training code that produced them. The third condition is what limits the list: the induced metric cannot be read from inference code alone. Five implementations satisfy all three and present six configurations between them.

Table 13: **The reading behind Table 1.** Rows are in the same order and the metric column uses the same terms.

implementation	source	the loss, and the metric it induces
EquiGraspFlow	`bdlim99/EquiGraspFlow`	unweighted mean-squared error, giving $M = cI$ exactly
DiffDock	`gcorso/DiffDock`	score-based diffusion, read as the metric its loss weights induce: three components under separate exponential schedules, $M = \mathrm{diag}(\sigma_{\mathrm{tr}}^{-2} I, \sigma_{\mathrm{rot}}^{-2} I, \sigma_{\mathrm{tor}}^{-2} I)$, each block a multiple of the identity
FrameFlow, $t < \frac{1}{4}$	`microsoft/protein-frame-flow`	rotational loss in a logarithm chart anchored at the current point, $J = \mathcal{J}_r(\phi)^{-\top} \mathcal{J}_r(\phi)^{-1}$ with $\phi = (1-\tau)\omega$, frozen at τ and aligned with ω at every anchor (Proposition 4)
FrameFlow, $t > \frac{1}{4}$	as above	auxiliary atom term enabled, with off-diagonal $+[\sum_a b_a]_\times$ coupling the blocks
SmolVLA	`huggingface/lerobot`	unweighted mean-squared error in one fixed chart of normalised increments, inducing the flat pullback $B(g)^\top W B(g)$ with $W = \mathrm{diag}(1/(\sigma_i s_i)^2)$, whose rotational eigenvalue ratio is 3.947 (Lemma 16)
3D FlowMatch	`nickgkan/3d_flowmatch_actor`	weighted sum of L_1 terms, whose Hessian vanishes almost everywhere, so no quadratic metric is induced

Where a repository ships more than one configuration the reading follows the one distributed with the released weights, which matters for three rows. FrameFlow's repository default sets the auxiliary atom loss to weight zero, under which the coupled metric never takes effect, while the configuration shipped with the released checkpoint sets it to one with the gate at $t = \frac{1}{4}$. These are the SCOPE weights of Zenodo record 12776473, which the protein rows use. The PDB weights in the same release put the gate at $t = \frac{1}{2}$. The two give opposite verdicts and we read the checkpoint, since the released model is the trained artefact and its configuration records how it was trained. Our own FrameFlow runs were trained under the repository default, so the two FrameFlow rows describe the released checkpoint's metric rather than the metric of the arms we trained. DiffDock's argparse defaults differ from the configuration shipped with the released weights, and we read the latter. 3D FlowMatch selects among three samplers by a configuration flag and we read its default. SmolVLA's verdict follows from Lemma 16 rather than from a measurement. On the pairs of its ladder in Table 2, the released path sits at 0.09 to 0.18 of the floor under the pullback at every rung, and at 2×10^{10} to 5×10^{12} times the floor under the left-invariant freeze.

DiffDock's block ratio is not constant. The three schedules are exponential in the diffusion time, $\sigma_c(t) = \sigma_{c,\min}^{1-t} \sigma_{c,\max}^{t}$, running from 0.1 to 19.0 Å in translation and from 0.03 to 1.55 rad in rotation. The ratio $\sigma_{\mathrm{tr}}/\sigma_{\mathrm{rot}}$ is therefore 3.33 at $t = 0$ and 12.26 at $t = 1$. The criterion is unaffected at every t, since each block remains a multiple of the identity whatever the ratio between them.

FrameFlow's coupling is a frame convention. The auxiliary term compares backbone atoms placed as $x_a = Rb_a + t$ over the idealised template $b = \{(-0.525, 1.363, 0), (0, 0, 0), (1.526, 0, 0)\}$ Å. Expanding $\sum_a \|\omega \times b_a + v\|^2$ to second order in a body perturbation gives the Gram operator

$$\begin{bmatrix} J & +[\sum_a b_a]_\times \\ +[\sum_a b_a]_\times^\top & 3I \end{bmatrix}, \qquad \sum_a b_a = (1.001,\ 1.363,\ 0)\ \text{Å},$$

so the off-diagonal is the hat map of the summed template, three times its centroid. It is a fixed structural constant, and the diffusion time enters only through an overall scale that grows as $t \to 1$. Centring the template so that $\sum_a b_a = 0$ removes the coupling exactly, which we confirm numerically at 3.9×10^{-18}, and leaves the rotational block unchanged. On 60 real pairs at $\tau = \frac{1}{4}$ it

lowers the gap by a factor of 26 under the template-coupled metric and 32 under the released chart metric frozen at τ. Centring also shifts the frame origin, $p \mapsto p + Rc$ with c the centroid, so the factorized interpolation taken in the centred coordinates is a different path from the released one rather than a relabelling of it.

D.2 The choice of path parameter

Deviations in this paper are reported at $\tau = \frac{1}{4}$, which is our choice and not any implementation's. What an implementation trains against is the deviation averaged over its own distribution for τ, and since the deviation carries a $\tau(1-\tau)$ profile that average depends on where the distribution puts its mass. The relevant quantity is the overlap $\mathbb{E}_\tau[\tau(1-\tau)]/\max_\tau \tau(1-\tau)$, which is $2/3$ for a uniform draw and 0.75 at $\tau = \frac{1}{4}$.

Table 14: The distribution each implementation draws τ from, and its overlap with the $\tau(1-\tau)$ profile.

implementation	distribution	overlap
3D FlowMatch	$\mathrm{sigmoid}(\mathcal{N}(0, 1.5))$	0.707
SmolVLA	$\mathrm{Beta}(1.5, 1)$, clipped at 10^{-3}	0.686
FrameFlow	$\mathcal{U}(10^{-2},\ 1 - 10^{-2})$	0.680
DiffDock	$\mathcal{U}(0, 1)$ at its defaults	0.667
EquiGraspFlow	$\mathcal{U}(0, 1)$	0.667

The five span 0.667 to 0.707, so none samples away from where the deviation lives and no row of Table 1 needs rescaling. Both departures from uniform go toward the middle rather than away from it, since SmolVLA's tilt and FrameFlow's trimming of the two endpoints each raise the overlap slightly. Read against an implementation's own distribution, a deviation reported at $\tau = \frac{1}{4}$ is 0.89 to 0.94 of its stated value.

D.3 Block weights do not change the optimal path

A metric that weights the rotational and translational blocks differently, $M = \mathrm{diag}(\alpha I_3, \beta I_3)$, gives the same generator as the isotropic one. Writing $\xi = (\omega, v)$ and expanding $\Lambda(\xi, M) = M^{-1}\mathrm{ad}^*_\xi(M\xi)$ on $\mathrm{SE}(3)$, the translational component is $-\beta(\omega \times v)/\beta$ and the rotational component vanishes, so both block weights cancel and $\Lambda(\xi, M) = (0, -\omega \times v)$ for every $\alpha, \beta > 0$.

Two checks support this. Across $(\alpha, \beta) \in \{(1,1), (1,7.3), (12.9,1), (0.03,41)\}$ and four seeds, the generator agrees with its reference value to at most 2.7×10^{-15} and the shot geodesics coincide as curves to at most 3.1×10^{-15}. And in the docking domain, whose implementation draws translation, rotation and torsion with three separate noise scales, the measurements under the induced block-scaled metric agree with the isotropic ones to within 3×10^{-13} relative at all three seeds and all three values of τ.

The consequence for practice is that the ratio of rotational to translational weight, whatever it is, leaves the optimal path and the correction unchanged. It enters only through the norm in which a deviation is reported. Anisotropy *within* a block is a different matter, and that is what a per-dimension normalised loss induces. Under $M = \mathrm{diag}(J, I)$ with $J = \mathrm{diag}(1, 2, 4)$, over 2000 pairs per amplitude at two seeds, the measured residuals are

ϵ	uncorrected	one level	two levels
0.20	7.160×10^{-3}	2.648×10^{-4}	3.179×10^{-5}
0.10	1.803×10^{-3}	3.253×10^{-5}	1.972×10^{-6}
0.05	4.506×10^{-4}	4.071×10^{-6}	1.224×10^{-7}

with local slopes $1.99, 2.00$ uncorrected, $3.02, 3.00$ after one level, and $4.01, 4.01$ after two. These reproduce an earlier run at five endpoint pairs per amplitude to within a factor of 0.99 to 1.12, the difference being sampling.

D.4 Alignment and anisotropy varied independently

Four released configurations satisfy the criterion and one violates it, which is five points. The claim that alignment rather than anisotropy separates them is stronger than five points can carry, so we measure it on a released metric with its eigenbasis rotated, and on a family constructed to vary the two independently.

A released metric, rotated. FrameFlow's chart metric below its gate, frozen at $\tau = \frac{1}{4}$, has rotational eigenvalue ratio 1.22 and the endpoint axis as an eigenvector. On $30\,000$ real protein pairs over three seeds its gap has median 1.8×10^{-13} Å against a floor of 5 to 10×10^{-11} Å. Replacing J by $QJQ^\top$ with $Q = \exp(45^\circ \cdot (1,1,1)/\sqrt{3})$, which keeps every eigenvalue, raises the median gap on the same pairs to 3.33, 3.25 and 3.30×10^{-2} Å across the three seeds and the normalised misalignment $\|\omega \times J\omega\|/(\|J\|\|\omega\|^2)$ from 10^{-16} to 0.078.

The family. Take $J(r,\phi) = R_x(\phi)\,\mathrm{diag}(1,\sqrt{r},r)\,R_x(\phi)^\top$ with the endpoint axis fixed to e_3, over seven eigenvalue ratios $r \in \{1, 1.2, 1.5, 2, 3, 5, 10\}$ and seven eigenbasis rotations $\phi \in \{0, 5, 15, 30, 45, 60, 90\}$ degrees. The eigenvalues are $\{1, \sqrt{r}, r\}$ rather than $\{1, 1, r\}$, which keeps $\lambda_{\max}/\lambda_{\min} = r$ exactly while leaving the geometry generic. The sweep is aligned at both ends: the endpoint axis is an eigenvector at $\phi = 0^\circ$ and again at $\phi = 90^\circ$, where its eigenvalue is $\sqrt{r}$, and the misalignment $|r - \sqrt{r}|\,|\sin\phi\cos\phi|$ is maximal at $\phi = 45^\circ$. Pairs are synthetic, $u = (\theta e_3,\, t)$ with $\theta = 0.4$ and $\|t\| = 1$ in the dimensionless units of $M = \mathrm{diag}(J, I)$, $2\,000$ per cell at $\tau = \frac{1}{4}$, three seeds, gap and floor measured per cell. The gap is the median six-vector norm of the deviation, unnormalised. At $r = 10$ it measures below the floor at $\phi = 0^\circ$ and 5.2×10^{-2} at $\phi = 45^\circ$, and over the forty-nine cells its rank correlation is 0.93 with the misalignment and 0.54 with the ratio r, which are the figures §5.2 cites.

Three controls, and they are not the same kind of zero. The isotropic row, $r = 1$, satisfies the criterion by algebra at every ϕ. With $J = cI$ the misalignment is $\|\omega \times c\omega\|/(c\|\omega\|^2) = 0$ identically, for every endpoint pair and every seed, and it measures below the floor throughout, at $\mathrm{gap/floor} = 0.0711$ constant to four figures. The aligned column, $\phi = 0$, satisfies it for the pairs drawn, since e_3 is an eigenvector of $\mathrm{diag}(1,\sqrt{r},r)$ at every r, and it measures below the floor throughout as well. The column $\phi = 90^\circ$ is a control of the same kind: the endpoint axis is an eigenvector there too, with eigenvalue $\sqrt{r}$, and its gap measures at 3.33×10^{-14} against 3.33×10^{-14} at $\phi = 0^\circ$, indistinguishable across all three seeds. The grid therefore carries nineteen aligned cells, seven at $r = 1$ and six at each of $\phi = 0^\circ$ and $\phi = 90^\circ$. The distinction matters beyond this grid: an exactly isotropic metric satisfies the criterion for every endpoint pair by construction, while a metric whose misalignment happens to measure at 10^{-17} satisfies it for the pairs drawn and a different draw could differ. Of the released configurations, EquiGraspFlow and DiffDock are of the first kind.

D.5 The ladders, and what the apparatus can resolve

Protocol. For each rung, g_0 is taken from the implementation's own endpoint draw in the top panel of Table 2, its source distribution for DiffDock and FrameFlow and a data grasp for EquiGraspFlow, and is a random pose in the bottom panel, $u = \epsilon\hat{u}$ with $\hat{u}$ a uniformly random unit direction in $\mathfrak{se}(3)$, and $g_1 = g_0 \exp(u)$, so that ϵ is the $\mathfrak{se}(3)$ norm of u. The bottom panel's arms are algebra interpolation and group interpolation corrected once and twice. Setting the twist rather than the two endpoint vectors matters: scaling the endpoints lets the relative rotation reach twice the intended amplitude, crossing π at half the nominal value and leaving the boundary-value problem without a unique solution. The geodesic is obtained by shooting and the residual reported is $\|\log(g_{\mathrm{riem}}^{-1} g_{\mathrm{arm}})\|/\|u\|$ at $\tau = \frac{1}{4}$, so a residual of order k appears with slope $k - 1$. Three hundred pairs per rung, seed 42, thirty-five cells, no non-converged solves, maximum shooting residual below 4.2×10^{-14}.

What the apparatus resolves. The floor is measured on a case whose answer is known in advance. On $\mathrm{SO}(3)$ with $M = I$ the generator vanishes identically, the geodesic is the one-parameter subgroup, and the shot solution reproduces it to 5.2×10^{-14}. That figure calibrates the integrator but overstates what the apparatus resolves on $\mathrm{SE}(3)$, where the generator $(0, -\omega \times v)$ does not vanish and the shooting has work to do. Under $M = cI$ on $\mathrm{SE}(3)$ the answer is again algebraically zero, and the solver returns 5.7×10^{-12}, drifting by a factor of 1.03 from 0.72° to 45.84°. That is the resolution quoted throughout, and it is the shooting solver's own convergence rather than a geometric quantity.

Why the flat rows report a residual and not a ratio. A natural presentation would divide each residual by its own floor. We do not, because that floor is the integrator's self-reproducibility, the same shot integrated at two tolerances, and it measures something other than the accuracy of the shooting solve. On EquiGraspFlow the residual drifts by a factor of 1.03 across the ladder while the floor moves by 205, so the ratio spans 10.6 to 2106 and does so non-monotonically, rising to its maximum at $\epsilon = 0.1$ and falling by two orders after. Nothing geometric behaves that way. The residual is the stable quantity and it is directly comparable to the violating rows.

The floor's own jump between two adjacent rungs is a crossover between two sources of error rather than a geometric event. Below $\epsilon \approx 0.3$ the self-reproducibility is dominated by rounding, which is flat in amplitude, and above it by truncation, which grows. The two are equal near that amplitude, and the measured floor tracks whichever is larger.

The ratio between consecutive levels. Under the template-coupled metric the product $(C_1/C_2)\,\epsilon$ is 6.803, 6.801, 6.797, 6.789, 6.765, 6.765 and 6.800 across a thirty-two-fold range, a drift of 0.56 per cent, so $C_1/C_2 = 6.79/\epsilon$. This tests the constant between levels, which the slopes on their own do not. Under SmolVLA's frozen weights the same product drifts 13 per cent, consistent with the 10 per cent its slopes of 2.023 and 3.001 predict over its wider range, and the agreement between predicted and observed drift is itself a check.

Agreement with the amplitude ladders. The ladder of Appendix C.9 gives local slopes converging to 2.01, 3.01 and 4.00 at $\tau = \frac{1}{4}$ over five pairs. Converting the fitted slopes of Table 2 by the $k - 1$ relation above gives residual orders of 2.000, 3.023 and 4.001 here. The two methods agree, and the present one is the stronger because its reference is a solved geodesic rather than another truncation.

D.6 Wall-clock cost

On $2\,000$ synthetic pairs with $M = I$ at $\epsilon = 1.2$, one closed-form level is $38.0\,\mu$s per pair against 98.9 ms for a boundary-value solve, a factor of 2601 at the median. The ratio of the ninety-fifth percentile to the median is 1.09 for the algebraic levels and 1.79 for the solver, so the cost of a level is nearly independent of the endpoints.

The solve is not. Measured on each domain's own pairs, under the frozen metric for the action head and the template-coupled metric for protein, it takes 21.85 ms at a median rotation of $3.38°$ and 158.13 ms at $115.96°$, a 7.24-fold rise across a 34.3-fold rise in amplitude, an exponent of 0.56. On forty cores that is 0.22 s per training step in the first case against a 1.08 s step, and 1.58 s in the second. The sublinear scaling is what makes the exact route usable in both regimes rather than only the cheap one.

D.7 Composition, and why the protocol matters

Appendix C.14 derives how a per-increment deviation composes. This appendix reports what it composes to.

Two protocols. A population measurement draws increments independently. A policy emits fifty consecutive increments from one trajectory. The two are different populations and they do not agree, so every figure in the main text described as a chunk uses the consecutive one.

Per increment we take $D_k = \log\big(g_{\mathrm{riem}}(\tau)^{-1} g_{\mathrm{path}}(\tau)\big)$ at $\tau = \frac{1}{4}$, with g_{riem} the numerically solved geodesic of the frozen metric and g_{path} the released algebra path or its corrections built on the group base, compose $\prod_k \exp(\xi_k)$ against $\prod_k \exp(\xi_k)\exp(D_k)$, and report the positional offset of the two endpoints. Note that τ indexes the flow-matching time within the generation of one action while k indexes actions along the chunk. They are different axes.

The tolerance the deviation is read against. The LIBERO-SPATIAL success predicate is (On bowl plate), which LIBERO implements as contact and vertical ordering together with a lateral centre-distance test, $\|(p_{\mathrm{bowl}} - p_{\mathrm{plate}})_{xy}\| < 0.03\,\mathrm{m}$ (`libero/envs/object_states/base_object_states.py:78--94`). All ten tasks in the suite carry the same goal, so the 30 mm is one number for every evaluation reported here, and it is a constant of the task rather than of any model, so ratios formed against it are comparable across the four arms. Two things it is not. It is a tolerance on the placed object, not on the end-effector, so comparing the composed end-effector deviation against it assumes the deviation is carried intact into the release position. And it bounds a lateral offset while the composed deviation is a three-dimensional norm, which bounds its own lateral part, so the comparison is conservative in that direction. Rotation does not enter the predicate, so the rotational deviation is reported without a comparable tolerance. The composed displacement, 4.481 mm averaged over three noise realisations, is 0.149 of the tolerance.

arm	random	consecutive	ratio
released	3.96×10^{-3}	6.91×10^{-3}	1.74
one level	4.55×10^{-5}	9.30×10^{-5}	2.04
two levels	5.94×10^{-6}	5.36×10^{-6}	0.90

Thirty chunks per protocol, $H = 50$, seed 42, identical per-chunk noise seeds, the two differing only in how the fifty increments are chosen. Values are the mixed $\mathfrak{se}(3)$ norm, and §D.8 gives the split into displacement and rotation.

Coherence. Define $\rho_c = \| \sum_k D_k \| \, / \, \sum_k \| D_k \|$, which is one when the fifty deviations point in a common direction and small when they cancel. It predicts the composed value within each arm:

arm	ρ_c consecutive	ρ_c random	ratio	Spearman, ρ_c vs endpoint
released	0.276	0.189	1.46	+0.68
one level	0.364	0.239	1.53	+0.41
two levels	0.303	0.299	1.01	+0.85

The coherence gain and the endpoint gain move together arm by arm. Where consecutive increments raise ρ_c they raise the composed value, and where they do not, they do not, so the twice-corrected arm has no ratio left to explain. The correlations are within an arm rather than between arms: the four values of ρ_c do not differ systematically, so the correction does not change how aligned the residuals are.

Rotational amplitude is a weaker predictor of the same quantity, at $+0.41$ to $+0.56$, and translational amplitude predicts nothing anywhere, every p above 0.33.

Exponents. Fitting the composed displacement against H over $\{5, 10, 20, 30, 50, 75, 100\}$, at 120 chunks per seed and three seeds, the 120 chunks within a seed sharing one noise realisation:

arm	consecutive	random	difference
released	1.01	0.72	+0.29
one level	1.07	0.73	+0.34
two levels	1.31	0.98	+0.33

The protocol shift is 0.29 to 0.34 on every arm, so both the sampling scheme and the correction level move the fitted exponent. Independent accumulation gives $\frac{1}{2}$ and coherent gives 1, and the random protocol sits nearer the first while the consecutive protocol sits nearer the second. The exponents rise with the level of correction under both protocols, which we record without an explanation. The arm measured against the geodesic of the block-isotropic metric, for which the factorized path satisfies the criterion, is omitted throughout, since its composed values are at 10^{-16}, so its exponent and its coherence are the slope and the direction of floating-point residuals.

D.8 Displacement, not the mixed norm

Composition returns an element of $\mathrm{SE}(3)$, and the quantity we report is the positional offset of its two endpoints. The norm of the full $\mathfrak{se}(3)$ logarithm adds a rotation in radians to a translation in metres and is not commensurable with a tolerance in millimetres. On the released arm at $H = 50$ the split is 4.5 mm of displacement against 0.21° of rotation, over three seeds giving 4.02, 4.16 and 5.27 mm.

The translation component of the logarithm and the positional offset differ by at most $\|V(\omega) - I\| \approx \|\omega\|/2$ in relative terms, about 2×10^{-3} at the composed rotation of 4×10^{-3} rad, since the left Jacobian is close to the identity there. The distinction matters in principle and not at this amplitude.

D.9 The four policies, and what the null rests on

Training. Four policies on the action head, identical but for the conditional path: the released algebra construction, one and two levels of the closed-form correction under the left-invariant freeze of the metric its loss induces, and the exact geodesic of that frozen metric solved by shooting at every training step. Twenty-five thousand steps each, chosen because two earlier policies were within 1.4 per cent of their final error by that point and their convergence curves agreed to four decimals.

Evaluation. One thousand paired chunks from 166 episodes of the training set, **with the flow-matching noise drawn once per chunk and handed to every policy**. Leaving each to draw its own inflates the paired standard deviation thirtyfold and produces apparent effects of several per cent that vanish under control.

What the comparison controls. The alternative is the geodesic of the left-invariant freeze of the loss's metric rather than of the identity, so it differs from the released path in the rotational block, where that freeze is misaligned. The reference is solved rather than expanded, so what is compared is the released path against an exact geodesic. The sampling noise is common across the four, so the comparison is not dominated by a term belonging to none of them. The intervals of Table 3 resample episodes with all their chunks; treating chunks as independent narrows the rotational intervals by a quarter to a third and leaves every interval covering zero. The minimum detectable effects of 0.10 to 0.26 per cent are the design's resolution, not a bound on the effect.

Scope. None of the four was trained with a held-out split, so the evaluation episodes are drawn from the training set. The effect of that overlap on the comparison is not determined by this design, and the result is reported as an in-sample comparison.

The weights. Two networks trained on identical targets differ because the optimiser is stochastic, so we train one additional policy at a different seed and report every contrast as a fraction of it (Table 15). No layer of the five hundred carries a target contrast exceeding its own seed value, and on the output projection the rotational rows show the smallest difference of the three, which is the block where the frozen metric is misaligned.

Table 15: Each contrast is $\|\Delta W\|$ normalised by the tensor's own norm. Left, the p90 and the maximum of the target contrast over all five hundred tensors, each divided by the same statistic of the seed-only control. Right, the four row groups of the action head's output projection, each divided by its seed-only value.

	all tensors		output rows			
contrast	p90	max	trans.	rot.	gripper	padding
l_0 vs l_1	0.463	0.400	0.214	0.189	0.196	0.518
l_0 vs l_4	0.458	0.396	0.207	0.187	0.199	0.513

The seed band is estimated from one pair, so its own spread is not characterised. The target contrasts sit at a fifth of it and the reading survives the band being wrong by a factor of two either way. The two contrasts are also not measured against the same reference, since both policies in a target contrast share a seed while the band does not. The padding rows are informative here: their targets are identical across the four, since the correction acts only on the six algebra dimensions, yet they differ by 0.52 of the band. Training at a fixed seed is therefore not bit-reproducible, and part of every target contrast is that irreducible variation. What the measurement supports is the comparative claim, that the choice of conditional path moves the parameters less than the choice of seed, and that claim does not need a reference the two share.

D.10 Two things the harness had to get right

Every evaluation before a repair to the state builder returned zero success on every task. `lerobot`'s environment wrapper builds `observation.state` from an `agent_pos` key while the LIBERO environment emits a nested `robot_state`, so the policy received no proprioceptive input and emitted near-constant actions. After the repair the published `smolvla_libero_spatial` checkpoint scores 38.5 per cent over 200 episodes, and our own policy trained from the same initialisation scores 35.5 per cent, within 0.62 standard errors of it. That agreement is what establishes that the policies compared in §D.9 were trained and evaluated correctly.

One convention that had to be resolved. The gripper points downward throughout a LIBERO episode, so the end-effector rotation sits at $179.97°$ for the whole trajectory, within 5.7×10^{-4} of π. There $w = \cos(\theta/2)$ vanishes and no rule keyed on its sign selects a branch. We require the dominant axis-angle component to be positive within 0.01 of π, which introduces $0.065°$ of error at the measured angle and matches the convention the training data uses. This is the branch locus of Proposition 19 appearing in a state encoding rather than in a path.

D.11 The correction is a training-time object

The correction is a closed-form function of the endpoints and can in principle be added to the sampling path of a model never trained with it. It cannot usefully be, and the reason is structural rather than an implementation detail.

What was measured. On the released checkpoint with unmodified weights, integrating the corrected path at sampling time in place of the released one degrades the policy, and the degradation is not a mismatch between training and sampling. Applying the corrected sampler to a policy trained with the corrected path degrades it by 14.265 per cent and applying it to one trained with the released path degrades it by 14.286, an interaction of -0.025 to $+0.020$ per cent with p from 0.34 to 0.88 over 300 chunks.

What that leaves. The conditional path is a training-time object. At training the paths may cross, at sampling they cannot, since the ODE solution is unique, and the sampler follows the marginal field rather than any conditional path. The correction is defined on the conditional path and is a function of both its endpoints, one of which is exactly what sampling is trying to produce. Estimating that endpoint and then using the conditional path is a known technique with known difficulties in marginal-conditioned bridge samplers, and the general remedy in that literature is ours: put the constraint in the training objective. Which is what the four policies of §D.9 test.