ASYMPTOTIC INCONSISTENCY OF INTERPOLATION IN TERMINAL-PROJECTED FLOWS FOR PDE INVERSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Zero-shot constrained generative models have been widely adopted for solving partial differential equation (PDE) inverse problems. A prominent class of methods, such as ECI and PCFM, enforces hard physical constraints through terminal projection coupled with reverse optimal-transport (OT) displacement interpolation. The validity of these methods relies on a generally accepted asymptotic consistency assumption of the flow interpolation, that the instantaneous predicted velocity asymptotically matches the interpolation secant velocity as the integration step size vanishes. However, whether the consistency remains valid after projection has not been verified. To address this, we establish a reference framework based on Minimal Step-wise Perturbation (MSP). We prove that the asymptotic consistency fails after terminal projection: the state discrepancy does not vanish under step refinement, inducing an extraneous surplus velocity field of order $\Theta(\Delta t^{-1})$. This unrecognized surplus velocity causes trajectory distortions and distribution deformation, which cannot be interpreted as principled posterior guidance. Across parameter sweeps on PDE benchmarks, we find that model performance exhibits strong task dependence, with substantial variations and even performance reversals across different task configurations. These results demonstrate that observed performance gains cannot be directly attributed to genuine posterior-aware optimization, but instead be influenced by the surplus dynamics.

1 INTRODUCTION

Flow Matching provides an ODE-based framework for generative modeling and scientific inference. In generative modeling, how to enforce hard constraints using pretrained model priors is particularly important for partial differential equation (PDE) systems. Recent zero-shot constrained samplers, including ECI (Cheng et al., 2025), PCFM (Utkarsh et al., 2026), and ProFlow (Yu et al., 2026), use the learned velocity field to extrapolate a terminal state from the current state, project it onto the constraint manifold, and then perform a pullback operation to construct the next-step state.

The validity of this terminal-projection and pullback construction relies on reverse optimal-transport (OT) displacement interpolation: as the integration step size Δt approaches zero, the instantaneous velocity predicted at the current state should asymptotically match the secant velocity represented by the interpolation. This relation is generally applied on the original unconstrained flow trajectory, where the interpolation is constructed from states lying on the same learned flow path. However, the corrected terminal state is usually not the terminal state of the original flow trajectory.

To answer whether this consistency remains valid after terminal projection, we construct a reference framework based on Minimal Step-wise Perturbation (MSP). The framework directly minimizes the perturbation of the immediate unconstrained next-state update subject to the terminal constraint, and crucially does not require the interpolation to satisfy the asymptotic consistency condition. We further recast this framework into the same extrapolation–projection–interpolation procedure and compare it with the original update. This reveals that, relative to the original models, MSP differs only in the projection center, which can isolate the state discrepancy introduced by the asymptotic consistency assumption. We prove that, under non-degenerate conditions, this discrepancy does not vanish under step refinement. Normalizing the discrepancy by the integration step yields a surplus velocity field with magnitude $\Theta(\Delta t^{-1})$. This demonstrates that the asymptotic consistency assumption becomes invalid after terminal projection.

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 further prove that the direction of this surplus velocity is determined by the historical evolution of the flow rather than the posterior target. The surplus acts as an uncontrolled intervention, and different task configurations can produce qualitatively opposite effects. This reveals a previously unrecognized false assumption in terminal-projection samplers, providing a unified mechanistic explanation for their erratic, task-dependent behavior and the distortion of distributions.

Contributions.

- **Audit of the State Interpolation Assumption.** We reveal that the optimal-transport state interpolation assumption breaks down after terminal projection: the secant velocity derived from the projected endpoints fails to asymptotically match the instantaneous generative velocity, thereby inducing a divergent surplus velocity field of order $\Theta(\Delta t^{-1})$.
- **Characterization of Surplus Dynamics.** We prove that this previously unrecognized surplus velocity severely distorts the generative trajectories and predictive distributions. Lacking intrinsic posterior directionality, its task-dependent nature compromises the interpretability of the constrained dynamics.
- **Plug-and-Play Surplus Elimination.** We propose the Minimal Step-wise Perturbation (MSP) framework, which replaces terminal perturbation with minimal disturbance to the next-step state. This formulation bypasses the invalid interpolation assumption while strictly preserving the extrapolation–projection–interpolation procedural structure by merely shifting the projection center, enabling its deployment as a seamless plug-and-play correction for existing samplers.

2 Related Works

Flow-Based Inverse Solvers. Inverse problems seek to recover unknown states, signals, or images from indirect, incomplete, or corrupted observations across scientific and engineering applications (Lailly, 1983; Huang et al., 2005; Moliner et al., 2023; Song et al., 2022; Quan et al., 2024). Flow-based generative models learn ODE-driven transports between source and target distributions (Lipman et al., 2022; Liu et al., 2022; Kerrigan et al., 2023), offering non-diffusive probability paths and flexible transport through optimal-transport interpolation (Albergo & Vanden-Eijnden, 2022; Albergo et al., 2025). Recent flow-based inverse solvers can be broadly categorized by how posterior information is incorporated into pretrained flows: trajectory-optimization methods modify the trajectory or source variable (Liu et al., 2023; Ben-Hamu et al., 2024); step-wise guidance and proximal methods inject local posterior corrections (Martin et al., 2025; Park & Ye, 2025; Kim et al., 2025); and trajectory-recoupling methods reshape transport toward measurement-consistent posterior trajectories (Zhang et al., 2026a). This line of work has also extended to scientific and physical inverse problems (Zhang et al., 2026b; Sherki et al., 2025; Leong et al., 2026).

Constraint-Aware Generation for PDEs. PDE inverse problems recover physical fields or unknown parameters from sparse or indirect observations and are commonly addressed with physics-constrained solvers and physics-informed learning, such as Neural Operators (Li et al., 2020; Lu et al., 2021) and PINNs (Raissi et al., 2019; Li et al., 2024). More recently, pretrained generative priors have been used for physics-constrained inference by incorporating PDE residuals or observational likelihoods during generation (Huang et al., 2024; Bastek et al., 2025). Existing approaches have also explored model fine-tuning (Tauberschmidt et al., 2026) and dual-energy formulations (Kurtz & Davydov, 2026; Sun et al., 2026), but these require task-specific adaptation. This has motivated *zero-shot* methods that enforce physical constraints directly at inference time while preserving the pretrained prior (Cheng et al., 2025; Utkarsh et al., 2026; Yu et al., 2026), followed by subsequent extensions (Cai et al., 2026; Kolli et al., 2026). Among them, ECI (Cheng et al., 2025) and PCFM (Utkarsh et al., 2026) use terminal projection with optimal-transport interpolation, while ProFlow (Yu et al., 2026) adopts proximal guidance. We focus on this zero-shot class and audit the dynamical consistency of its projection–interpolation construction.

Existing Observations. Recent studies have exposed related limitations in terminal-projection models: Xu et al. (2026) highlights that measure-zero PDE manifolds necessitate co-area corrections for posterior consistency, while Gao et al. (2026) identifies dynamical artifacts such as sharp trajectory kinks. Whereas these works focus on measure- or trajectory-level symptoms, our audit examines the core consistency assumption underlying the projection and OT interpolation.

3 BACKGROUND AND PROBLEM SETUP

PDE inverse problems and constraint manifolds. Let $\Omega \subset \mathbb{R}^d$ denote a bounded physical domain, and $u \in \mathcal{U} \subseteq \mathbb{R}^D$ an unknown physical state field. Physical laws, boundary conditions, and observational data are encoded by a continuously-differentiable constraint operator $h\colon \mathbb{R}^D \to \mathbb{R}^m$ with $m < D$. The admissible configuration space forms the corresponding zero-level set:

$$\mathcal{M} \coloneqq \left\{ u \in \mathbb{R}^D \mid h(u) = \mathbf{0} \right\}. \tag{1}$$

We assume h is C^2-smooth and its Jacobian $J(u) \coloneqq \nabla h(u)^\top$ has full row-rank m on $\mathcal{M}$, so that $\mathcal{M}$ is a closed, smooth embedded submanifold of dimension $D - m$.

Post-hoc terminal projection and the unified pullback update. To enforce the terminal physical constraint $h(u_1) = \mathbf{0}$ at sampling time without retraining v_θ, post-hoc methods first compute the unconstrained terminal extrapolation $\hat{u}_1(u_t, t) \coloneqq u_t + (1-t)v_\theta(u_t, t)$ and project it onto $\mathcal{M}$ via Euclidean-norm minimization:

$$u_1^{\text{corr}} \coloneqq \Pi_{\mathcal{M}}(\hat{u}_1(u_t, t)) = \arg\min_{u_1 \in \mathbb{R}^D} \frac{1}{2}\|u_1 - \hat{u}_1(u_t, t)\|_2^2 \quad \text{s.t.} \quad h(u_1) = \mathbf{0}. \tag{2}$$

To propagate the trajectory from time t to $t + \Delta t$, existing methods typically pull this corrected terminal state back via reverse optimal-transport (OT) displacement interpolation, bypassing costly backward integration. The operational and physical validity of this shortcut relies on a foundational premise: the average secant velocity of the pullback interpolation must asymptotically match the instantaneous predicted velocity at the corresponding state as the step size vanishes. To formally audit this prerequisite across post-hoc samplers, we explicitly extract and formulate this requirement as follows:

Definition 1 (Asymptotic Secant–Tangent Consistency (ASTC) Assumption)**.** Let $\boldsymbol{u}_t$ denote the trajectory state at flow time t generated by a discrete numerical solver with step size Δt. Under the linear optimal-transport parameterization $\boldsymbol{u}_t = (1-t)\boldsymbol{u}_0 + t\boldsymbol{u}_1$, the **Asymptotic Secant–Tangent Consistency (ASTC)** requires the finite-step secant velocity between endpoints to asymptotically recover the learned instantaneous velocity as the step size vanishes:

$$\boldsymbol{e}_t(\Delta t) \coloneqq \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - \frac{\boldsymbol{u}_t - \boldsymbol{u}_0}{t}, \qquad \lim_{\Delta t \to 0} \boldsymbol{e}_t(\Delta t) = \mathbf{0}. \tag{3}$$

Granted this assumption, the terminal correction can be consistently folded into the next-state update as

$$u_{t+\Delta t} = (1 - t - \Delta t)u_{\text{src}} + (t + \Delta t)\Pi_{\mathcal{M}}\big(\hat{u}_1(u_t, t)\big). \tag{4}$$

Eq. 4 unifies the update rule across existing post-hoc samplers, which differ primarily in their trajectory origin u_{src} and constraint handling: PCFM (Utkarsh et al., 2026) and its linear specialization ECI (Cheng et al., 2025) adopt the deterministic origin $u_{\text{src}} = u_0$, while ProFlow (Yu et al., 2026) employs a resampled noise vector $u_{\text{src}} = \varepsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.

4 THE MINIMAL NEXT-STEP CORRECTION FRAMEWORK

Among existing post-hoc constraint enforcement methods, we take the most representative one, PCFM as the primary audit target.

4.1 STEP-WISE EUCLIDEAN MINIMALITY

To audit the validity of the interpolation-consistency assumption, we construct a reference framework that circumvents it entirely. Instead of seeking minimal perturbation at the terminal state $\boldsymbol{u}_1$ as in PCFM, we formulate the local variational problem directly over the state space to minimize the disturbance to the immediate next-step state $\boldsymbol{u}_{t+\Delta t}$ subject to the physical constraint. We designate this formulation the **Minimal Step-wise Perturbation (MSP)** framework, which acts as a mathematical probe to isolate dynamical distortions induced by reverse OT interpolation.

Definition 2 (MSP Local Variational Objective)**.** Let $\boldsymbol{u}_{t+\Delta t}^{\mathrm{VFM}} \coloneqq \boldsymbol{u}_t + \Delta t \boldsymbol{v}_\theta(\boldsymbol{u}_t, t)$ denote the unconstrained push-forward state under the learned velocity field. We define the Minimal Step-wise Perturbation (MSP) state $\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}}$ as the solution to the local variational problem:

$$\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}} \coloneqq \underset{\boldsymbol{z}\in\mathbb{R}^D}{\arg\min} \frac{1}{2}\left\|\boldsymbol{z} - \boldsymbol{u}_{t+\Delta t}^{\mathrm{VFM}}\right\|_2^2 \quad \text{s.t.} \quad \boldsymbol{z} = (1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\boldsymbol{u}_1,\ h(\boldsymbol{u}_1) = \boldsymbol{0}. \tag{5}$$

Although defined over the next-step state rather than the terminal state, we prove that MSP retains the identical three-step structure of extrapolation, projection, and interpolation. Structurally, the only discrepancy between MSP and PCFM lies in a shift of the projection center.

Proposition 1 (Invariance of Process and Projection Center Shift; Proof in Appendix D.1)**.** Under the linear optimal-transport parameterization, the discrete MSP state update $u_{t+\Delta t}^{\mathrm{MSP}}$ defined identically preserves the extrapolation–projection–interpolation structure, with the projection center uniquely shifted from the unconstrained terminal prediction $\hat{u}_1$ to the equivalent reference center u_1^{ref}:

$$u_{t+\Delta t}^{\mathrm{MSP}} = (1-t-\Delta t)u_0 + (t+\Delta t)\Pi_{\mathcal{M}}(u_1^{\mathrm{ref}}), \tag{6}$$

where the equivalent projection center u_1^{ref} is uniquely given in closed form by:

$$u_1^{\mathrm{ref}} \coloneqq \frac{u_t + \Delta t v_\theta(u_t, t) - (1-t-\Delta t)u_0}{t+\Delta t}. \tag{7}$$

By subtracting the standard extrapolation $\hat{u}_1$ from the equivalent center u_1^{ref} in Eq. 7, direct algebraic simplification yields:

$$u_1^{\mathrm{ref}} - \hat{u}_1 = -\frac{1-t-\Delta t}{t+\Delta t}\Big[u_t - u_0 - t v_\theta(u_t, t)\Big] = -\frac{t(1-t-\Delta t)}{t+\Delta t} e_t, \tag{8}$$

where $e_t \coloneqq v_\theta(u_t, t) - \frac{u_t - u_0}{t}$ is precisely the secant–tangent defect defined in Definition 1. Eq. 8 establishes a bridge between the shift of the projection center and the instantaneous interpolation defect. Taking the step-refinement limit immediately leads to the following consequence:

Corollary 1 (Asymptotic Consequence of Interpolation Consistency; Proof in Appendix D.1)**.** For any fixed flow time $t \in (0,1)$, taking the continuum limit $\Delta t \to 0$ in Eq. 8 gives:

$$\lim_{\Delta t\to 0}\left(u_1^{\mathrm{ref}} - \hat{u}_1\right) = -(1-t)\lim_{\Delta t\to 0} e_t. \tag{9}$$

Consequently, if the asymptotic consistency condition holds ($\lim_{\Delta t\to 0} e_t = \boldsymbol{0}$) and the metric projection $\Pi_{\mathcal{M}}$ is locally continuous, the MSP and PCFM updates become asymptotically equivalent.

4.2 Asymptotic Order Analysis of the Surplus Velocity Field

To examine whether the reverse interpolation satisfies the ASTC Assumption under step refinement, we introduce the surplus velocity field as the normalized discrepancy between the two updates.

Definition 3 (Surplus Velocity Field)**.** Let $\boldsymbol{u}_t$ be the trajectory state at time t, and let $\boldsymbol{u}_{t+\Delta t}^{\mathrm{PCFM}}$ and $\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}}$ denote the one-step updates evaluated from the same state $(\boldsymbol{u}_0, \boldsymbol{u}_t, t, \Delta t)$. We define the surplus velocity field as

$$\boldsymbol{r}_t(\boldsymbol{u}_t, t) \coloneqq \frac{\boldsymbol{u}_{t+\Delta t}^{\mathrm{PCFM}} - \boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}}}{\Delta t} = \frac{t+\Delta t}{\Delta t}\left[\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) - \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})\right]. \tag{10}$$

$\boldsymbol{r}_t$ isolates the additional velocity induced by the reverse interpolation. We prove that r_t is asymptotically dominated by its component tangent to the constraint manifold $\mathcal{M}$ (Proposition 5 in Appendix A). In addition, under step refinement, this surplus velocity does not asymptotically vanish:

Proposition 2 (Divergence of the Surplus Velocity; Proof in Appendix D.2)**.** For any compact interval $t \in [a,b] \subset (0,1)$, under the non-degeneracy condition on the tangential defect $\|\mathcal{P}_{T_{\boldsymbol{p}_t}\mathcal{M}}\boldsymbol{e}_t(\Delta t)\|_2 \geq \gamma_T > 0$, the surplus velocity field satisfies:

$$\|\boldsymbol{r}_t(\boldsymbol{u}_t, t)\|_2 = \Theta\left(\Delta t^{-1}\right). \tag{11}$$

Correspondingly, the discrepancy between the equivalent reference center and the unconstrained terminal prediction remains non-vanishing in the continuum limit.

Corollary 2 (Non-Vanishing Center Discrepancy; Proof in Appendix D.2)**.** For any $t \in (0, 1)$, the discrepancy between reference centers remains of zeroth order:

$$\left\| \boldsymbol{u}_1^{\mathrm{ref}} - \hat{\boldsymbol{u}}_1 \right\|_2 = \Theta(1). \tag{12}$$

which contradicts the requirement $\lim_{\Delta t \to 0}(\boldsymbol{u}_1^{\mathrm{ref}} - \hat{\boldsymbol{u}}_1) = \mathbf{0}$ of the ASTC assumption, leading to the following remark.

Remark 1 (Invalidation of the ASTC Assumption after Projection)**.** The preceding analysis establishes that, the ASTC assumption fails after projection. In other words, the discrepancy between the interpolation secant velocity and the instantaneous predicted velocity cannot be regarded as a vanishing discretization error under step refinement.

Consequently, the reverse optimal-transport (OT) displacement interpolation assumption relied upon by existing methods is invalid after projection. Importantly, we do not challenge the OT assumption underlying flow matching itself; indeed, our analysis allows the OT construction to hold. Rather, we audit whether reverse OT interpolation remains valid after the terminal state has been projected away from the original flow manifold.

5 Impact of Surplus Velocity Field

Having shown that the ASTC assumption fails after projection, we further examine how the surplus dynamics distorts the generation process. We elaborate on its consequences from three perspectives: trajectory behavior, conditional information alignment, and posterior measure distortion.

5.1 Module I: Non-Vanishing Dynamical Intervention

The preceding analysis establishes that the surplus velocity remains non-vanishing under step refinement and induces a zeroth-order discrepancy between the reverse interpolation and the locally minimal next-state update (Proposition 2; Corollary 2). When the generated states remain in a bounded region Ω_B, this persistent displacement cannot be sustained indefinitely along a locally consistent transport direction as $\Delta t \to 0$. The resulting trajectories therefore exhibit non-monotone geometric deviations, including path bending, fluctuations, detours, or high frequency oscillations.

5.2 Module II: Failure of Posterior-Consistent Guidance

We next examine whether the tangential surplus displacement $\boldsymbol{w}_T^{\mathrm{PCFM}}(\boldsymbol{u}_t, t) \coloneqq \Delta t\, \mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{r}_t(\boldsymbol{u}_t, t)$ provides a posterior-directed steering mechanism. Let $c \in \mathcal{C}$ denote the conditioning variable, and define the local tangential displacement of the target posterior as $\boldsymbol{d}_{\mathrm{GT}}(c) \coloneqq \mathcal{P}_T(\hat{\boldsymbol{u}}_1)(\boldsymbol{\mu}_{\mathrm{GT}}(c) - \boldsymbol{u}_t)$, where $\boldsymbol{\mu}_{\mathrm{GT}}(c)$ is the conditional posterior mean.

Proposition 3 (History-Induced Surplus Dynamic; Proof in Appendix D.3)**.** Let $\dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau) \coloneqq \frac{d}{d\tau}\boldsymbol{v}_\theta(\boldsymbol{u}_\tau, \tau)$ denote the material derivative of the pretrained velocity field along the local unperturbed reference trajectory connecting $\boldsymbol{u}_0$ and $\boldsymbol{u}_t$. Then the secant–tangent defect satisfies the integral:

$$\boldsymbol{e}_t = \frac{1}{t}\int_0^t \tau\, \dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau)\, d\tau. \tag{13}$$

Its direction is determined entirely by the historical acceleration of the pretrained flow rather than by an explicit specified posterior correction. For fixed local trajectory variables $(\boldsymbol{u}_0, \boldsymbol{u}_t, t)$, the surplus exhibits zero functional sensitivity to the conditioning variable: $\left.\frac{\delta \boldsymbol{w}_T^{\mathrm{PCFM}}(\boldsymbol{u}_t, t)}{\delta c}\right|_{(\boldsymbol{u}_0, \boldsymbol{u}_t, t)} \equiv \mathbf{0}$. Consequently, the surplus displacement lacks intrinsic mechanism to adapt its direction to the posterior target $\boldsymbol{d}_{\mathrm{GT}}(c)$, causing its alignment to vary arbitrarily across task instances. When $\langle \boldsymbol{w}_T^{\mathrm{PCFM}}, \boldsymbol{d}_{\mathrm{GT}}(c) \rangle > 0$, the surplus incidentally nudges the state toward the posterior mean and improves reconstruction; otherwise, it steers the trajectory away from the target, actively degrading performance. This arbitrary alignment demonstrates that model performance is heavily governed by this unrecognized surplus dynamics.

5.3 MODULE III: DISTRIBUTIONAL DISTORTION

Let $\mathcal{M}_t \coloneqq (1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\mathcal{M}$ denote the admissible configuration manifold at time $t+\Delta t$. Let $\rho_E(\boldsymbol{z})$ denote the ambient probability density of the unconstrained push-forward state. Due to the spatial shift $t(1-t-\Delta t)\boldsymbol{e}_t$ in Eq. 8, the ambient density entering PCFM's projection is transported to $\rho_P(\boldsymbol{z})$. We define the surplus-induced density factor as $R_t(\boldsymbol{z}) \coloneqq \rho_P(\boldsymbol{z})/\rho_E(\boldsymbol{z})$.

Proposition 4 (Surplus-Induced Measure Deformation; Proof in Appendix D.4)**.** Under standard Euclidean tube conditioning onto $\mathcal{M}_t$ with Hausdorff measure $\mathcal{H}^k$ ($k = D-m$), the local constrained posterior densities realized by MSP and PCFM satisfy the exact closed-form expressions:

$$p_E^{\mathrm{MSP}}(\boldsymbol{z}) = \frac{\rho_E(\boldsymbol{z})}{\displaystyle\int_{\mathcal{M}_t} \rho_E(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi})}, \qquad \boldsymbol{z} \in \mathcal{M}_t, \tag{14}$$

$$p_E^{\mathrm{PCFM}}(\boldsymbol{z}) = \frac{\rho_E(\boldsymbol{z}) R_t(\boldsymbol{z})}{\displaystyle\int_{\mathcal{M}_t} \rho_E(\boldsymbol{\xi}) R_t(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi})} = p_E^{\mathrm{MSP}}(\boldsymbol{z}) \cdot \frac{R_t(\boldsymbol{z})}{\mathbb{E}_{p_E^{\mathrm{MSP}}}[R_t]}, \qquad \boldsymbol{z} \in \mathcal{M}_t. \tag{15}$$

Because the constraint geometry is non-uniform, the true hard-constraint posterior requires co-area information:

$$p_\star(\boldsymbol{z}) = \frac{\rho_E(\boldsymbol{z}) \left[\det\big(J(\boldsymbol{z})J(\boldsymbol{z})^\top\big)\right]^{-1/2}}{\displaystyle\int_{\mathcal{M}_t} \rho_E(\boldsymbol{\xi}) \left[\det\big(J(\boldsymbol{\xi})J(\boldsymbol{\xi})^\top\big)\right]^{-1/2} d\mathcal{H}^k(\boldsymbol{\xi})}, \qquad \boldsymbol{z} \in \mathcal{M}_t, \tag{16}$$

where J denotes the constraint Jacobian. In contrast, when such co-area information is not incorporated, the Euclidean posterior in Eq. 14 is the appropriate constrained law induced by the available ambient density and Euclidean projection geometry.

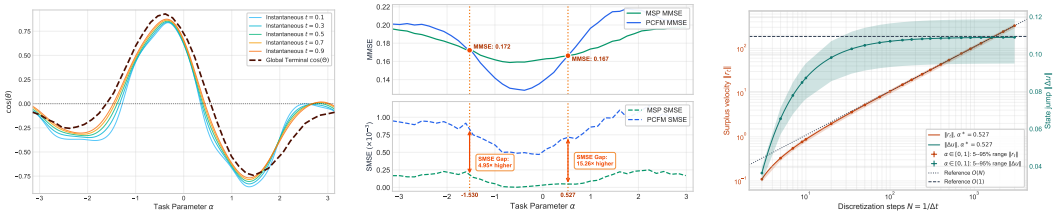
Crucially, R_t is an unrecognized, uncontrolled factor generated solely by the projection-induced surplus, not posterior objective. Even if the co-area correction is supplied afterwards, the fact R_t intrinsically distorts the distribution, preventing PCFM from recovering the co-area posterior. As empirically revealed by Xu et al. (2026), the bias incurred by terminal-projection methods is substantially more severe than the mere omission of the geometric co-area (Fixman) factor.

This result also clarifies the relationship with Section 5.2: when the surplus factor $R_t(\boldsymbol{z})$ happens to approximate the missing co-area factor $\left[\det\big(J(\boldsymbol{z})J(\boldsymbol{z})^\top\big)\right]^{-1/2}$, the surplus may partially bring the distribution closer to the co-area posterior, improving empirical performance; however, because R_t itself carries no co-area information, the gain is uninterpretable and lacks theoretical justification.

5.4 FURTHER DISCUSSION

For ProFlow, while a fresh random initialization is drawn at each step, our qualitative conclusions remain unchanged. Although ProFlow is motivated via ancestral sampling, conditioning on the measure-zero constraint set $\mathcal{C} = \{\boldsymbol{u} \mid h(\boldsymbol{u}) = \boldsymbol{0}\}$ is ill-posed under the ambient Lebesgue measure without an explicit limiting metric, as formalized by the Borel–Kolmogorov paradox (Trésor & Lukashchuk, 2025). Moreover, we prove that even if ProFlow's heuristic Gaussian surrogate assumption is accepted, MSP strictly preserves its formal Gaussian transition and ancestral-sampling structure without disrupting its probabilistic machinery, operating solely as a center rectification (see Appendix B for details). Thus, the applicability of MSP to ProFlow cannot be dismissed merely on the basis of alternative interpretations.

In PCFM, the inconsistency following OT interpolation is attributed to finite-step discretization. However, we have proved that applying OT interpolation after terminal projection inherently violates the asymptotic consistency condition, rather than being caused merely by discretization. Furthermore, PCFM can selectively introduce a relaxed penalty term for correction, which essentially restores manifold feasibility through a constraint-normal update. In contrast, the MSP framework eliminates the reliance on this invalid assumption at its source, requiring no additional feasibility correction. Further detailed analysis is provided in Appendix C.



(a) Instantaneous & global alignment.

(b) MMSE & SMSE at isomean roots.

(c) Step refinement scaling of the surplus velocity and state jump.

Figure 1: **Comprehensive dynamical and distributional audit in the 8D decay benchmark.** (a) tracks the instantaneous steering angle and global terminal alignment across task parameter $\alpha \in [-\pi, \pi]$; (b) reflects the trends and highlights the SMSE gap where MMSE match; (c) confirms the $\Theta(\Delta t^{-1}) = \Theta(N)$ scaling of the surplus velocity norm $\|\boldsymbol{r}_t\|_2$ alongside the $\Theta(1)$ plateau of the state jump $\|\Delta \boldsymbol{u}\|_2$ across step refinement. Shaded ribbons depict the 5–95% empirical range over the window $\alpha \in [0, 1]$, and the solid curve highlights the isomean crossing root $\alpha^* = 0.527$.

6 EXPERIMENTS

6.1 8D DECAY BENCHMARK

We consider a nonlinear 8D decay benchmark with state $\boldsymbol{u} \in \mathbb{R}^8$: the first four coordinates are constrained to an affine subspace $\mathcal{C} := \{\boldsymbol{u} \in \mathbb{R}^8 \mid \boldsymbol{u}_{1:4} = \boldsymbol{b}_{1:4}\}$, while the remaining four evolve freely under cubic decay dynamics $\boldsymbol{v}_{\mathrm{VFM}}(\boldsymbol{u}) = -\boldsymbol{\lambda} \odot \boldsymbol{u} - \boldsymbol{\beta} \odot \boldsymbol{u}^{\odot 3}$. Additional experimental details are provided in Appendix F.1

Tasks are parameterized by an angle α, which modulates the initial mean on the free subspace under fixed source covariance. To characterize the surplus dynamics across scales, we define two alignment metrics: instantaneous alignment $\cos(\theta_t) := \cos \angle(\boldsymbol{r}t, \boldsymbol{g}t)$, measuring the instantaneous alignment between the surplus velocity $\boldsymbol{r}t$ and the constraint-induced reference velocity $\boldsymbol{g}t := \boldsymbol{v}\mathrm{VFM}, \mathrm{cond}(\boldsymbol{u}t) - \boldsymbol{v}\mathrm{VFM}(\boldsymbol{u}t)$ at flow time t; and global alignment $\cos(\Theta) := \cos \angle(\boldsymbol{\mu}\mathrm{PCFM} - \boldsymbol{\mu}\mathrm{MSP}, \boldsymbol{\mu}\mathrm{GT} - \boldsymbol{\mu}\mathrm{VFM})$, measuring whether the net surplus-induced mean displacement aligns with the target displacement relative to the unconstrained prior.

Propagating samples under PCFM and MSP across tasks with identical discretizations, we report the directional alignment and the resulting first- and second-order statistical errors in Figs. 1a and 1b. To avoid favoring either method, we focus on the task parameter window $\alpha \in [0, 1]$, which contains the equal-MMSE crossing root $\alpha^* \approx 0.527$ and thus covers both regimes where MSP ($\alpha < \alpha^*$) and PCFM ($\alpha > \alpha^*$) achieve lower first-moment reconstruction error. Within this window, we further evaluate the one-step discrepancy under discretization refinement (Fig. 1c). The corresponding tangential/normal decomposition of the surplus velocity is shown in Fig. 4 in Appendix H.2.

Task-Dependent Directional Alignment. As shown in Fig. 1a, both instantaneous and global alignment vary substantially across tasks, with no consistent directional alignment. This indicates that the effect of the surplus is task-dependent rather than determined solely by the physical constraint, directly validating the absence of universal posterior alignment proved in Proposition 3.

Trends in Performance Gaps. The corresponding performance gaps also vary systematically across tasks, as shown in Fig. 1b. ΔMMSE changes sign, indicating that the surplus can either improve or degrade first-moment reconstruction, whereas PCFM consistently yields larger SMSE than MSP. At the two MMSE-crossing points, where first-moment errors are equal, the PCFM SMSE remains approximately $4.95\times$ and $15.26\times$ larger, respectively, indicating that the surplus can affect second-order statistics beyond mean reconstruction.

Correlation between Surplus Alignment and Statistical Discrepancies. We further correlate the global alignment $\cos(\Theta)$ with the performance discrepancies across the full task sweep (Fig. 3 in Appendix H.1). The correlation with ΔMMSE yields an almost perfect linear relationship ($r = -0.999, R^2 = 0.998$), providing clear empirical proof that the first-moment gap is directly

governed by the inner product between the surplus displacement and the target shift. The second-moment gap $\Delta\mathrm{SMSE}$ likewise exhibits a strong correlation ($r = -0.919, R^2 = 0.845$), with the residual spread reflecting higher-order nonlinear deformation across the manifold.

Asymptotic Scaling and Structure of the Surplus Velocity Field. As illustrated in Fig. 1c, the surplus velocity norm $\|\boldsymbol{r}_t\|_2$ exhibits strict $\Theta(\Delta t^{-1}) = \Theta(N)$ scaling and the underlying one-step state discrepancy $\|\Delta\boldsymbol{u}\|_2$ plateaus at a non-vanishing $\Theta(1)$ constant. This zeroth-order state deviation directly contradicts the ASTC assumption and confirms the Proposition 2. Furthermore, orthogonal decomposition across flow time t (Fig. 4 in Appendix H.2) reveals that the tangential component $\|\mathcal{P}_T\boldsymbol{r}_t\|_2$ accounts for nearly 100.0% of the total surplus magnitude across all tested $\alpha \in [0, 1]$, which validates Proposition 5 in Appendix A.

6.2 Inverse Measure Benchmark

To assess distributional distortion in a realistic PDE inverse problem, we adopt the codimension-3 Darcy inverse benchmark from CoCoS (Xu et al., 2026), where constraint sensitivity is heterogeneous across the posterior manifold. We reuse its problem setup, residual-band rejection arbiter, and evaluation protocol, and compare MSP and PCFM against the same gold-standard posterior. We report the average per-coordinate W_1 distance and its ratio to the arbiter's two-sample noise floor in Table 1 (full setup follow CoCoS; details in Appendix F.2). D denotes the dimensionality of the parameter space.

Table 1: Average per-coordinate Wasserstein distance ($\overline{W}_1 \downarrow$) and ratio to the noise floor ($\times$floor) on the Darcy benchmark across $D = 8$ and $D = 16$.

Method	$D = 8$ (3 seeds)		$D = 16$ (3 seeds)	
	$\overline{W}_1 \downarrow$	$\times$floor	$\overline{W}_1 \downarrow$	$\times$floor
With Co-Area Information				
Rej. Floor	0.012	$1.0\times$	0.018	$1.0\times$
CoCoS	0.036 ± 0.017	$3.0\times$	0.029 ± 0.009	$1.6\times$
Without Co-Area Information				
No Fixman	0.083 ± 0.024	$6.9\times$	0.062 ± 0.016	$3.4\times$
MSP	0.086 ± 0.025	$7.2\times$	0.063 ± 0.018	$3.5\times$
PCFM	0.402 ± 0.167	$\sim 34\times$	0.188 ± 0.093	$\sim 10\times$

Measure Distortion from Projection-Induced Surplus. Because neither MSP nor PCFM incorporates co-area information, we evaluate their intrinsic distributional behavior relative to the Euclidean constrained posterior, represented by the No-Fixman baseline. Across both dimensions, MSP closely matches this reference, yielding nearly identical $\overline{W}_1$ distances ($7.2\times$ and $3.5\times$ the noise floor). This agreement directly confirms that MSP realizes the uncorrupted Euclidean constrained law in Eq. 14. In contrast, PCFM incurs substantially larger deviations ($\sim 34\times$ and $\sim 10\times$), driven by the accumulation of the projection-induced surplus velocity that introduces the uncontrolled state-dependent density factor R_t. These results directly corroborate the surplus-induced measure deformation established in Proposition 4.

6.3 High-Dimensional Heat Benchmark

To evaluate whether the audited surplus dynamics persist in high-dimensional PDE systems and account for reported baseline discrepancies, we adopt the 1D Heat equation benchmark, sweeping the initial sinusoidal phase across multiple task settings. We specifically focus on a linear PDE system to establish a fair and rigorous comparison, as ECI is restricted to linear constraints. We test PCFM, ECI, and ProFlow alongside their respective surplus-removed counterparts, where the projection center is corrected via the MSP update. The quantitative evaluations are summarized in Table 2.

Prevalent Impacts of the Surplus Velocity. Comparing each method with its surplus-removed counterpart confirms the prevalence of surplus-induced dynamical effects. Shaded backgrounds highlight the model group achieving superior overall performance under each phase configuration. Notably, the effects induced by the surplus velocity fields of PCFM, ECI, and ProFlow exhibit a highly consistent cross-task pattern: both reconstruction and uncertainty errors peak around $\phi = \pi/2$ and reach their minimum at $\phi = 5\pi/4$. This shared trend indicates that the bias introduced by the surplus dynamics does not depend strongly on the specific formulation of each model, but is instead governed by the pretrained generative prior and the task itself.

Cross-Method Discrepancy Reduction. The performance gap between methods is particularly pronounced in uncertainty estimation. Across the five phase configurations, the ratio of the largest to

Table 2: Quantitative results on the 1D Linear Heat Benchmark for varying initial phase configurations $\phi \in [\pi/4, 5\pi/4]$. Columns labeled **Full Models (+)** denote original post-hoc methods; **Surplus-Removed (-)** denote variants with surplus dynamics eliminated via the MSP update. The **Ratio** is the maximum-to-minimum error ($\max / \min$) over the three models per metric. **Shaded background** marks the collectively better-performing model group for each phase.

Metric	Full Models (+)				Surplus-Removed (-)			
	PCFM	ECI	ProFlow	Ratio	PCFM	ECI	ProFlow	Ratio
Initial Phase $\phi = \pi/4$ *(45°)*								
MMSE ($\times 10^{-2}$) ↓	2.073	6.551	3.495	3.16×	**0.407**	**0.424**	**0.406**	**1.04×**
SMSE ($\times 10^{-2}$) ↓	1.218	7.410	**0.081**	91.8×	**0.376**	**0.365**	0.294	**1.28×**
FPD ↓	8.130	12.783	9.465	1.57×	**6.730**	**6.707**	**6.439**	**1.05×**
Initial Phase $\phi = \pi/2$ *(90°)*								
MMSE ($\times 10^{-2}$) ↓	2.748	10.378	5.199	3.78×	**0.115**	**0.145**	**0.110**	**1.31×**
SMSE ($\times 10^{-2}$) ↓	1.299	12.984	**0.103**	126.2×	**0.387**	**0.370**	0.311	**1.24×**
FPD ↓	7.363	11.723	11.450	1.59×	**6.264**	**6.465**	**5.838**	**1.11×**
Initial Phase $\phi = 3\pi/4$ *(135°)*								
MMSE ($\times 10^{-2}$) ↓	1.889	8.342	2.377	4.42×	**0.608**	**0.628**	**0.580**	**1.08×**
SMSE ($\times 10^{-2}$) ↓	1.453	12.294	**0.040**	307.3×	**0.399**	**0.386**	0.342	**1.17×**
FPD ↓	7.446	12.655	8.149	1.70×	**7.518**	**7.528**	**6.988**	**1.08×**
Initial Phase $\phi = \pi$ *(180°)*								
MMSE ($\times 10^{-2}$) ↓	**1.262**	8.129	**0.709**	11.46×	1.989	**1.989**	2.006	**1.01×**
SMSE ($\times 10^{-2}$) ↓	1.499	9.969	**0.016**	635.0×	**0.574**	**0.574**	0.467	**1.23×**
FPD ↓	**6.476**	12.084	**4.409**	2.74×	**10.213**	**10.213**	9.835	**1.04×**
Initial Phase $\phi = 5\pi/4$ *(225°, Out-of-Distribution)*								
MMSE ($\times 10^{-2}$) ↓	**0.506**	4.546	**0.298**	15.28×	3.875	**3.884**	4.047	**1.04×**
SMSE ($\times 10^{-2}$) ↓	1.493	6.718	**0.061**	110.1×	**1.204**	**1.252**	**1**.055	**1.19×**
FPD ↓	**5.946**	**11.099**	**3.042**	3.65×	12.893	12.981	12.780	**1.02×**

smallest SMSE among the original methods reaches $635.0\times$, whereas after removing the surplus term, it remains below $1.28\times$ for all tested phases. The corresponding MMSE and FPD ratios are likewise reduced to at most $1.31\times$ and $1.11\times$, respectively, making the three methods largely consistent. This indicates that, in the linear setting where all three methods are applicable, a substantial portion of the observed performance differences arises from the uncontrolled, task-dependent surplus term, which cannot provide stable performance improvements.

ProFlow and the Mechanism of Apparent Gains. Despite its stochastic re-initialization, ProFlow exhibits cross-task performance trends qualitatively identical to PCFM and ECI. Once the surplus dynamics is eliminated via MSP, its uncertainty metrics converge closely to those of PCFM and ECI. This confirms that, without co-area corrections, ProFlow lacks an intrinsic posterior-directed sampling mechanism, even if interpreted through ancestral sampling. Its favorable baseline SMSE likely stems in part from the stochastic variance of fresh random initializations, which serendipitously offsets the projection-induced uncertainty contraction.

7 Conclusion

In this work, we audit terminal projection methods for PDE inverse problems and demonstrate that the widely adopted reversibility under OT displacement interpolant fails after projection. The resulting surplus velocity field disrupts trajectories and lacks principled posterior-guidance capability. This unrecognized and uncontrolled degree of freedom compromises model interpretability and can lead to spurious performance gains. To resolve this issue, we derive the Minimal Step-wise Perturbation (MSP) as a minimal correction to this dynamical defect. While faithfully preserving the operational mechanics of the original models, MSP provides an interpretable and dynamically consistent alternative for constrained generative sampling.

ETHICS STATEMENT.

This work uses only synthetic PDE systems and simulated data and does not involve human subjects, personal data, or other sensitive information. We have no known ethical concerns beyond the responsible and accurate reporting of scientific results. All claims and empirical findings are presented with their stated assumptions and limitations, and prior methods and results are appropriately acknowledged and cited.

REPRODUCIBILITY STATEMENT.

For the theoretical results, all assumptions and complete proofs of the claims are provided in the Appendix, together with the corresponding derivations and definitions needed to reproduce the theoretical analysis. Detailed experimental configurations and evaluation procedures are also documented in the Appendix.

AI USE STATEMENT

In this work, we used generative AI tools for qualitative and thematic data analysis and for interpreting results. We did not use generative AI tools to develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in writing proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, or assist with translation. The generation of synthetic datasets and dataset cleaning or reformatting were not applicable to this work.

Additionally, we used generative AI tools to edit the research paper for readability and to identify relevant literature. We have reviewed all AI-assisted work. We validated the AI-assisted data analysis by comparing it against the actual experimental results, and assessed the accuracy of AI-assisted interpretations by comparing them with the expected findings. We iteratively revised the manuscript to ensure that AI-assisted interpretations and language editing remained consistent with the intended content of the paper. All relevant literature identified with AI assistance was independently read and reviewed by the authors to verify its relevance and accuracy and to avoid errors caused by hallucinated or inaccurate references. We take responsibility for the final content of this work, including any text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

Jan-Hendrik Bastek, WaiChing Sun, and Dennis Kochmann. Physics-informed diffusion models. In *International Conference on Learning Representations*, volume 2025, pp. 3360–3385, 2025.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. *arXiv preprint arXiv:2402.14017*, 2024.

Pengfei Cai, Utkarsh Utkarsh, Nofit Segal, Akshay Subramanian, Katharina Jager, Elton Pan, Alan Edelman, Christopher Vincent Rackauckas, and Rafael Gomez-Bombarelli. Enforcing constraints in molecular and crystalline generative models via physics-constrained flow matching. *AI for Accelerated Materials Design-ICLR*, 2026, 2026.

Chaoran Cheng, Boran Han, Danielle Maddix, Abdul Fatir Ansari, Andrew Stuart, Michael W Mahoney, and Bernie Wang. Gradient-free generation for hard-constrained systems. In *International Conference on Learning Representations*, volume 2025, pp. 100510–100539, 2025.

Weiguo Gao, Ming Li, and Qianxiao Li. Terminally constrained flow-based generative models from an optimal control perspective. *arXiv preprint arXiv:2601.09474*, 2026.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. Diffusionpde: Generative pde-solving under partial observation. *Advances in Neural Information Processing Systems*, 37: 130291–130323, 2024.

Sixun Huang, Jie Xiang, Huadong Du, and Xiaoqun Cao. Inverse problems in atmospheric science and their application. *Journal of Physics: Conference Series*, 12(1):45, jan 2005. doi: 10.1088/1742-6596/12/1/005. URL https://doi.org/10.1088/1742-6596/12/1/005.

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. *arXiv preprint arXiv:2305.17209*, 2023.

Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. Flowdps: Flow-driven posterior sampling for inverse problems. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12328–12337. IEEE, 2025.

Alaina Kolli, Theodoros Xenakis, Utkarsh Utkarsh, Pengfei Cai, Rafael Gomez-Bombarelli, Alan Edelman, and Christopher Vincent Rackauckas. Snap-fm: Sparse nonlinear accelerated projection for physics-constrained generative modeling. *arXiv preprint arXiv:2607.00095*, 2026.

Vince Kurtz and Alexander Davydov. Constrained flow matching via lagrangian dual flows. *arXiv preprint arXiv:2607.04513*, 2026.

Patrick Lailly. The seismic inverse problem as a sequence of before stack migrations. In *Conference on Inverse Scattering: Theory and Application*, pp. 206–220. SIAM, 1983.

Alfred Wei Lun Leong, Yatao Bian, and Bryan Kian Hsiang Low. Beyond equality: Inequality-constrained flow matching for physically admissible solutions. In *ICML 2026 Workshop on AI for Physics*, 2026.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Xingchao Liu, Lemeng Wu, Shujian Zhang, Chengyue Gong, Wei Ping, and Qiang Liu. Flowgrad: Controlling the output of generative odes with gradients. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24335–24344. IEEE, 2023.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Ségolène Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. Pnp-flow: Plug-and-play image restoration with flow matching. In *International Conference on Learning Representations*, volume 2025, pp. 45466–45492, 2025.

Eloi Moliner, Jaakko Lehtinen, and Vesa Välimäki. Solving audio inverse problems with a diffusion model. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023. doi: 10.1109/ICASSP49357.2023.10095637.

Jonghyun Park and Jong Chul Ye. Flowlps: Langevin-proximal sampling for flow-based inverse problem solvers. *arXiv preprint arXiv:2512.07150*, 2025.

Weize Quan, Jiaxi Chen, Yanli Liu, Dong-Ming Yan, and Peter Wonka. Deep learning-based image and video inpainting: A survey. *International Journal of Computer Vision*, 132(7):2367–2400, 2024.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Daniil Sherki, Ivan Oseledets, and Ekaterina Muravleva. Combining flow matching and transformers for efficient solution of bayesian inverse problems. In *AI4X 2025 International Conference*, 2025.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models, 2022. URL `https://arxiv.org/abs/2111.08005`.

Yixuan Sun, Anirban Samaddar, and Sandeep Madireddy. Composing flow-matching energies with known physics: Generation, ood detection, and inversion on pde fields. *arXiv preprint arXiv:2608.18004*, 2026.

Jan Tauberschmidt, Sophie Fellenz, Sebastian Vollmer, and Andrew Duncan. Physics-constrained fine-tuning of flow-matching models for generation and inverse problems. In *International Conference on Learning Representations*, volume 2026, pp. 143992–144037, 2026.

Raphaël Trésor and Mykola Lukashchuk. Resolution of the borel-kolmogorov paradox via the maximum entropy principle. *arXiv preprint arXiv:2509.24735*, 2025.

Utkarsh Utkarsh, Pengfei Cai, Alan Edelman, Rafael Gomez-Bombarelli, and Christopher Rackauckas. Physics-constrained flow matching: Sampling generative models with hard constraints. *Advances in Neural Information Processing Systems*, 38:160217–160252, 2026.

Jian Xu, Yanning Wu, Delu Zeng, John Paisley, and Qibin Zhao. The right measure for physics-constrained generation: A co-area correction for posterior-consistent pde inverse problems. *arXiv preprint arXiv:2606.04804*, 2026.

Zichao Yu, Ming Li, Wenyi Zhang, Difan Zou, and Weiguo Gao. Proflow: Zero-shot physics-consistent sampling via proximal flow guidance. *arXiv preprint arXiv:2601.20227*, 2026.

Jiaqi Zhang, Yiqi Wang, Hongjie Wu, Bohan Guo, Xinan Wang, Zichen Luo, Taotao Cai, Zhi Chen, and Mingkai Zheng. Rebridge-flow: Re-coupling posterior bridges in flow matching for image restoration. *arXiv preprint arXiv:2609.00811*, 2026a.

Zheng Zhang, Hao Tang, Yingying Hu, Zhanli Hu, and Jing Qin. Flowpet: Physics-informed symplectic flow matching for low-count pet reconstruction. *arXiv preprint arXiv:2607.11104*, 2026b.

A GEOMETRIC ANALYSIS OF THE SURPLUS VELOCITY

Proposition 5 (One-Step Tangent Dominance; Proof in Appendix D.5)**.** Assume the constraint manifold $\mathcal{M}$, defined by $h(\boldsymbol{u}) = \mathbf{0}$, is a C^2 embedded submanifold with full-rank constraint Jacobian on $\mathcal{M}$. Let $\boldsymbol{p}_t \coloneqq \Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1)$ and $\rho_t \coloneqq \|\Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}}) - \boldsymbol{p}_t\|_2$ denote the local chord distance between the projected endpoints. Let $\mathcal{P}_T(\hat{\boldsymbol{u}}_1) \equiv \mathcal{P}_{T_{\boldsymbol{p}_t}\mathcal{M}}$ denote the orthogonal projector onto the tangent space $T_{\boldsymbol{p}_t}\mathcal{M}$. Then, the surplus velocity admits the relative component scaling:

$$\frac{\|(\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1))\,\boldsymbol{r}_t\|_2}{\|\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{r}_t\|_2} = \mathcal{O}(\rho_t). \tag{17}$$

Corollary 3 (Exact Tangentiality under Affine Constraints; Proof in Appendix D.5)**.** When the constraint operator is affine, the constraint manifold has vanishing curvature ($\kappa \equiv 0$), and the normal component vanishes identically: $(\boldsymbol{I} - \mathcal{P}_T)\boldsymbol{r}_t \equiv \mathbf{0}$. For general nonlinear submanifolds, the normal component acts purely as a curvature-induced higher-order correction with respect to the projected chord separation ρ_t, as governed by Proposition 5.

B THEORETICAL ANALYSIS OF PROFLOW

For ProFlow (Yu et al., 2026), surplus-induced discrepancies take a distinct form due to the stochastic interpolation that samples a fresh initialization at each step. Even so, our qualitative conclusions hold unchanged. First, stochastic re-initialization yields non-vanishing zeroth-order intervention in the refinement limit; the resulting dynamics do not converge to the locally minimal update as $\Delta t \to 0$. Second, despite the terminal correction being condition-aware, its alignment with the true posterior direction is not guaranteed by the proximal formulation alone. Third, variance decomposition and the associated SMSE identity remain algebraically valid, yet the induced uncertainty change reflects both terminal correction and extra stochasticity from re-initialization, instead of arising purely from a deterministic surplus velocity field. Furthermore, ProFlow's random starting point introduces stochastic variance that may offset projection-induced uncertainty degradation.

ProFlow admits an interpretation via posterior-consistent ancestral sampling. Nevertheless, the physical constraint set $\mathcal{C} = \{\boldsymbol{u} \mid h(\boldsymbol{u}) = \mathbf{0}\}$ is typically a zero-measure set under the ambient Lebesgue measure, and no standard ambient density exists for the conditional distribution supported on $\mathcal{C}$. As illustrated by the Borel–Kolmogorov paradox, conditioning on null events is not uniquely defined without specifying an induced limiting metric. Hence, without rigorous measure-theoretic justification, direct application of the Gaussian ancestral-sampling framework to hard-constrained terminal distributions casts doubt on the well-posedness of the constructed probabilistic model. We do not reject the viability of this theoretical interpretation; we merely include ProFlow within our framework for joint analysis.

B.1 MSP UNDER PROFLOW'S GAUSSIAN SURROGATE ASSUMPTION

The purpose of this section is not to endorse the theoretical validity of this heuristic surrogate. Rather, we establish a strict conditional compatibility theorem: even if one accepts the foundational premise of ProFlow that replacing the intractable constrained transition with a local Gaussian surrogate is legitimate, the proposed Minimal Step-wise Perturbation (MSP) framework strictly preserves the exact Gaussian transition mechanics, the covariance geometry, and the sequential ancestral-sampling factorization. MSP merely rectifies the dynamical center within that assumed framework, confirming that our audit and correction do not compromise the operational or formal probabilistic structure of stochastic samplers even on their own theoretical terms.

Let $E \coloneqq \mathcal{C} \cap \mathcal{O}$ denote the joint physical-consistency and observational event on the state space $\mathbb{R}^D$, where $\mathcal{C} = \{\boldsymbol{u} \in \mathbb{R}^D \mid h(\boldsymbol{u}) = \mathbf{0}\}$ and $\mathcal{O}$ denotes downstream observations. In the stochastic bridge setting of ProFlow, a fresh Gaussian reference sample $\boldsymbol{u}_{0,t} \sim \mu_0 = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_D)$ is drawn at each integration step t. Conditioned on the local state pair $(\boldsymbol{u}_t, \boldsymbol{u}_{0,t})$, the MSP framework derives the unique, dynamically consistent reference center:

$$\boldsymbol{u}_{1,t}^{\mathrm{ref}} \coloneqq \frac{\boldsymbol{u}_t + \Delta t \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - (1 - t - \Delta t)\boldsymbol{u}_{0,t}}{t + \Delta t}. \tag{18}$$

Assumption 1 (ProFlow Gaussian Terminal Surrogate Assumption)**.** Assume, strictly following the axiomatic premise of ProFlow (Yu et al., 2026), that the intractable conditioned transition law $p(\boldsymbol{u}_1 \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E)$ can be validly approximated by a non-degenerate Gaussian density:

$$\widetilde{p}_t^{\mathrm{MSP}}(\boldsymbol{u}_1 \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E) \coloneqq \mathcal{N}\left(\boldsymbol{u}_1;\, \boldsymbol{\mu}_t^{\mathrm{MSP}},\, \boldsymbol{\Sigma}_t^{\mathrm{MSP}}\right), \tag{19}$$

where the prior anchor enters through the MSP-rectified center $\boldsymbol{u}_{1,t}^{\mathrm{ref}}$ via $p(\boldsymbol{u}_1 \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}) \approx \mathcal{N}(\boldsymbol{u}_{1,t}^{\mathrm{ref}}, \sigma_t^2 \boldsymbol{I}_D)$, and $\boldsymbol{\mu}_t^{\mathrm{MSP}} \in \mathbb{R}^D$ denotes the corresponding local maximum a posteriori (MAP) estimate.

Under Assumption 1, we establish the formal ancestral sampling properties of MSP-ProFlow:

Proposition 6 (Preservation of Gaussian Transition Form; Proof in Appendix D.6)**.** Under Assumption 1, let the MSP-ProFlow discrete update be constructed by interpolating a terminal sample $\widetilde{\boldsymbol{U}}_1 \sim \widetilde{p}_t^{\mathrm{MSP}}(\cdot \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E)$ with a freshly sampled independent Gaussian noise vector $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_D)$:

$$\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}} \coloneqq (1 - t - \Delta t)\boldsymbol{\varepsilon} + (t + \Delta t)\widetilde{\boldsymbol{U}}_1. \tag{20}$$

Then, the resulting one-step transition density $\widetilde{p}_t^{\mathrm{MSP}}(\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}} \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E)$ is strictly Gaussian:

$$\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}} \mid (\boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E) \sim \mathcal{N}\left((t + \Delta t)\boldsymbol{\mu}_t^{\mathrm{MSP}},\ (t + \Delta t)^2 \boldsymbol{\Sigma}_t^{\mathrm{MSP}} + (1 - t - \Delta t)^2 \boldsymbol{I}_D\right). \tag{21}$$

Proposition 7 (Ancestral-Sampling Consistency; Proof in Appendix D.6)**.** Under Assumption 1, let $\{\widetilde{p}_{t_n}^{\mathrm{MSP}}(\boldsymbol{u}_{t_n} \mid E)\}_{n=0}^{N}$ be the sequence of marginal distributions generated by initializing $\boldsymbol{u}_0 \sim \mu_0$ and iteratively applying the transition kernel established in Proposition 6. Then, the sequential state sequence generated by MSP-ProFlow satisfies the exact Chapman–Kolmogorov ancestral factorization:

$$\widetilde{p}_{t_{n+1}}^{\mathrm{MSP}}(\boldsymbol{u}_{t_{n+1}} \mid E) = \int_{\mathbb{R}^D} \widetilde{p}_{t_n}^{\mathrm{MSP}}(\boldsymbol{u}_{t_n} \mid E)\, \bar{K}_{t_n}^{\mathrm{MSP}}(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, E)\, d\boldsymbol{u}_{t_n}, \tag{22}$$

where $\bar{K}_{t_n}^{\mathrm{MSP}}(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, E) \coloneqq \int_{\mathbb{R}^D} \widetilde{p}_{t_n}^{\mathrm{MSP}}(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, E)\, \mu_0(d\boldsymbol{u}_{0,t_n})$ is the marginalized spatial transition kernel. Consequently, MSP-ProFlow constitutes a valid ancestral simulation chain within the surrogate Gaussian model family.

Corollary 4 (Exact Invariance of Covariance Geometry; Proof in Appendix D.6)**.** Under Assumption 1 and conditioned on accepting the Gaussian surrogate premise, suppose the observational condition is specified by a linear forward model $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{u}_1 + \boldsymbol{\eta}$ with Gaussian measurement noise $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \sigma_{\mathrm{obs}}^2 \boldsymbol{I}_m)$. Then, MSP-ProFlow preserves the exact posterior covariance matrix of the uncorrected baseline identically:

$$\boldsymbol{\Sigma}_t^{\mathrm{MSP}} = \left(\sigma_t^{-2}\boldsymbol{I}_D + \sigma_{\mathrm{obs}}^{-2}\boldsymbol{H}^\top \boldsymbol{H}\right)^{-1} \equiv \boldsymbol{\Sigma}_t^{\mathrm{ProFlow}}, \tag{23}$$

while modifying exclusively the posterior mean via the dynamically consistent reference center:

$$\boldsymbol{\mu}_t^{\mathrm{MSP}} = \boldsymbol{\Sigma}_t^{\mathrm{MSP}}\left(\sigma_t^{-2}\boldsymbol{u}_{1,t}^{\mathrm{ref}} + \sigma_{\mathrm{obs}}^{-2}\boldsymbol{H}^\top \boldsymbol{y}\right). \tag{24}$$

Remark 2 (Preservation of Probabilistic Machinery under Hypothetical Concession)**.** Proposition 6, Proposition 7, and Corollary 4 prove that the Minimal Step-wise Perturbation framework does not alter the mathematical structure of the ancestral-sampling machinery. Even if one sets aside the measure-theoretic concerns regarding Gaussian conditioning on null sets, the algebraic apparatus of ProFlow remains completely intact under MSP. MSP does not compete with ancestral sampling; rather, it provides the uniquely consistent dynamical center $\boldsymbol{u}_{1,t}^{\mathrm{ref}}$ to be consumed by that sampling machinery.

C Relaxed Penalty Correction in PCFM

In this section, we provide a detailed geometric and dynamical analysis of the relaxed penalty correction introduced in PCFM (Utkarsh et al., 2026). We formalize its mathematical formulation, characterize the subspace in which the penalty correction acts, and explain its relationship to the tangentially dominant surplus identified in our analysis. We then show why the Minimal Step-wise Perturbation (MSP) framework resolves the underlying dynamical incompatibility a priori, without relying on the secant–tangent consistency required by the OT interpolation or on an auxiliary penalty optimization.

C.1 Introduction of the Penalty Correction

In the PCFM framework, following the terminal projection $\boldsymbol{u}_{\text{proj}} \coloneqq \Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1)$ and the subsequent reverse Optimal Transport (OT) displacement interpolation, an intermediate pullback state is generated via a backward Euler step:

$$\hat{\boldsymbol{u}}_{\tau'} = \boldsymbol{u}_{\text{proj}} - (1-\tau')(\boldsymbol{u}_{\text{proj}} - \boldsymbol{u}_0) = (1-\tau')\boldsymbol{u}_0 + \tau'\boldsymbol{u}_{\text{proj}}, \tag{25}$$

where $\tau' \coloneqq \tau + \Delta\tau$. This OT interpolation relies on the secant–tangent consistency used in PCFM to justify replacing the backward flow by the OT displacement interpolant. After terminal projection, however, the current state is no longer generally on the original interpolation trajectory, so the corresponding instantaneous velocity need not coincide with the secant displacement.

PCFM attributes the resulting residual constraint violation to finite-step discretization and the nonlinearity of the constraint, and introduces an auxiliary optimization step termed *Relaxed Constraint Correction*:

$$\boldsymbol{u}_{\tau'} = \arg\min_{\boldsymbol{u}\in\mathbb{R}^D} \Psi(\boldsymbol{u}), \qquad \text{where} \qquad \Psi(\boldsymbol{u}) \coloneqq \frac{1}{2}\|\boldsymbol{u} - \hat{\boldsymbol{u}}_{\tau'}\|_2^2 + \frac{\lambda}{2}\|h(T(\boldsymbol{u}))\|_2^2. \tag{26}$$

Here, $\lambda > 0$ is a penalty parameter, and $T(\boldsymbol{u}) \coloneqq \boldsymbol{u} + (1-\tau')\boldsymbol{v}_\theta(\boldsymbol{u}, \tau')$ denotes the one-step forward extrapolation operator under the learned velocity field.

The objective in Eq. 26 has two roles: the first term keeps the corrected state close to the OT pullback state $\hat{\boldsymbol{u}}_{\tau'}$, while the second penalizes violation of the physical constraint after forward extrapolation under the current velocity field. Thus, the penalty does not enforce the secant–tangent consistency itself; rather, it provides a state-space correction that compensates for the mismatch between the OT interpolation and the forward constraint condition. In practice, this non-convex problem is solved through several inner-loop gradient steps.

C.2 Constraint-Normal Character of the Penalty Correction

We now characterize the geometric direction of the correction induced by Eq. 26. Define $\delta\boldsymbol{u}_\lambda \coloneqq \boldsymbol{u}_{\tau'} - \hat{\boldsymbol{u}}_{\tau'}$. The exact first-order optimality condition shows that this correction lies in the range of the composite constraint Jacobian transpose.

Proposition 8 (Normal Subspace Confinement of the Penalty Correction; Proof in Appendix D.7)**.** Let the constraint operator $h\colon \mathbb{R}^D \to \mathbb{R}^m$ be C^2-smooth with Jacobian $J_h(\boldsymbol{x}) \coloneqq \nabla h(\boldsymbol{x})^\top \in \mathbb{R}^{m\times D}$ having full row-rank on $\mathcal{M}$. Consider the local linearization of the extrapolation operator $DT(\boldsymbol{u}) \coloneqq \boldsymbol{I} + (1-\tau')\nabla_{\boldsymbol{u}}\boldsymbol{v}_\theta(\boldsymbol{u}, \tau')$, and define the composite constraint Jacobian:

$$\mathcal{J}(\boldsymbol{u}) \coloneqq J_h(T(\boldsymbol{u}))DT(\boldsymbol{u}) \in \mathbb{R}^{m\times D}. \tag{27}$$

Then, the corrective update satisfies:

$$\delta\boldsymbol{u}_\lambda \in \operatorname{range}\big(\mathcal{J}(\boldsymbol{u}_{\tau'})^\top\big) = \big(\ker \mathcal{J}(\boldsymbol{u}_{\tau'})\big)^\perp. \tag{28}$$

Furthermore, to first order around the reference state, the correction is given by:

$$\delta\boldsymbol{u}_\lambda = -\lambda\mathcal{J}^\top\left(\boldsymbol{I}_m + \lambda\mathcal{J}\mathcal{J}^\top\right)^{-1} h(T(\hat{\boldsymbol{u}}_{\tau'})) + \mathcal{O}\left(\|h(T(\hat{\boldsymbol{u}}_{\tau'}))\|_2^2\right). \tag{29}$$

Orthogonality to the Tangential Surplus. Proposition 8 shows that the λ-correction is confined to the range of the composite Jacobian transpose. For affine constraints, or locally in regimes where DT is close to the identity, this is close to the constraint-normal subspace. By contrast, the surplus velocity identified in Proposition 5 is dominated by its tangential component. Hence, in the affine or locally flat regime,

$$\langle \delta\boldsymbol{u}_\lambda,\, \mathcal{P}_T\boldsymbol{r}_t\rangle = 0, \tag{30}$$

while for nonlinear constraints the corresponding orthogonality holds to first order in the local curvature approximation.

Thus, the λ-penalty is an effective mechanism for restoring manifold feasibility by suppressing constraint-visible departures, but it does not directly remove the dominant tangential component of the surplus velocity. This distinction is important: the penalty correction compensates the constraint violation produced by the interpolation, whereas it does not reconstruct the locally minimal constrained dynamics identified by MSP.

C.3 Reason that MSP Eliminates the Need for Auxiliary Penalties

We now explain why the MSP framework requires neither the λ-penalty term nor inner-loop gradient iterations to obtain a compatible constrained update.

A-Priori Extrapolation Consistency. Let $\boldsymbol{u}_{\mathrm{proj}}^{\mathrm{MSP}} \coloneqq \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})$. By construction of the metric projection, $h(\boldsymbol{u}_{\mathrm{proj}}^{\mathrm{MSP}}) = \mathbf{0}$. The MSP feasible set is defined through the same affine pullback relation. Hence, for $\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}} = (1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\boldsymbol{u}_{\mathrm{proj}}^{\mathrm{MSP}}$, the corresponding terminal state recovered from this affine relation is identically:

$$\frac{\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}} - (1-t-\Delta t)\boldsymbol{u}_0}{t+\Delta t} = \frac{\big[(1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\boldsymbol{u}_{\mathrm{proj}}^{\mathrm{MSP}}\big] - (1-t-\Delta t)\boldsymbol{u}_0}{t+\Delta t} = \boldsymbol{u}_{\mathrm{proj}}^{\mathrm{MSP}}. \tag{31}$$

Consequently:

$$h\left(\frac{\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}} - (1-t-\Delta t)\boldsymbol{u}_0}{t+\Delta t}\right) = h\big(\boldsymbol{u}_{\mathrm{proj}}^{\mathrm{MSP}}\big) \equiv \mathbf{0}. \tag{32}$$

This identity follows directly from the MSP variational construction and does not invoke the invalid interpolation-consistency assumption or require $\boldsymbol{v}_\theta = \bar{\boldsymbol{v}}$. In contrast, the PCFM penalty is defined through the velocity-based extrapolation $T(\boldsymbol{u}) = \boldsymbol{u} + (1-\tau')\boldsymbol{v}_\theta(\boldsymbol{u}, \tau')$ and is introduced precisely to reduce the residual constraint violation of the OT interpolation under finite-step approximation.

Comparison and Takeaway. The distinction can therefore be summarized as follows. PCFM retains the terminal projection and OT pullback construction and subsequently applies an auxiliary penalty to reduce the constraint residual of the resulting state under the learned velocity field. This correction is mathematically effective for restoring constraint feasibility and, in the local affine regime, acts in the constraint-normal subspace. However, it does not remove the tangentially dominant surplus generated by the underlying interpolation mismatch.

MSP instead shifts the projection center to $\boldsymbol{u}_1^{\mathrm{ref}}$ *before* the constrained update is constructed. The hard constraint is therefore built directly into the local variational solution, rather than restored through a subsequent penalty optimization. Consequently, MSP simultaneously provides:

- exact constraint satisfaction of the terminal state recovered from the MSP affine pullback relation;
- elimination of the surplus velocity by construction, since the MSP state is itself the minimum-perturbation reference;
- elimination of the auxiliary inner-loop penalty optimization.

Thus, the two mechanisms should not be interpreted as competing forms of the same correction. The PCFM penalty is an effective post-hoc mechanism for compensating constraint-visible deviations caused by the approximate interpolation, whereas MSP provides an a-priori compatible construction that removes the source of the tangentially dominant surplus itself.

D Complete Mathematical Proofs

D.1 Proof of the MSP projection-center equivalence

Proof of Proposition 1.

Proof. Let $\boldsymbol{u}_t$ be the state at time t, and let $\boldsymbol{u}_{t+\Delta t}^{\mathrm{VFM}} \coloneqq \boldsymbol{u}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t)$ denote the unconstrained push-forward state. By Definition 2, the Minimal Step-wise Perturbation (MSP) update $\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}}$ is defined as the solution to the constrained variational problem:

$$\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}} = \operatorname*{arg\,min}_{\boldsymbol{z}\in\mathbb{R}^D} \frac{1}{2}\|\boldsymbol{z} - (\boldsymbol{u}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t))\|_2^2 \tag{33}$$

subject to the parameterized interpolation and physical feasibility constraints:

$$\boldsymbol{z} = (1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\boldsymbol{u}_1, \qquad h(\boldsymbol{u}_1) = \mathbf{0}. \tag{34}$$

Substituting the affine parameterization of $\boldsymbol{z}$ directly into the objective function Eq. 33, the optimization over $\boldsymbol{z}$ reduces to an optimization over the terminal state $\boldsymbol{u}_1$:

$$\boldsymbol{u}_1^{\mathrm{MSP}} = \operatorname*{arg\,min}_{\boldsymbol{u}_1 \in \mathbb{R}^D} \frac{1}{2} \|(1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\boldsymbol{u}_1 - \boldsymbol{u}_t - \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t)\|_2^2 \quad \text{s.t.} \quad h(\boldsymbol{u}_1) = \mathbf{0}. \tag{35}$$

Since $t \in (0,1)$ and $\Delta t > 0$, the scalar factor $t + \Delta t$ is strictly positive. Grouping terms with respect to $\boldsymbol{u}_1$ and factoring out $(t + \Delta t)$ from the Euclidean norm yields:

$$\begin{aligned} &\|(t+\Delta t)\boldsymbol{u}_1 - [\boldsymbol{u}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - (1-t-\Delta t)\boldsymbol{u}_0]\|_2^2 \\ &\quad = (t+\Delta t)^2 \left\| \boldsymbol{u}_1 - \frac{\boldsymbol{u}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - (1-t-\Delta t)\boldsymbol{u}_0}{t+\Delta t} \right\|_2^2. \end{aligned} \tag{36}$$

We define the equivalent reference center $\boldsymbol{u}_1^{\mathrm{ref}}$ as:

$$\boldsymbol{u}_1^{\mathrm{ref}} := \frac{\boldsymbol{u}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - (1-t-\Delta t)\boldsymbol{u}_0}{t+\Delta t}. \tag{37}$$

Substituting Eq. 36 and Eq. 37 into Eq. 35, the terminal variational problem simplifies to:

$$\boldsymbol{u}_1^{\mathrm{MSP}} = \operatorname*{arg\,min}_{\boldsymbol{u}_1 \in \mathbb{R}^D} \frac{1}{2}(t+\Delta t)^2 \left\|\boldsymbol{u}_1 - \boldsymbol{u}_1^{\mathrm{ref}}\right\|_2^2 \quad \text{s.t.} \quad h(\boldsymbol{u}_1) = \mathbf{0}. \tag{38}$$

Because $(t + \Delta t)^2 > 0$ is a strictly positive scalar constant, scaling the objective by $(t + \Delta t)^{-2}$ preserves the set of minimizers identically. Consequently, the optimization problem in Eq. 38 is strictly equivalent to the Euclidean metric projection of the center $\boldsymbol{u}_1^{\mathrm{ref}}$ onto the constraint manifold $\mathcal{M} := \{\boldsymbol{u} \in \mathbb{R}^D \mid h(\boldsymbol{u}) = \mathbf{0}\}$:

$$\boldsymbol{u}_1^{\mathrm{MSP}} = \operatorname*{arg\,min}_{\boldsymbol{u}_1 \in \mathcal{M}} \frac{1}{2}\left\|\boldsymbol{u}_1 - \boldsymbol{u}_1^{\mathrm{ref}}\right\|_2^2 =: \Pi_{\mathcal{M}}\left(\boldsymbol{u}_1^{\mathrm{ref}}\right). \tag{39}$$

Finally, substituting the optimal terminal state $\boldsymbol{u}_1^{\mathrm{MSP}} = \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})$ back into the linear interpolation formula Eq. 34, we obtain the discrete state update at time $t + \Delta t$:

$$\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}} = (1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\Pi_{\mathcal{M}}\left(\boldsymbol{u}_1^{\mathrm{ref}}\right). \tag{40}$$

This establishes that the local variational problem preserves the structural cycle of terminal extrapolation, metric projection, and linear pullback, with the projection target shifted uniquely to $\boldsymbol{u}_1^{\mathrm{ref}}$, completing the proof. □

Proof of Corollary 1.

Proof. Under the linear optimal-transport trajectory parameterization $\boldsymbol{u}_t = (1-t)\boldsymbol{u}_0 + t\boldsymbol{u}_1$, the learned instantaneous velocity $\boldsymbol{v}_\theta(\boldsymbol{u}_t, t)$ and the secant displacement velocity of the unperturbed interpolation trajectory are related via the secant–tangent defect vector:

$$\boldsymbol{e}_t := \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - \frac{\boldsymbol{u}_t - \boldsymbol{u}_0}{t}. \tag{41}$$

The unconstrained forward terminal prediction employed by standard terminal-projection samplers is given by:

$$\hat{\boldsymbol{u}}_1 := \boldsymbol{u}_t + (1-t)\boldsymbol{v}_\theta(\boldsymbol{u}_t, t). \tag{42}$$

From Proposition 1, the equivalent reference center $\boldsymbol{u}_1^{\mathrm{ref}}$ for the step-wise minimal update is defined as:

$$\boldsymbol{u}_1^{\mathrm{ref}} := \frac{\boldsymbol{u}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - (1-t-\Delta t)\boldsymbol{u}_0}{t+\Delta t}. \tag{43}$$

We express $\boldsymbol{u}_1^{\mathrm{ref}}$ directly in terms of the unconstrained terminal prediction $\hat{\boldsymbol{u}}_1$ and the defect vector $\boldsymbol{e}_t$. Combining terms over the common denominator $(t + \Delta t)$ gives:

$$\begin{aligned} \boldsymbol{u}_1^{\mathrm{ref}} &= \frac{\boldsymbol{u}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - (1-t-\Delta t)\boldsymbol{u}_0}{t+\Delta t} \\ &= \hat{\boldsymbol{u}}_1 - \frac{(t+\Delta t)\hat{\boldsymbol{u}}_1 - [\boldsymbol{u}_t + \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - (1-t-\Delta t)\boldsymbol{u}_0]}{t+\Delta t}. \end{aligned} \tag{44}$$

Substituting $\hat{\boldsymbol{u}}_1 = \boldsymbol{u}_t + (1-t)\boldsymbol{v}_\theta(\boldsymbol{u}_t, t)$ into the bracketed term in the numerator yields:

$$\begin{aligned}
&(t+\Delta t)\big[\boldsymbol{u}_t + (1-t)\boldsymbol{v}_\theta(\boldsymbol{u}_t,t)\big] - \boldsymbol{u}_t - \Delta t\, \boldsymbol{v}_\theta(\boldsymbol{u}_t,t) + (1-t-\Delta t)\boldsymbol{u}_0 \\
&\quad = (t+\Delta t-1)\boldsymbol{u}_t + \big[(t+\Delta t)(1-t) - \Delta t\big]\boldsymbol{v}_\theta(\boldsymbol{u}_t,t) + (1-t-\Delta t)\boldsymbol{u}_0 \\
&\quad = -(1-t-\Delta t)\boldsymbol{u}_t + t(1-t-\Delta t)\boldsymbol{v}_\theta(\boldsymbol{u}_t,t) + (1-t-\Delta t)\boldsymbol{u}_0 \\
&\quad = (1-t-\Delta t)\left[\boldsymbol{u}_0 - \boldsymbol{u}_t + t\,\boldsymbol{v}_\theta(\boldsymbol{u}_t,t)\right]. \qquad (45)
\end{aligned}$$

Recalling the definition of $\boldsymbol{e}_t$ in Eq. 41, we have the exact identity:

$$\boldsymbol{u}_0 - \boldsymbol{u}_t + t\,\boldsymbol{v}_\theta(\boldsymbol{u}_t,t) = t\left[\boldsymbol{v}_\theta(\boldsymbol{u}_t,t) - \frac{\boldsymbol{u}_t - \boldsymbol{u}_0}{t}\right] = t\boldsymbol{e}_t. \qquad (46)$$

Substituting this relation back into Eq. 45 and Eq. 44 establishes the exact algebraic relation between the two projection centers:

$$\boldsymbol{u}_1^{\mathrm{ref}} = \hat{\boldsymbol{u}}_1 - \frac{t(1-t-\Delta t)}{t+\Delta t}\boldsymbol{e}_t. \qquad (47)$$

Taking the step-refinement limit $\Delta t \to 0$ for any fixed flow homotopy time $t \in (0,1)$, we directly evaluate:

$$\lim_{\Delta t\to 0} \boldsymbol{u}_1^{\mathrm{ref}} = \lim_{\Delta t\to 0}\left(\hat{\boldsymbol{u}}_1 - \frac{t(1-t-\Delta t)}{t+\Delta t}\boldsymbol{e}_t\right). \qquad (48)$$

Under the ASTC assumption, the learned velocity recovers the secant velocity in the continuum limit, i.e., $\lim_{\Delta t\to 0}\boldsymbol{e}_t = \boldsymbol{0}$. Since the scalar prefactor converges to a finite value:

$$\lim_{\Delta t\to 0}\frac{t(1-t-\Delta t)}{t+\Delta t} = 1-t \in (0,1), \qquad (49)$$

Eq. 48 establishes the direct asymptotic equivalence of the projection centers:

$$\lim_{\Delta t\to 0}\boldsymbol{u}_1^{\mathrm{ref}} = \lim_{\Delta t\to 0}\hat{\boldsymbol{u}}_1. \qquad (50)$$

Now, recall that the standard PCFM update and the MSP update from the identical state $(\boldsymbol{u}_0, \boldsymbol{u}_t, t, \Delta t)$ are given by:

$$\boldsymbol{u}_{t+\Delta t}^{\mathrm{PCFM}} = (1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1), \qquad (51)$$

$$\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}} = (1-t-\Delta t)\boldsymbol{u}_0 + (t+\Delta t)\Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}}). \qquad (52)$$

Assuming continuity of the metric projection operator $\Pi_{\mathcal{M}}$ in a tubular neighborhood of $\mathcal{M}$, the limit of the projection operator commutes with the center limit:

$$\lim_{\Delta t\to 0}\Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}}) = \Pi_{\mathcal{M}}\left(\lim_{\Delta t\to 0}\boldsymbol{u}_1^{\mathrm{ref}}\right) = \Pi_{\mathcal{M}}\left(\lim_{\Delta t\to 0}\hat{\boldsymbol{u}}_1\right) = \lim_{\Delta t\to 0}\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1). \qquad (53)$$

Applying this equivalence directly to Eq. 51 and Eq. 52, and noting that the affine prefactor satisfies $\lim_{\Delta t\to 0}(t+\Delta t) = t$, we obtain:

$$\begin{aligned}
\lim_{\Delta t\to 0}\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}} &= (1-t)\boldsymbol{u}_0 + t\lim_{\Delta t\to 0}\Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}}) \\
&= (1-t)\boldsymbol{u}_0 + t\lim_{\Delta t\to 0}\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) \\
&= \lim_{\Delta t\to 0}\boldsymbol{u}_{t+\Delta t}^{\mathrm{PCFM}}. \qquad (54)
\end{aligned}$$

This proves directly that the step-wise minimal update $\boldsymbol{u}_{t+\Delta t}^{\mathrm{MSP}}$ asymptotically recovers the standard terminal-projection update $\boldsymbol{u}_{t+\Delta t}^{\mathrm{PCFM}}$ whenever the ASTC assumption holds. □

D.2 Proof of surplus divergence

Proof of Proposition 2.

Proof. By Definition 3, the surplus velocity field evaluated from the identical state $(\boldsymbol{u}_0, \boldsymbol{u}_t, t, \Delta t)$ is given by:

$$\boldsymbol{r}_t(\boldsymbol{u}_t, t) = \frac{t+\Delta t}{\Delta t}\left[\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) - \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})\right]. \qquad (55)$$

Taking the Euclidean norm on both sides yields:

$$\|\boldsymbol{r}_t(\boldsymbol{u}_t, t)\|_2 = \frac{t+\Delta t}{\Delta t}\left\|\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) - \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})\right\|_2. \tag{56}$$

To establish the tight $\Theta(\Delta t^{-1})$ scaling on any compact interval $t \in [a, b] \subset (0, 1)$, we evaluate matching upper and lower bounds for the projected chord distance $\left\|\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) - \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})\right\|_2$ as $\Delta t \to 0$.

By standard trajectory regularity, the state $\boldsymbol{u}_t$ remains within a compact set $\Omega_B \subset \mathbb{R}^D$, and the learned vector field is uniformly bounded: $\sup_{t\in[0,1],\boldsymbol{u}\in\Omega_B} \|\boldsymbol{v}_\theta(\boldsymbol{u}, t)\|_2 \le M < \infty$. Consequently, the unconstrained terminal prediction $\hat{\boldsymbol{u}}_1 = \boldsymbol{u}_t + (1-t)\boldsymbol{v}_\theta(\boldsymbol{u}_t, t)$ satisfies $\|\hat{\boldsymbol{u}}_1\|_2 = \mathcal{O}(1)$. Similarly, for the equivalent reference center $\boldsymbol{u}_1^{\mathrm{ref}} = \frac{\boldsymbol{u}_t + \Delta t \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - (1-t-\Delta t)\boldsymbol{u}_0}{t+\Delta t}$, the denominator is strictly bounded below by $t + \Delta t \ge a > 0$ for sufficiently small $\Delta t < a/2$, ensuring $\left\|\boldsymbol{u}_1^{\mathrm{ref}}\right\|_2 = \mathcal{O}(1)$.

Because h is C^2-smooth, the constraint manifold $\mathcal{M}$ is a non-empty closed embedded submanifold. Fix an arbitrary reference point $\boldsymbol{u}^\star \in \mathcal{M}$. By the defining nearest-point property of the metric projection $\Pi_{\mathcal{M}}$, for any $\boldsymbol{x} \in \mathbb{R}^D$, the distance satisfies $\|\Pi_{\mathcal{M}}(\boldsymbol{x}) - \boldsymbol{x}\|_2 \le \|\boldsymbol{u}^\star - \boldsymbol{x}\|_2$. Applying the triangle inequality:

$$\begin{aligned}\|\Pi_{\mathcal{M}}(\boldsymbol{x})\|_2 &\le \|\Pi_{\mathcal{M}}(\boldsymbol{x}) - \boldsymbol{x}\|_2 + \|\boldsymbol{x}\|_2 \\ &\le \|\boldsymbol{u}^\star - \boldsymbol{x}\|_2 + \|\boldsymbol{x}\|_2 \le \|\boldsymbol{u}^\star\|_2 + 2\|\boldsymbol{x}\|_2.\end{aligned} \tag{57}$$

Substituting $\boldsymbol{x} = \hat{\boldsymbol{u}}_1$ and $\boldsymbol{x} = \boldsymbol{u}_1^{\mathrm{ref}}$ into Eq. 57, both projected endpoints remain uniformly bounded. The triangle inequality directly gives the uniform upper bound:

$$\left\|\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) - \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})\right\|_2 \le \|\Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1)\|_2 + \left\|\Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})\right\|_2 \le C_{\max} < \infty. \tag{58}$$

We establish the lower bound via contradiction using compactness and the continuity of the normal bundle of C^2 submanifolds. Let $\boldsymbol{x}_{\Delta t} \coloneqq \hat{\boldsymbol{u}}_1$, $\boldsymbol{y}_{\Delta t} \coloneqq \boldsymbol{u}_1^{\mathrm{ref}}$, and denote their respective metric projections by $\boldsymbol{p}_{\Delta t} \coloneqq \Pi_{\mathcal{M}}(\boldsymbol{x}_{\Delta t})$ and $\boldsymbol{q}_{\Delta t} \coloneqq \Pi_{\mathcal{M}}(\boldsymbol{y}_{\Delta t})$. From Eq. 8, the pre-projection center discrepancy satisfies the exact algebraic identity:

$$\boldsymbol{x}_{\Delta t} - \boldsymbol{y}_{\Delta t} = \frac{t(1-t-\Delta t)}{t+\Delta t}\boldsymbol{e}_t(\Delta t). \tag{59}$$

We consider the non-degenerate regime where the tangential component of the interpolation defect on the projected tangent space is strictly bounded away from zero:

$$\left\|\mathcal{P}_{T_{\boldsymbol{p}_{\Delta t}}\mathcal{M}}\boldsymbol{e}_t(\Delta t)\right\|_2 \ge \gamma_T > 0. \tag{60}$$

Combining Eq. 59 and Eq. 60, for any $t \in [a, b]$ and sufficiently small Δt, the tangential center separation is strictly positive:

$$\left\|\mathcal{P}_{T_{\boldsymbol{p}_{\Delta t}}\mathcal{M}}(\boldsymbol{x}_{\Delta t} - \boldsymbol{y}_{\Delta t})\right\|_2 \ge \frac{a(1-b-\Delta t)}{b+\Delta t}\gamma_T \ge c_0\gamma_T > 0, \tag{61}$$

where $c_0 > 0$ is an independent scalar constant.

Now, assume for the sake of contradiction that the projected chord distance collapses along a subsequence $\Delta t_k \to 0$:

$$\lim_{k\to\infty} \|\boldsymbol{p}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k}\|_2 = 0. \tag{62}$$

Since $\boldsymbol{p}_{\Delta t_k}$ lies in a bounded subset of $\mathcal{M}$, by the Bolzano–Weierstrass theorem, we can extract a further subsequence (still indexed by k) converging to a limit point $\boldsymbol{p}^\star \in \mathcal{M}$, which implies:

$$\lim_{k\to\infty} \boldsymbol{p}_{\Delta t_k} = \boldsymbol{p}^\star, \qquad \lim_{k\to\infty} \boldsymbol{q}_{\Delta t_k} = \boldsymbol{p}^\star. \tag{63}$$

Because $\mathcal{M}$ is a C^2 embedded submanifold, the metric projection minimizers satisfy the first-order necessary optimality conditions, which require the projection residuals to lie strictly in the respective normal spaces:

$$\boldsymbol{x}_{\Delta t_k} - \boldsymbol{p}_{\Delta t_k} \in N_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}, \qquad \boldsymbol{y}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k} \in N_{\boldsymbol{q}_{\Delta t_k}}\mathcal{M}. \tag{64}$$

Projecting onto the tangent space $T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}$, the first relation yields:

$$\mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{x}_{\Delta t_k} - \boldsymbol{p}_{\Delta t_k}) = \mathbf{0}. \tag{65}$$

For the second relation, since $\boldsymbol{p}_{\Delta t_k} \to \boldsymbol{p}^\star$ and $\boldsymbol{q}_{\Delta t_k} \to \boldsymbol{p}^\star$, the C^1 smoothness of the tangent and normal bundles ensures that the orthogonal projector varies continuously:

$$\lim_{k\to\infty} \left\| \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}} - \mathcal{P}_{T_{\boldsymbol{q}_{\Delta t_k}}\mathcal{M}} \right\|_{\mathrm{op}} = 0. \tag{66}$$

Because the unconstrained residual $\|\boldsymbol{y}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k}\|_2$ remains uniformly bounded by Eq. 58, it follows that:

$$\begin{aligned} \left\| \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{y}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k}) \right\|_2 &\le \left\| \left(\mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}} - \mathcal{P}_{T_{\boldsymbol{q}_{\Delta t_k}}\mathcal{M}} \right) (\boldsymbol{y}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k}) \right\|_2 \\ &\quad + \left\| \mathcal{P}_{T_{\boldsymbol{q}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{y}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k}) \right\|_2 \\ &\to 0 + 0 = 0 \quad \text{as } k \to \infty. \end{aligned} \tag{67}$$

Now, decomposing the tangential difference between the two pre-projection centers:

$$\begin{aligned} \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{x}_{\Delta t_k} - \boldsymbol{y}_{\Delta t_k}) &= \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{x}_{\Delta t_k} - \boldsymbol{p}_{\Delta t_k}) \\ &\quad + \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{p}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k}) \\ &\quad + \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{q}_{\Delta t_k} - \boldsymbol{y}_{\Delta t_k}). \end{aligned} \tag{68}$$

Applying the triangle inequality and substituting Eq. 65 and Eq. 67:

$$\begin{aligned} \left\| \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{x}_{\Delta t_k} - \boldsymbol{y}_{\Delta t_k}) \right\|_2 &\le 0 + \left\| \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{p}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k}) \right\|_2 + \left\| \mathcal{P}_{T_{\boldsymbol{p}_{\Delta t_k}}\mathcal{M}}(\boldsymbol{q}_{\Delta t_k} - \boldsymbol{y}_{\Delta t_k}) \right\|_2 \\ &\le \|\boldsymbol{p}_{\Delta t_k} - \boldsymbol{q}_{\Delta t_k}\|_2 + o(1) \\ &\to 0 \quad \text{as } k \to \infty. \end{aligned} \tag{69}$$

This directly contradicts Eq. 61, which bounded this tangential norm below by $c_0\gamma_T > 0$. Therefore, the projected chord distance cannot collapse to zero, establishing the uniform positive lower bound:

$$\liminf_{\Delta t\to 0} \left\| \Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) - \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}}) \right\|_2 \ge C_{\min} > 0. \tag{70}$$

Combining the uniform upper bound Eq. 58 and the strictly positive lower bound Eq. 70 directly yields:

$$\left\| \Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) - \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}}) \right\|_2 = \Theta(1). \tag{71}$$

Finally, substituting this exact order into Eq. 56, and noting that the prefactor satisfies $a \le t + \Delta t \le b + \Delta t = \Theta(1)$ on $[a, b]$, we conclude:

$$\|\boldsymbol{r}_t(\boldsymbol{u}_t, t)\|_2 = \frac{t + \Delta t}{\Delta t} \cdot \Theta(1) = \Theta\left(\Delta t^{-1}\right). \tag{72}$$

This completes the rigorous proof of the first-order kinetic singularity of the surplus velocity field. □

Proof of Corollary 2.

Proof. From Eq. 47, the exact algebraic difference between the equivalent reference center $\boldsymbol{u}_1^{\mathrm{ref}}$ and the unconstrained terminal prediction $\hat{\boldsymbol{u}}_1$ satisfies:

$$\boldsymbol{u}_1^{\mathrm{ref}} - \hat{\boldsymbol{u}}_1 = -\frac{t(1 - t - \Delta t)}{t + \Delta t} \boldsymbol{e}_t(\Delta t). \tag{73}$$

Taking the Euclidean norm on both sides for any fixed $t \in (0, 1)$:

$$\left\| \boldsymbol{u}_1^{\mathrm{ref}} - \hat{\boldsymbol{u}}_1 \right\|_2 = \frac{t(1 - t - \Delta t)}{t + \Delta t} \|\boldsymbol{e}_t(\Delta t)\|_2. \tag{74}$$

As the temporal step size vanishes ($\Delta t \to 0$), the scalar prefactor converges to a strictly positive constant:

$$\lim_{\Delta t\to 0} \frac{t(1 - t - \Delta t)}{t + \Delta t} = 1 - t \in (0, 1). \tag{75}$$

In any non-degenerate regime where the learned vector field possesses curvature (such that the secant–tangent defect remains bounded away from zero, $\|\boldsymbol{e}_t\|_2 \geq \gamma > 0$), the center discrepancy is lower-bounded by $(1-t)\gamma > 0$ and upper-bounded by standard regularity.

Consequently, the center discrepancy does not vanish as $\Delta t \to 0$, but strictly satisfies:

$$\left\|\boldsymbol{u}_1^{\mathrm{ref}} - \hat{\boldsymbol{u}}_1\right\|_2 = \Theta(1). \tag{76}$$

This establishes that the terminal center mismatch is of non-vanishing zeroth order, completing the proof. □

D.3 Proof of History-Induced Surplus

Proof of Proposition 3.

Proof. To characterize the physical origin of the secant–tangent defect $\boldsymbol{e}_t$, we consider the local unperturbed reference characteristic connecting the trajectory origin $\boldsymbol{u}_0$ and the evaluated state $\boldsymbol{u}_t$. Let $\boldsymbol{u}_s$ for $s \in [0, t]$ denote the continuous reference flow path governed by the pretrained dynamics $\dot{\boldsymbol{u}}(s) = \boldsymbol{v}_\theta(\boldsymbol{u}_s, s)$. Along this reference characteristic, the net displacement satisfies:

$$\boldsymbol{u}_t - \boldsymbol{u}_0 = \int_0^t \boldsymbol{v}_\theta(\boldsymbol{u}_s, s)\, ds. \tag{77}$$

Substituting this relation into the definition of the ASTC defect vector $\boldsymbol{e}_t$ in Eq. 3:

$$\boldsymbol{e}_t \coloneqq \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - \frac{\boldsymbol{u}_t - \boldsymbol{u}_0}{t} = \boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - \frac{1}{t}\int_0^t \boldsymbol{v}_\theta(\boldsymbol{u}_s, s)\, ds. \tag{78}$$

Expressing the instantaneous velocity $\boldsymbol{v}_\theta(\boldsymbol{u}_t, t)$ through the trivial identity $\boldsymbol{v}_\theta(\boldsymbol{u}_t, t) = \frac{1}{t}\int_0^t \boldsymbol{v}_\theta(\boldsymbol{u}_t, t)\, ds$, Eq. 78 rewrites as:

$$\boldsymbol{e}_t = \frac{1}{t}\int_0^t \left[\boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - \boldsymbol{v}_\theta(\boldsymbol{u}_s, s)\right] ds. \tag{79}$$

Now, let $\dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau) \coloneqq \frac{d}{d\tau}\boldsymbol{v}_\theta(\boldsymbol{u}_\tau, \tau) = \partial_\tau \boldsymbol{v}_\theta + (\boldsymbol{v}_\theta \cdot \nabla_{\boldsymbol{u}})\boldsymbol{v}_\theta$ denote the convective material acceleration along the reference flow. The velocity discrepancy between time s and time t is:

$$\boldsymbol{v}_\theta(\boldsymbol{u}_t, t) - \boldsymbol{v}_\theta(\boldsymbol{u}_s, s) = \int_s^t \dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau)\, d\tau. \tag{80}$$

Substituting this relation into Eq. 79 yields a double integral over the triangular domain $0 \leq s \leq \tau \leq t$:

$$\boldsymbol{e}_t = \frac{1}{t}\int_0^t \left(\int_s^t \dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau)\, d\tau\right) ds. \tag{81}$$

Applying Fubini's theorem to interchange the order of integration:

$$\begin{aligned} \boldsymbol{e}_t &= \frac{1}{t}\int_0^t \left(\int_0^\tau ds\right) \dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau)\, d\tau \\ &= \frac{1}{t}\int_0^t \tau\, \dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau)\, d\tau. \end{aligned} \tag{82}$$

This establishes the exact reference-flow representation in Eq. 13: the defect vector $\boldsymbol{e}_t$ is precisely the time-weighted historical average of the prior flow acceleration.

Next, from the proof of Corollary 3, the surplus velocity is given by $\boldsymbol{r}_t = \frac{t(1-t-\Delta t)}{\Delta t}\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{e}_t$. Substituting Eq. 82 into the definition of the tangential step intervention $\boldsymbol{w}_T^{\mathrm{PCFM}}(\boldsymbol{u}_t, t) \coloneqq \Delta t\, \mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{r}_t$ yields:

$$\begin{aligned} \boldsymbol{w}_T^{\mathrm{PCFM}}(\boldsymbol{u}_t, t) &= \Delta t \cdot \left[\frac{t(1-t-\Delta t)}{\Delta t}\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{e}_t\right] \\ &= t(1-t-\Delta t)\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\left[\frac{1}{t}\int_0^t \tau\, \dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau)\, d\tau\right] \\ &= (1-t-\Delta t)\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\int_0^t \tau\, \dot{\boldsymbol{v}}_\theta(\boldsymbol{u}_\tau, \tau)\, d\tau. \end{aligned} \tag{83}$$

The functional inputs to the right-hand side of Eq. 83 consist solely of the trajectory history $\{\boldsymbol{u}_\tau\}_{\tau\in[0,t]}$ generated by the unconstrained prior vector field $\boldsymbol{v}_\theta$ and the local manifold projection $\mathcal{P}_T$. In particular, the integrand contains no functional dependence on the downstream task conditioning variable c or observational data y. Computing the functional derivative with respect to c at any fixed instantaneous trajectory state $(\boldsymbol{u}_0, \boldsymbol{u}_t, t)$ immediately gives:

$$\left.\frac{\delta \boldsymbol{w}_T^{\mathrm{PCFM}}(\boldsymbol{u}_t, t)}{\delta c}\right|_{(\boldsymbol{u}_0,\boldsymbol{u}_t,t)} \equiv \mathbf{0}. \tag{84}$$

Consequently, $\boldsymbol{w}_T^{\mathrm{PCFM}}$ is determined entirely by prior trajectory curvature rather than the target posterior direction $\boldsymbol{d}_{\mathrm{GT}}(c) \coloneqq \mathcal{P}_T(\hat{\boldsymbol{u}}_1)(\boldsymbol{\mu}_{\mathrm{GT}}(c) - \boldsymbol{u}_t)$, precluding any universal or condition-aware posterior alignment mechanism. □

D.4 Proof of Measure Deformation

Proof of Proposition 4.

Proof. Let $\mathcal{T}_\epsilon(\mathcal{M}_t) \coloneqq \{\boldsymbol{z} \in \mathbb{R}^D \mid \mathrm{dist}(\boldsymbol{z}, \mathcal{M}_t) \le \epsilon\}$ denote the tubular neighborhood of the k-dimensional smooth embedded submanifold $\mathcal{M}_t$ of radius $\epsilon > 0$. Under Euclidean metric projection onto $\mathcal{M}_t$, the uniform conditional expectation of any bounded continuous test function $\varphi \in C_b(\mathcal{M}_t)$ with respect to an ambient density ρ is defined via the tubular limit:

$$\mathbb{E}_{p_E}[\varphi] \coloneqq \lim_{\epsilon\to 0} \frac{\displaystyle\int_{\mathcal{T}_\epsilon(\mathcal{M}_t)} \varphi(\Pi_{\mathcal{M}_t}(\boldsymbol{z}))\, \rho(\boldsymbol{z})\, d\boldsymbol{z}}{\displaystyle\int_{\mathcal{T}_\epsilon(\mathcal{M}_t)} \rho(\boldsymbol{z})\, d\boldsymbol{z}}. \tag{85}$$

For a sufficiently small radius $\epsilon > 0$ within the reach of $\mathcal{M}_t$, the projection $\boldsymbol{\xi} = \Pi_{\mathcal{M}_t}(\boldsymbol{z}) \in \mathcal{M}_t$ and the normal vector $\boldsymbol{\eta} \coloneqq \boldsymbol{z} - \boldsymbol{\xi} \in (T_{\boldsymbol{\xi}}\mathcal{M}_t)^\perp$ define a diffeomorphic tubular coordinate chart $(\boldsymbol{\xi}, \boldsymbol{\eta}) \in \mathcal{M}_t \times B_\epsilon^{D-k}(\mathbf{0})$. In these normal coordinates, the ambient Lebesgue volume element decomposes as:

$$d\boldsymbol{z} = \theta(\boldsymbol{\xi}, \boldsymbol{\eta})\, d\boldsymbol{\eta}\, d\mathcal{H}^k(\boldsymbol{\xi}), \tag{86}$$

where $\theta(\boldsymbol{\xi}, \boldsymbol{\eta}) = \det(\boldsymbol{I} - S_{\boldsymbol{\eta}}(\boldsymbol{\xi})) = 1 + \mathcal{O}(\|\boldsymbol{\eta}\|_2)$ is the Jacobian determinant of the normal bundle trivialization, and $S_{\boldsymbol{\eta}}$ denotes the shape operator of $\mathcal{M}_t$ along the normal direction $\boldsymbol{\eta}$.

Since ρ is continuously differentiable and φ depends solely on the projected base point $\boldsymbol{\xi}$, a Taylor expansion along the normal slice yields $\rho(\boldsymbol{z})\theta(\boldsymbol{\xi}, \boldsymbol{\eta}) = \rho(\boldsymbol{\xi}) + \mathcal{O}(\|\boldsymbol{\eta}\|_2)$. Integrating out the normal coordinate $\boldsymbol{\eta}$ over the $(D-k)$-dimensional Euclidean ball $B_\epsilon^{D-k}(\mathbf{0})$:

$$\begin{aligned}\int_{\mathcal{T}_\epsilon(\mathcal{M}_t)} \varphi(\Pi_{\mathcal{M}_t}(\boldsymbol{z}))\, \rho(\boldsymbol{z})\, d\boldsymbol{z} &= \int_{\mathcal{M}_t} \varphi(\boldsymbol{\xi}) \left[\int_{B_\epsilon^{D-k}(\mathbf{0})} (\rho(\boldsymbol{\xi}) + \mathcal{O}(\|\boldsymbol{\eta}\|_2))\, d\boldsymbol{\eta}\right] d\mathcal{H}^k(\boldsymbol{\xi}) \\ &= \omega_{D-k}\epsilon^{D-k} \int_{\mathcal{M}_t} \varphi(\boldsymbol{\xi})\, \rho(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi}) + \mathcal{O}\left(\epsilon^{D-k+1}\right),\end{aligned} \tag{87}$$

where ω_{D-k} denotes the volume of the unit $(D-k)$-dimensional ball. Setting $\varphi \equiv 1$ gives the corresponding denominator expansion:

$$\int_{\mathcal{T}_\epsilon(\mathcal{M}_t)} \rho(\boldsymbol{z})\, d\boldsymbol{z} = \omega_{D-k}\epsilon^{D-k} \int_{\mathcal{M}_t} \rho(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi}) + \mathcal{O}\left(\epsilon^{D-k+1}\right). \tag{88}$$

Substituting Eq. 87 and Eq. 88 into Eq. 85, the leading normal volume prefactor $\omega_{D-k}\epsilon^{D-k}$ cancels identically, and the higher-order curvature terms vanish in the limit $\epsilon \to 0$, yielding the exact Euclidean constrained probability density:

$$p_E(\boldsymbol{z}) = \frac{\rho(\boldsymbol{z})}{\displaystyle\int_{\mathcal{M}_t} \rho(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi})}, \qquad \forall \boldsymbol{z} \in \mathcal{M}_t. \tag{89}$$

For the MSP update, the pre-projection state is the unconstrained push-forward state governed by the ambient density $\rho_E(\boldsymbol{z})$. Directly evaluating Eq. 89 with $\rho = \rho_E$ yields Eq. 14:

$$p_E^{\mathrm{MSP}}(\boldsymbol{z}) = \frac{\rho_E(\boldsymbol{z})}{\displaystyle\int_{\mathcal{M}_t} \rho_E(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi})}. \tag{90}$$

For PCFM, the pre-projection state is shifted by the surplus vector $\boldsymbol{s}_t \coloneqq t(1-t-\Delta t)\boldsymbol{e}_t$. Under this spatial transformation $S_t(\boldsymbol{z}) \coloneqq \boldsymbol{z} + \boldsymbol{s}_t(\boldsymbol{z})$, the ambient density entering the projection is transported to $\rho_P(\boldsymbol{z}) = \rho_E(S_t^{-1}(\boldsymbol{z}))|\det DS_t^{-1}(\boldsymbol{z})| = \rho_E(\boldsymbol{z})R_t(\boldsymbol{z})$, where $R_t(\boldsymbol{z}) \coloneqq \rho_P(\boldsymbol{z})/\rho_E(\boldsymbol{z})$. Applying the Euclidean conditioning formula Eq. 89 with the transported ambient density $\rho = \rho_P$:

$$p_E^{\mathrm{PCFM}}(\boldsymbol{z}) = \frac{\rho_P(\boldsymbol{z})}{\displaystyle\int_{\mathcal{M}_t} \rho_P(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi})} = \frac{\rho_E(\boldsymbol{z})R_t(\boldsymbol{z})}{\displaystyle\int_{\mathcal{M}_t} \rho_E(\boldsymbol{\xi})R_t(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi})}. \tag{91}$$

Dividing the numerator and denominator by the normalization constant $\int_{\mathcal{M}_t} \rho_E(\boldsymbol{\xi})\, d\mathcal{H}^k(\boldsymbol{\xi})$ directly establishes:

$$p_E^{\mathrm{PCFM}}(\boldsymbol{z}) = \frac{\frac{\rho_E(\boldsymbol{z})}{\int_{\mathcal{M}_t} \rho_E\, d\mathcal{H}^k} \cdot R_t(\boldsymbol{z})}{\displaystyle\int_{\mathcal{M}_t} R_t(\boldsymbol{\xi}) \frac{\rho_E(\boldsymbol{\xi})}{\int_{\mathcal{M}_t} \rho_E\, d\mathcal{H}^k}\, d\mathcal{H}^k(\boldsymbol{\xi})} = p_E^{\mathrm{MSP}}(\boldsymbol{z}) \cdot \frac{R_t(\boldsymbol{z})}{\mathbb{E}_{p_E^{\mathrm{MSP}}}[R_t]}. \tag{92}$$

This completes the proof. □

D.5 Proof of surplus geometry

Proof of Proposition 5.

Proof. Let $\boldsymbol{p}_t \coloneqq \Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1)$ and $\boldsymbol{q}_t \coloneqq \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})$. By definition of the metric projection onto the constraint manifold $\mathcal{M}$, both points satisfy the physical constraint identically:

$$h(\boldsymbol{p}_t) = \mathbf{0}, \qquad h(\boldsymbol{q}_t) = \mathbf{0}. \tag{93}$$

From Definition 3, the surplus velocity field evaluated from the identical state $(\boldsymbol{u}_0, \boldsymbol{u}_t, t, \Delta t)$ is given by:

$$\boldsymbol{r}_t(\boldsymbol{u}_t, t) = \frac{t+\Delta t}{\Delta t}\left(\boldsymbol{p}_t - \boldsymbol{q}_t\right). \tag{94}$$

We analyze the local geometry of the displacement between these two points on the constraint manifold. Define the chord displacement vector and its Euclidean norm:

$$\boldsymbol{\delta}_t \coloneqq \boldsymbol{q}_t - \boldsymbol{p}_t, \qquad \rho_t \coloneqq \|\boldsymbol{\delta}_t\|_2. \tag{95}$$

Because $\mathcal{M}$ is a C^2 embedded submanifold and the constraint Jacobian $J(\boldsymbol{u}) \coloneqq \nabla h(\boldsymbol{u})^\top \in \mathbb{R}^{m\times D}$ has full row rank on $\mathcal{M}$, the tangent space $T_{\boldsymbol{p}_t}\mathcal{M} = \ker J(\boldsymbol{p}_t)$ is well-defined. Let $\mathcal{P}_T(\hat{\boldsymbol{u}}_1)$ denote the orthogonal projector onto $T_{\boldsymbol{p}_t}\mathcal{M}$. The chord vector $\boldsymbol{\delta}_t$ admits the unique orthogonal decomposition:

$$\boldsymbol{\delta}_t = \mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{\delta}_t + \left(\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1)\right)\boldsymbol{\delta}_t. \tag{96}$$

Since $h(\boldsymbol{p}_t) = h(\boldsymbol{q}_t) = \mathbf{0}$ and h is twice continuously differentiable, a Taylor expansion of h around $\boldsymbol{p}_t$ yields:

$$\mathbf{0} = h(\boldsymbol{q}_t) = h(\boldsymbol{p}_t) + J(\boldsymbol{p}_t)\boldsymbol{\delta}_t + \boldsymbol{R}_t(\boldsymbol{\delta}_t), \tag{97}$$

where the second-order remainder $\boldsymbol{R}_t$ is controlled by the local extrinsic curvature of $\mathcal{M}$. For $\boldsymbol{q}_t$ in a sufficiently small tubular neighborhood of $\boldsymbol{p}_t$ (ρ_t sufficiently small), there exists a local geometric constant $C > 0$ such that:

$$\|\boldsymbol{R}_t(\boldsymbol{\delta}_t)\|_2 \le C\|\boldsymbol{\delta}_t\|_2^2 = C\rho_t^2. \tag{98}$$

Using $h(\boldsymbol{p}_t) = \mathbf{0}$, we obtain the direct Jacobian action bound:

$$\|J(\boldsymbol{p}_t)\boldsymbol{\delta}_t\|_2 = \|-\boldsymbol{R}_t(\boldsymbol{\delta}_t)\|_2 \le C\rho_t^2 = \mathcal{O}(\rho_t^2). \tag{99}$$

The orthogonal projector onto the normal subspace $(T_{\boldsymbol{p}_t}\mathcal{M})^\perp$ is explicitly given by:

$$\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1) = J(\boldsymbol{p}_t)^\top \left[J(\boldsymbol{p}_t)J(\boldsymbol{p}_t)^\top\right]^{-1} J(\boldsymbol{p}_t). \tag{100}$$

By the full-row-rank assumption, the Gram operator $\left[J(\boldsymbol{p}_t)J(\boldsymbol{p}_t)^\top\right]$ is strictly positive definite and invertible. Applying the normal projector to $\boldsymbol{\delta}_t$ and substituting Eq. 99 yields the second-order normal bound:

$$\begin{aligned}\|(\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1))\,\boldsymbol{\delta}_t\|_2 &= \left\| J(\boldsymbol{p}_t)^\top \left[J(\boldsymbol{p}_t)J(\boldsymbol{p}_t)^\top\right]^{-1} J(\boldsymbol{p}_t)\boldsymbol{\delta}_t \right\|_2 \\ &\le \left\| J(\boldsymbol{p}_t)^\top \left[J(\boldsymbol{p}_t)J(\boldsymbol{p}_t)^\top\right]^{-1} \right\|_2 \|J(\boldsymbol{p}_t)\boldsymbol{\delta}_t\|_2 = \mathcal{O}(\rho_t^2).\end{aligned} \tag{101}$$

We next establish the matching first-order lower bound for the tangential component. Applying the reverse triangle inequality to the orthogonal decomposition Eq. 96 and substituting the normal bound Eq. 101, we have:

$$\|\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{\delta}_t\|_2 = \|\boldsymbol{\delta}_t - (\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1))\,\boldsymbol{\delta}_t\|_2 \ge \|\boldsymbol{\delta}_t\|_2 - \|(\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1))\,\boldsymbol{\delta}_t\|_2 = \rho_t - \mathcal{O}(\rho_t^2). \tag{102}$$

Combined with the standard non-expansiveness upper bound $\|\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{\delta}_t\|_2 \le \|\boldsymbol{\delta}_t\|_2 = \rho_t$, Eq. 102 guarantees that the tangential chord displacement scales strictly as:

$$\|\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{\delta}_t\|_2 = \Theta(\rho_t). \tag{103}$$

This confirms that the displacement between two proximate points on a C^2 submanifold is strictly first-order tangential and second-order normal.

Substituting $\boldsymbol{p}_t - \boldsymbol{q}_t = -\boldsymbol{\delta}_t$ into Eq. 94, we scale the decomposed displacement by $(t + \Delta t)/\Delta t$. Since $t \in (0, 1)$ is fixed and $\Delta t \to 0$, the scalar prefactor satisfies $(t + \Delta t)/\Delta t = \Theta(\Delta t^{-1})$. For the tangential component of the surplus velocity, we have:

$$\|\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{r}_t\|_2 = \frac{t + \Delta t}{\Delta t}\|\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{\delta}_t\|_2 = \Theta\left(\frac{\rho_t}{\Delta t}\right). \tag{104}$$

Similarly, for the normal component:

$$\|(\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1))\,\boldsymbol{r}_t\|_2 = \frac{t + \Delta t}{\Delta t}\|(\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1))\,\boldsymbol{\delta}_t\|_2 = \mathcal{O}\left(\frac{\rho_t^2}{\Delta t}\right). \tag{105}$$

Taking the ratio between the normal and tangential norms, since the denominator is strictly $\Theta(\rho_t/\Delta t)$, the division is well-posed and yields:

$$\frac{\|(\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1))\,\boldsymbol{r}_t\|_2}{\|\mathcal{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{r}_t\|_2} = \frac{\mathcal{O}(\rho_t^2/\Delta t)}{\Theta(\rho_t/\Delta t)} = \mathcal{O}(\rho_t). \tag{106}$$

Hence, the normal departure is of strictly higher geometric order in terms of the chord separation ρ_t, confirming that any non-tangential deviation on a C^2 manifold is purely a curvature-induced effect governed by $\mathcal{O}(\rho_t)$. □

Proof of Corollary 3.

Proof. First, for general nonlinear submanifolds, the statement follows directly from Proposition 5, where the relative normal-to-tangential ratio scales as $\mathcal{O}(\rho_t)$, confirming that any deviation from the tangent space is purely a curvature-induced higher-order effect.

Next, suppose the constraint operator h is affine, expressed as $h(\boldsymbol{u}) = \boldsymbol{J}\boldsymbol{u} - \boldsymbol{b}$ where $\boldsymbol{J} \in \mathbb{R}^{m\times D}$ has full row rank. Under affine constraints, the constraint Jacobian $\boldsymbol{J}$ is constant across $\mathbb{R}^D$, and the orthogonal tangent projector:

$$\mathcal{P}_T = \boldsymbol{I} - \boldsymbol{J}^\top(\boldsymbol{J}\boldsymbol{J}^\top)^{-1}\boldsymbol{J} \tag{107}$$

is globally invariant and independent of the base point.

Since both $\boldsymbol{p}_t = \Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1)$ and $\boldsymbol{q}_t = \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}})$ strictly reside on the affine manifold $\mathcal{M}$, they satisfy the linear feasibility conditions $\boldsymbol{J}\boldsymbol{p}_t = \boldsymbol{b}$ and $\boldsymbol{J}\boldsymbol{q}_t = \boldsymbol{b}$. Their difference therefore satisfies the homogeneous linear constraint:

$$\boldsymbol{J}(\boldsymbol{p}_t - \boldsymbol{q}_t) = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}. \tag{108}$$

Applying the normal-space projector $(\boldsymbol{I} - \mathcal{P}_T)$ directly to the displacement vector $(\boldsymbol{p}_t - \boldsymbol{q}_t)$:

$$(\boldsymbol{I} - \mathcal{P}_T)(\boldsymbol{p}_t - \boldsymbol{q}_t) = \boldsymbol{J}^\top(\boldsymbol{J}\boldsymbol{J}^\top)^{-1}\boldsymbol{J}(\boldsymbol{p}_t - \boldsymbol{q}_t) = \boldsymbol{J}^\top(\boldsymbol{J}\boldsymbol{J}^\top)^{-1}\boldsymbol{0} = \boldsymbol{0}. \tag{109}$$

Multiplying by the non-zero scalar $(t + \Delta t)/\Delta t$, the normal component of the surplus velocity vanishes identically:

$$(\boldsymbol{I} - \mathcal{P}_T(\hat{\boldsymbol{u}}_1))\, \boldsymbol{r}_t = \boldsymbol{0}. \tag{110}$$

This establishes that under affine constraints, the surplus velocity is strictly and purely tangential to the manifold.

We next derive the exact closed-form expression for $\boldsymbol{r}_t$ in terms of the ASTC defect $\boldsymbol{e}_t$. Because Euclidean projection onto an affine subspace is an affine mapping whose linear derivative is $\mathcal{P}_T$, the projection difference simplifies to:

$$\boldsymbol{p}_t - \boldsymbol{q}_t = \Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1) - \Pi_{\mathcal{M}}(\boldsymbol{u}_1^{\mathrm{ref}}) = \mathcal{P}_T\left(\hat{\boldsymbol{u}}_1 - \boldsymbol{u}_1^{\mathrm{ref}}\right). \tag{111}$$

From Eq. 47, the exact center difference satisfies:

$$\hat{\boldsymbol{u}}_1 - \boldsymbol{u}_1^{\mathrm{ref}} = +\frac{t(1 - t - \Delta t)}{t + \Delta t}\boldsymbol{e}_t. \tag{112}$$

Substituting Eq. 112 into Eq. 111, and substituting the resulting expression into the definition of $\boldsymbol{r}_t$ in Eq. 94, we obtain:

$$\begin{aligned}
\boldsymbol{r}_t &= \frac{t + \Delta t}{\Delta t}\mathcal{P}_T\left(\hat{\boldsymbol{u}}_1 - \boldsymbol{u}_1^{\mathrm{ref}}\right) \\
&= \frac{t + \Delta t}{\Delta t}\mathcal{P}_T\left(+\frac{t(1 - t - \Delta t)}{t + \Delta t}\boldsymbol{e}_t\right) \\
&= \frac{t(1 - t - \Delta t)}{\Delta t}\mathcal{P}_T\boldsymbol{e}_t.
\end{aligned} \tag{113}$$

□

D.6 Proof of MSP-ProFlow compatibility

Proof of Proposition 6.

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space. We evaluate the distribution of the random vector $\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}}$ conditioned on the σ-algebra $\mathcal{F}_t \coloneqq \sigma(\boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E)$. By the MSP-ProFlow update rule in Eq. 20, we have the affine decomposition:

$$\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}} = \boldsymbol{Z}_1 + \boldsymbol{Z}_2, \quad \text{where} \quad \boldsymbol{Z}_1 \coloneqq (t + \Delta t)\widetilde{\boldsymbol{U}}_1, \quad \boldsymbol{Z}_2 \coloneqq (1 - t - \Delta t)\boldsymbol{\varepsilon}. \tag{114}$$

We characterize the joint distribution of $(\boldsymbol{Z}_1, \boldsymbol{Z}_2)$ through their conditional characteristic functions:

$$\varphi_{\boldsymbol{Z}_1|\mathcal{F}_t}(\boldsymbol{\xi}) \coloneqq \mathbb{E}\left[\exp\left(i\boldsymbol{\xi}^\top \boldsymbol{Z}_1\right) \mid \mathcal{F}_t\right], \qquad \varphi_{\boldsymbol{Z}_2|\mathcal{F}_t}(\boldsymbol{\xi}) \coloneqq \mathbb{E}\left[\exp\left(i\boldsymbol{\xi}^\top \boldsymbol{Z}_2\right) \mid \mathcal{F}_t\right], \quad \forall \boldsymbol{\xi} \in \mathbb{R}^D. \tag{115}$$

By Assumption 1, the conditional distribution of $\widetilde{\boldsymbol{U}}_1$ is given by $\mathcal{N}(\boldsymbol{\mu}_t^{\mathrm{MSP}}, \boldsymbol{\Sigma}_t^{\mathrm{MSP}})$. By the standard property of characteristic functions under linear transformations:

$$\begin{aligned}
\varphi_{\boldsymbol{Z}_1|\mathcal{F}_t}(\boldsymbol{\xi}) &= \mathbb{E}\left[\exp\left(i\left((t + \Delta t)\boldsymbol{\xi}\right)^\top \widetilde{\boldsymbol{U}}_1\right) \mid \mathcal{F}_t\right] \\
&= \exp\left(i(t + \Delta t)\boldsymbol{\xi}^\top \boldsymbol{\mu}_t^{\mathrm{MSP}} - \frac{1}{2}(t + \Delta t)^2 \boldsymbol{\xi}^\top \boldsymbol{\Sigma}_t^{\mathrm{MSP}} \boldsymbol{\xi}\right).
\end{aligned} \tag{116}$$

Next, by the operational construction of the algorithm, the noise vector $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_D)$ is sampled independently of the trajectory history up to time t, the current state $\boldsymbol{u}_t$, the bridge origin $\boldsymbol{u}_{0,t}$, and the terminal candidate $\widetilde{\boldsymbol{U}}_1$. Consequently, $\boldsymbol{\varepsilon}$ is conditionally independent of $\widetilde{\boldsymbol{U}}_1$ given $\mathcal{F}_t$. The conditional characteristic function of $\boldsymbol{Z}_2$ evaluates to:

$$\begin{aligned}
\varphi_{\boldsymbol{Z}_2|\mathcal{F}_t}(\boldsymbol{\xi}) &= \mathbb{E}\left[\exp\left(i\left((1 - t - \Delta t)\boldsymbol{\xi}\right)^\top \boldsymbol{\varepsilon}\right) \mid \mathcal{F}_t\right] \\
&= \exp\left(-\frac{1}{2}(1 - t - \Delta t)^2 \boldsymbol{\xi}^\top \boldsymbol{I}_D \boldsymbol{\xi}\right).
\end{aligned} \tag{117}$$

By conditional independence, the conditional characteristic function of the sum $\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}} = \boldsymbol{Z}_1 + \boldsymbol{Z}_2$ factorizes as the product of the individual characteristic functions:

$$\varphi_{\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}}|\mathcal{F}_t}(\boldsymbol{\xi}) = \varphi_{\boldsymbol{Z}_1|\mathcal{F}_t}(\boldsymbol{\xi}) \cdot \varphi_{\boldsymbol{Z}_2|\mathcal{F}_t}(\boldsymbol{\xi}). \tag{118}$$

Substituting Eq. 116 and Eq. 117 into this product yields:

$$\begin{aligned}\varphi_{\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}}|\mathcal{F}_t}(\boldsymbol{\xi}) &= \exp\left(i(t+\Delta t)\boldsymbol{\xi}^\top \boldsymbol{\mu}_t^{\text{MSP}} - \frac{1}{2}(t+\Delta t)^2 \boldsymbol{\xi}^\top \boldsymbol{\Sigma}_t^{\text{MSP}} \boldsymbol{\xi} - \frac{1}{2}(1-t-\Delta t)^2 \boldsymbol{\xi}^\top \boldsymbol{\xi}\right) \\ &= \exp\left(i\boldsymbol{\xi}^\top \left[(t+\Delta t)\boldsymbol{\mu}_t^{\text{MSP}}\right] - \frac{1}{2}\boldsymbol{\xi}^\top \left[(t+\Delta t)^2 \boldsymbol{\Sigma}_t^{\text{MSP}} + (1-t-\Delta t)^2 \boldsymbol{I}_D\right] \boldsymbol{\xi}\right).\end{aligned} \tag{119}$$

By the Lévy Inversion Theorem, a probability distribution is uniquely determined by its characteristic function. The expression in Eq. 119 is identically the characteristic function of a multivariate Gaussian distribution with mean vector:

$$\mathbb{E}\left[\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}} \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E\right] = (t+\Delta t)\boldsymbol{\mu}_t^{\text{MSP}}, \tag{120}$$

and covariance matrix:

$$\text{Cov}\left(\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}} \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E\right) = (t+\Delta t)^2 \boldsymbol{\Sigma}_t^{\text{MSP}} + (1-t-\Delta t)^2 \boldsymbol{I}_D. \tag{121}$$

This establishes that the transition density $\widetilde{p}_t^{\text{MSP}}(\boldsymbol{u}_{t+\Delta t}^{\text{MSP-PF}} \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, E)$ is strictly Gaussian, completing the proof. □

Proof of Proposition 7.

Proof. Let $0 = t_0 < t_1 < \cdots < t_N = 1$ denote the discrete temporal grid. We establish by mathematical induction that for all $n \in \{0, 1, \ldots, N\}$, the discrete state $\boldsymbol{u}_{t_n}$ distributed under the sequential MSP-ProFlow algorithm has marginal distribution given by $\widetilde{p}_{t_n}^{\text{MSP}}(\boldsymbol{u}_{t_n} \mid E)$.

For the base step ($n = 0$), by algorithmic definition, the initial state is sampled directly from the prior: $\boldsymbol{u}_{t_0} \sim \mu_0 = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_D) \equiv \widetilde{p}_0^{\text{MSP}}(\cdot \mid E)$. Hence, the induction base holds identically.

For the inductive step, assume that the hypothesis holds at step n, meaning the state $\boldsymbol{u}_{t_n}$ admits the marginal probability density $\widetilde{p}_{t_n}^{\text{MSP}}(\boldsymbol{u}_{t_n} \mid E)$ with respect to the Lebesgue measure on $\mathbb{R}^D$. At step n, a stochastic bridge origin $\boldsymbol{u}_{0,t_n}$ is drawn independently from μ_0. The joint distribution of the current state and the fresh source origin factorizes as $p(\boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n} \mid E) = \widetilde{p}_{t_n}^{\text{MSP}}(\boldsymbol{u}_{t_n} \mid E)\, \mu_0(\boldsymbol{u}_{0,t_n})$. According to Proposition 6, the transition probability to the next state $\boldsymbol{u}_{t_{n+1}}$ conditional on $(\boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, E)$ is governed by the regular Markov Gaussian transition kernel:

$$\begin{aligned}&K_{t_n}^{\text{MSP}}\left(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, E\right) \\ &\quad \coloneqq \mathcal{N}\left(\boldsymbol{u}_{t_{n+1}};\, (t_n+\Delta t)\boldsymbol{\mu}_{t_n}^{\text{MSP}},\, (t_n+\Delta t)^2 \boldsymbol{\Sigma}_{t_n}^{\text{MSP}} + (1-t_n-\Delta t)^2 \boldsymbol{I}_D\right).\end{aligned} \tag{122}$$

By the law of total probability, the joint distribution of $(\boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, \boldsymbol{u}_{t_{n+1}})$ under the algorithm is:

$$p\left(\boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, \boldsymbol{u}_{t_{n+1}} \mid E\right) = \widetilde{p}_{t_n}^{\text{MSP}}(\boldsymbol{u}_{t_n} \mid E)\, \mu_0(\boldsymbol{u}_{0,t_n})\, K_{t_n}^{\text{MSP}}\left(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, E\right). \tag{123}$$

To obtain the marginal distribution of $\boldsymbol{u}_{t_{n+1}}$, we integrate out the auxiliary source variable $\boldsymbol{u}_{0,t_n}$ and the preceding trajectory state $\boldsymbol{u}_{t_n}$. First, integrating over the bridge origin $\boldsymbol{u}_{0,t_n} \in \mathbb{R}^D$ defines the spatial Markov transition kernel:

$$\bar{K}_{t_n}^{\text{MSP}}\left(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, E\right) \coloneqq \int_{\mathbb{R}^D} K_{t_n}^{\text{MSP}}\left(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, E\right)\, \mu_0(\boldsymbol{u}_{0,t_n})\, d\boldsymbol{u}_{0,t_n}. \tag{124}$$

By Tonelli–Fubini's theorem (since all probability densities are non-negative and integrable), the order of integration can be interchanged freely:

$$\begin{aligned}p\left(\boldsymbol{u}_{t_{n+1}} \mid E\right) &= \int_{\mathbb{R}^D}\int_{\mathbb{R}^D} p\left(\boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, \boldsymbol{u}_{t_{n+1}} \mid E\right)\, d\boldsymbol{u}_{0,t_n}\, d\boldsymbol{u}_{t_n} \\ &= \int_{\mathbb{R}^D} \widetilde{p}_{t_n}^{\text{MSP}}(\boldsymbol{u}_{t_n} \mid E) \left[\int_{\mathbb{R}^D} K_{t_n}^{\text{MSP}}\left(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, \boldsymbol{u}_{0,t_n}, E\right) \mu_0(\boldsymbol{u}_{0,t_n})\, d\boldsymbol{u}_{0,t_n}\right] d\boldsymbol{u}_{t_n} \\ &= \int_{\mathbb{R}^D} \widetilde{p}_{t_n}^{\text{MSP}}(\boldsymbol{u}_{t_n} \mid E)\, \bar{K}_{t_n}^{\text{MSP}}\left(\boldsymbol{u}_{t_{n+1}} \mid \boldsymbol{u}_{t_n}, E\right)\, d\boldsymbol{u}_{t_n}.\end{aligned} \tag{125}$$

Eq. 125 is identically the Chapman–Kolmogorov evolution equation governing the discrete ancestral simulation of the Markov chain. By definition of the approximate conditional marginal path, the right-hand side of Eq. 125 evaluates to $\widetilde{p}^{\mathrm{MSP}}_{t_{n+1}}(\boldsymbol{u}_{t_{n+1}} \mid E)$. Therefore:

$$\boldsymbol{u}_{t_{n+1}} \sim \widetilde{p}^{\mathrm{MSP}}_{t_{n+1}}(\cdot \mid E). \tag{126}$$

By mathematical induction, this marginal identity holds for all $n \in \{0, 1, \ldots, N\}$. This completes the proof that MSP-ProFlow realizes an exact ancestral sampling chain under the Gaussian surrogate model family. □

Proof of Corollary 4.

Proof. Consider the linear observation equation $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{u}_1 + \boldsymbol{\eta}$, where $\boldsymbol{H} \in \mathbb{R}^{m\times D}$ is the measurement operator and $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2_{\mathrm{obs}}\boldsymbol{I}_m)$. The observational likelihood is given by the Gaussian density:

$$p(\boldsymbol{y} \mid \boldsymbol{u}_1) = \left(2\pi\sigma^2_{\mathrm{obs}}\right)^{-m/2} \exp\left(-\frac{1}{2\sigma^2_{\mathrm{obs}}}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{u}_1\|_2^2\right). \tag{127}$$

Under Assumption 1, the MSP local prior surrogate evaluated at the bridge pair $(\boldsymbol{u}_t, \boldsymbol{u}_{0,t})$ is:

$$p(\boldsymbol{u}_1 \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}) = \left(2\pi\sigma_t^2\right)^{-D/2} \exp\left(-\frac{1}{2\sigma_t^2}\left\|\boldsymbol{u}_1 - \boldsymbol{u}^{\mathrm{ref}}_{1,t}\right\|_2^2\right). \tag{128}$$

By Bayes' theorem, the log-posterior density of the terminal state conditioned on the observation $\boldsymbol{y}$ expands as:

$$\begin{aligned}\log \widetilde{p}^{\mathrm{MSP}}_t(\boldsymbol{u}_1 \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, \boldsymbol{y}) &= \log p(\boldsymbol{u}_1 \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}) + \log p(\boldsymbol{y} \mid \boldsymbol{u}_1) + C_0 \\ &= -\frac{1}{2\sigma_t^2}\left\|\boldsymbol{u}_1 - \boldsymbol{u}^{\mathrm{ref}}_{1,t}\right\|_2^2 - \frac{1}{2\sigma^2_{\mathrm{obs}}}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{u}_1\|_2^2 + C_1,\end{aligned} \tag{129}$$

where C_0, C_1 are constants independent of $\boldsymbol{u}_1$. Expanding the squared Euclidean norms:

$$\left\|\boldsymbol{u}_1 - \boldsymbol{u}^{\mathrm{ref}}_{1,t}\right\|_2^2 = \boldsymbol{u}_1^\top\boldsymbol{u}_1 - 2\boldsymbol{u}_1^\top\boldsymbol{u}^{\mathrm{ref}}_{1,t} + \left\|\boldsymbol{u}^{\mathrm{ref}}_{1,t}\right\|_2^2, \tag{130}$$

$$\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{u}_1\|_2^2 = \boldsymbol{u}_1^\top\boldsymbol{H}^\top\boldsymbol{H}\boldsymbol{u}_1 - 2\boldsymbol{u}_1^\top\boldsymbol{H}^\top\boldsymbol{y} + \|\boldsymbol{y}\|_2^2. \tag{131}$$

Substituting these expansions into the log-posterior and grouping terms with respect to the powers of $\boldsymbol{u}_1$:

$$\begin{aligned}\log \widetilde{p}^{\mathrm{MSP}}_t(\boldsymbol{u}_1 \mid \boldsymbol{u}_t, \boldsymbol{u}_{0,t}, \boldsymbol{y}) = &-\frac{1}{2}\boldsymbol{u}_1^\top\left[\sigma_t^{-2}\boldsymbol{I}_D + \sigma^{-2}_{\mathrm{obs}}\boldsymbol{H}^\top\boldsymbol{H}\right]\boldsymbol{u}_1 \\ &+ \boldsymbol{u}_1^\top\left[\sigma_t^{-2}\boldsymbol{u}^{\mathrm{ref}}_{1,t} + \sigma^{-2}_{\mathrm{obs}}\boldsymbol{H}^\top\boldsymbol{y}\right] + C_2,\end{aligned} \tag{132}$$

where C_2 absorbs all terms independent of $\boldsymbol{u}_1$. Now, recall the general form of a multivariate Gaussian log-density $\log\mathcal{N}(\boldsymbol{u}_1; \boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$\log\mathcal{N}(\boldsymbol{u}_1; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{1}{2}\boldsymbol{u}_1^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_1 + \boldsymbol{u}_1^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\log\det(2\pi\boldsymbol{\Sigma}). \tag{133}$$

Comparing the quadratic coefficients of $\boldsymbol{u}_1$ directly establishes the inverse covariance (precision) matrix:

$$\left(\boldsymbol{\Sigma}_t^{\mathrm{MSP}}\right)^{-1} = \sigma_t^{-2}\boldsymbol{I}_D + \sigma^{-2}_{\mathrm{obs}}\boldsymbol{H}^\top\boldsymbol{H}. \tag{134}$$

Notice that the quadratic coefficient depends exclusively on the noise variances $(\sigma_t^2, \sigma^2_{\mathrm{obs}})$ and the measurement matrix $\boldsymbol{H}$, and contains no functional dependence on the prior center. Consequently, evaluating the uncorrected baseline ProFlow under prior center $\hat{\boldsymbol{u}}_1$ yields the identical quadratic matrix:

$$\left(\boldsymbol{\Sigma}_t^{\mathrm{ProFlow}}\right)^{-1} = \sigma_t^{-2}\boldsymbol{I}_D + \sigma^{-2}_{\mathrm{obs}}\boldsymbol{H}^\top\boldsymbol{H} \equiv \left(\boldsymbol{\Sigma}_t^{\mathrm{MSP}}\right)^{-1}. \tag{135}$$

Since $\sigma_t^{-2}\boldsymbol{I}_D \succ 0$ and $\sigma^{-2}_{\mathrm{obs}}\boldsymbol{H}^\top\boldsymbol{H} \succeq 0$, the precision matrix is strictly positive definite and invertible. Inverting both sides yields the exact covariance identity in Eq. 23:

$$\boldsymbol{\Sigma}_t^{\mathrm{MSP}} = \left(\sigma_t^{-2}\boldsymbol{I}_D + \sigma^{-2}_{\mathrm{obs}}\boldsymbol{H}^\top\boldsymbol{H}\right)^{-1} \equiv \boldsymbol{\Sigma}_t^{\mathrm{ProFlow}}. \tag{136}$$

Next, matching the linear coefficients of $\boldsymbol{u}_1$:

$$\left(\boldsymbol{\Sigma}_t^{\mathrm{MSP}}\right)^{-1} \boldsymbol{\mu}_t^{\mathrm{MSP}} = \sigma_t^{-2} \boldsymbol{u}_{1,t}^{\mathrm{ref}} + \sigma_{\mathrm{obs}}^{-2} \boldsymbol{H}^\top \boldsymbol{y}. \tag{137}$$

Multiplying both sides by the covariance matrix $\boldsymbol{\Sigma}_t^{\mathrm{MSP}}$ explicitly gives the posterior mean:

$$\boldsymbol{\mu}_t^{\mathrm{MSP}} = \boldsymbol{\Sigma}_t^{\mathrm{MSP}} \left(\sigma_t^{-2} \boldsymbol{u}_{1,t}^{\mathrm{ref}} + \sigma_{\mathrm{obs}}^{-2} \boldsymbol{H}^\top \boldsymbol{y}\right). \tag{138}$$

This establishes Eq. 24 and confirms that the only discrepancy between the baseline and MSP is the substitution of $\hat{\boldsymbol{u}}_1$ by $\boldsymbol{u}_{1,t}^{\mathrm{ref}}$ in the linear vector term. This completes the proof. □

D.7 Proof of the PCFM penalty geometry

Proof of Proposition 8.

Proof. By the first-order necessary optimality condition for the unconstrained minimization in Eq. 26:

$$\nabla \Psi(\boldsymbol{u}_{\tau'}) = (\boldsymbol{u}_{\tau'} - \hat{\boldsymbol{u}}_{\tau'}) + \lambda \left[DT(\boldsymbol{u}_{\tau'})\right]^\top \left[J_h(T(\boldsymbol{u}_{\tau'}))\right]^\top h(T(\boldsymbol{u}_{\tau'})) = \mathbf{0}. \tag{139}$$

Recalling that $\delta \boldsymbol{u}_\lambda \coloneqq \boldsymbol{u}_{\tau'} - \hat{\boldsymbol{u}}_{\tau'}$ and substituting the composite constraint Jacobian definition $\mathcal{J}(\boldsymbol{u}_{\tau'}) = J_h(T(\boldsymbol{u}_{\tau'}))DT(\boldsymbol{u}_{\tau'})$, we obtain:

$$\delta \boldsymbol{u}_\lambda = -\lambda \mathcal{J}(\boldsymbol{u}_{\tau'})^\top h(T(\boldsymbol{u}_{\tau'})). \tag{140}$$

Hence:

$$\delta \boldsymbol{u}_\lambda \in \mathrm{range}\big(\mathcal{J}(\boldsymbol{u}_{\tau'})^\top\big) = \big(\ker \mathcal{J}(\boldsymbol{u}_{\tau'})\big)^\perp. \tag{141}$$

To obtain the first-order resolvent representation, we linearize the constraint residual around $\hat{\boldsymbol{u}}_{\tau'}$. Since h is C^2-smooth, Taylor's theorem yields:

$$h(T(\boldsymbol{u}_{\tau'})) = h(T(\hat{\boldsymbol{u}}_{\tau'})) + \mathcal{J}(\hat{\boldsymbol{u}}_{\tau'}) \delta \boldsymbol{u}_\lambda + \mathcal{O}\left(\|\delta \boldsymbol{u}_\lambda\|_2^2\right). \tag{142}$$

Writing $\mathcal{J} \equiv \mathcal{J}(\hat{\boldsymbol{u}}_{\tau'})$ for brevity and substituting into Eq. 140 gives:

$$\delta \boldsymbol{u}_\lambda = -\lambda \mathcal{J}^\top \left[h(T(\hat{\boldsymbol{u}}_{\tau'})) + \mathcal{J} \delta \boldsymbol{u}_\lambda\right] + \mathcal{O}\left(\|\delta \boldsymbol{u}_\lambda\|_2^2\right). \tag{143}$$

Rearranging terms with respect to $\delta \boldsymbol{u}_\lambda$:

$$\left(\boldsymbol{I}_D + \lambda \mathcal{J}^\top \mathcal{J}\right) \delta \boldsymbol{u}_\lambda = -\lambda \mathcal{J}^\top h(T(\hat{\boldsymbol{u}}_{\tau'})) + \mathcal{O}\left(\|\delta \boldsymbol{u}_\lambda\|_2^2\right). \tag{144}$$

Since $\mathcal{J}^\top \mathcal{J} \succeq 0$ is positive semi-definite, the matrix $\boldsymbol{I}_D + \lambda \mathcal{J}^\top \mathcal{J}$ is strictly positive definite and invertible with $\left\|(\boldsymbol{I}_D + \lambda \mathcal{J}^\top \mathcal{J})^{-1}\right\|_2 \leq 1$ for all $\lambda > 0$. Taking the Euclidean norm on both sides of Eq. 144 directly establishes the first-order bound on the displacement:

$$\|\delta \boldsymbol{u}_\lambda\|_2 = \mathcal{O}\left(\|h(T(\hat{\boldsymbol{u}}_{\tau'}))\|_2\right). \tag{145}$$

Consequently, the second-order remainder satisfies $\mathcal{O}(\|\delta \boldsymbol{u}_\lambda\|_2^2) = \mathcal{O}(\|h(T(\hat{\boldsymbol{u}}_{\tau'}))\|_2^2)$. Finally, applying the fundamental operator identity:

$$(\boldsymbol{I}_D + \lambda \mathcal{J}^\top \mathcal{J})^{-1} \mathcal{J}^\top = \mathcal{J}^\top (\boldsymbol{I}_m + \lambda \mathcal{J} \mathcal{J}^\top)^{-1} \tag{146}$$

to Eq. 144 yields:

$$\delta \boldsymbol{u}_\lambda = -\lambda \mathcal{J}^\top \left(\boldsymbol{I}_m + \lambda \mathcal{J} \mathcal{J}^\top\right)^{-1} h(T(\hat{\boldsymbol{u}}_{\tau'})) + \mathcal{O}\left(\|h(T(\hat{\boldsymbol{u}}_{\tau'}))\|_2^2\right). \tag{147}$$

This completes the proof. □

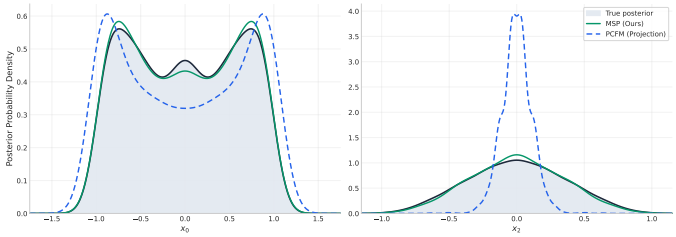


Figure 2: Posterior marginals x_0 and x_2 for the 4D benchmark

Metric	MSP	PCFM
MMSE ($\times 10^{-3}$)	**0.00**	**0.00**
SMSE ($\times 10^{-3}$)	**0.215**	37.089
RCFE	**4.96%**	43.63%

Table 3: Quantitative metrics for the 4D spherical experiment. RCFE: relative covariance Frobenius error.

E More Experiments

E.1 Posterior Measure Separation Benchmark

To isolate distributional distortion from mean displacement, we construct a 4D spherical generative system ($S^3 \subset \mathbb{R}^4$) targeting a centrally symmetric Bingham distribution. By employing antipodal sample pairing, the empirical means of the ground truth, MSP, and PCFM are strictly fixed at zero under identical rotational dynamics and projection. We report second-order moments and covariance errors in Table 3, with marginal distributions displayed in Fig. 2 (full setup in Appendix F.4).

As shown in Table 3, PCFM still incurs severe second-order distortion despite sharing an identical zero mean with the ground truth: its SMSE is $172.7\times$ higher than that of MSP, and its relative covariance Frobenius error (RCFE) escalates from 4.96% to 43.63%. The marginal profiles in Fig. 2 illustrate the geometric manifestation of this distortion: PCFM systematically transfers probability mass from the low-variance subspace (x_2, x_3) toward the high-variance subspace (x_0, x_1). This loss of volume preservation directly corroborates the theoretical uncertainty distortion analyzed in Section 5.3.

E.2 Stress-Test on the 1D Inviscid Burgers' Benchmark

Table 4: **Quantitative audit of PCFM on the 1D Inviscid Burgers' Benchmark across varying shock locations p_{loc}.** We evaluate the representative PCFM sampler: **Full (+)** denotes the original post-hoc model with terminal-pullback interpolation, whereas **MSP (-)** represents its surplus-removed counterpart where the projection center is corrected via the MSP variational formulation. Best results between $(+)$ and $(-)$ for each metric are in **bold**.

Shock Location (p_{loc})	**MMSE** ($\times 10^{-2}$) ↓		**SMSE** ($\times 10^{-2}$) ↓		**FPD** ↓	
	Full (+)	**MSP (-)**	**Full (+)**	**MSP (-)**	**Full (+)**	**MSP (-)**
$p_{\text{loc}} = 0.40$	9.3886	**3.4492**	**3.8525**	4.3756	7.5531	**4.4648**
$p_{\text{loc}} = 0.50$	10.1190	**3.5287**	**3.2947**	3.8137	**7.3802**	9.2462
$p_{\text{loc}} = 0.60$	7.1697	**2.7216**	**1.9155**	3.1065	**6.7004**	7.2080
$p_{\text{loc}} = 0.70$	**1.3982**	2.5269	**1.0977**	2.8029	**3.7650**	6.3083
$p_{\text{loc}} = 0.80$	**0.0941**	4.4741	**0.1052**	3.0748	**1.1003**	8.5781

To demonstrate that even if PCFM performs well on second-order metrics or distributional metrics, it cannot guarantee that the direction of the surplus velocity field aligns with the true posterior direction, we evaluate the 1D Inviscid Burgers' benchmark under a controlled parameter sweep across shock locations $p_{\text{loc}} \in [0.40, 0.80]$. The detailed problem setup and numerical configurations are documented in Appendix F.5. The quantitative results are reported in Table 4.

As shown in the table, even under parameter configurations where PCFM achieves lower SMSE and FPD than MSP, it fails to consistently provide lower first-moment reconstruction error (MMSE). Instead, the performance landscape exhibits pronounced task dependence: MSP achieves lower MMSE at $p_{\text{loc}} = 0.40, 0.50$, and 0.60, whereas PCFM achieves lower MMSE at $p_{\text{loc}} = 0.70$ and 0.80. Furthermore, the MMSE error exhibits pronounced sensitivity to the task parameter, with a sharp transition between $p_{\text{loc}} = 0.60$ and 0.70.

This is because the tangential surplus is not inherently task-directed, its directional alignment with the target posterior displacement varies substantially as the shock location shifts. Because the first-moment reconstruction discrepancy is directly modulated by the inner product between the surplus displacement and the target displacement, small variations in task configuration can induce substantial shifts in the surplus contribution. Consequently, favorable results observed on specific individual metrics should not be interpreted as principled posterior-aware optimization, but rather as incidental effects governed by this uncontrolled dynamical surplus.

F Experiments Setup

F.1 8D Nonlinear-Decay Benchmark

We construct a synthetic eight-dimensional nonlinear dynamical system $\boldsymbol{u} \in \mathbb{R}^8$ with a 4D constrained subspace and a 4D tangent subspace. The first four coordinates are fixed by an affine boundary constraint $H_{8\mathrm{D}}(\boldsymbol{u}) \coloneqq \boldsymbol{u}_{1:4} - \boldsymbol{b}_{1:4} = \mathbf{0}$, defining the feasible affine subspace $\mathcal{M}_{8\mathrm{D}} \coloneqq \{\boldsymbol{u} \in \mathbb{R}^8 \mid \boldsymbol{u}_{1:4} = \boldsymbol{b}_{1:4}\}$. The remaining four unconstrained coordinates evolve according to shifted nonlinear cubic decay dynamics:

$$\frac{du_i}{dt} = -\lambda_i(u_i - b_i) - \beta_i(u_i - b_i)^3, \qquad i = 5, \dots, 8, \tag{148}$$

subject to $\boldsymbol{u}_{1:4}(t) \equiv \boldsymbol{b}_{1:4}$. The prescribed boundary and equilibrium parameters are specified as $\boldsymbol{b} = (0.8, -0.5, 0.3, -0.2, 1.0, -0.8, 0.6, -0.4)^\top$, decay rates $\boldsymbol{\lambda} = (0.8, 1.2, 1.6, 2.0, 1.0, 1.5, 2.0, 2.5)^\top$, and cubic coefficients $\boldsymbol{\beta} = (0.8, 0.8, 0.8, 0.8, 1.0, 1.0, 1.0, 1.0)^\top$. The corresponding unconstrained reference vector field used by the baseline samplers is given analytically by $v_{\mathrm{VFM},i}(\boldsymbol{u}) = -\lambda_i u_i - \beta_i u_i^3$. Along the constrained ground-truth flow, the transformed variables $z_i(t) \coloneqq u_i(t) - b_i$ admit the exact closed-form solution:

$$z(t) = \frac{z_0 e^{-\lambda t}}{\sqrt{1 + \frac{\beta}{\lambda} z_0^2 \left(1 - e^{-2\lambda t}\right)}}, \tag{149}$$

enabling the analytical evaluation of ground-truth moments without numerical ODE integration error.

Controlled Task Parameterization. To introduce controlled directional variation, we define a one-parameter family of initial-condition means indexed by $\alpha \in [-\pi, \pi)$. The constrained coordinates remain fixed at $\boldsymbol{b}_{1:4}$, while the initial mean over the tangent subspace is modulated as:

$$\boldsymbol{\mu}_{\mathrm{IC}}^{\tan}(\alpha) = \frac{0.8}{\sqrt{2}} \begin{bmatrix} \cos\alpha \\ \sin\alpha \\ \cos(2\alpha) \\ \sin(2\alpha) \end{bmatrix}. \tag{150}$$

For each task configuration α, initial source states are generated by injecting independent Gaussian noise with standard deviation 0.18 into the four tangent coordinates, while keeping the constrained coordinates strictly at $\boldsymbol{b}_{1:4}$. We evaluate 41 uniformly spaced task angles over $[-\pi, \pi)$ and sample 1024 independent source trajectories per task.

Analytical Ground-Truth Moments. The ground-truth terminal moments are computed analytically for each task: for the four constrained coordinates, the terminal mean is identically $\boldsymbol{b}_{1:4}$ with zero variance; for each tangent coordinate, the initial Gaussian distribution is propagated through the closed-form decay solution, with the resulting first and second moments evaluated via 64-point Gauss–Hermite quadrature.

Sampling and Evaluation Protocol. All evaluated samplers perform $N = 100$ uniform Euler integration steps over $t \in [0, 1]$ ($\Delta t = 0.01$). At each step, PCFM extrapolates the terminal state, projects it onto the affine subspace by resetting the first four coordinates to $\boldsymbol{b}_{1:4}$, and performs linear pullback from the source state $\boldsymbol{u}_0$. MSP instead evaluates the locally minimal unconstrained next-state update to construct the rectified projection center $\boldsymbol{u}_1^{\mathrm{ref}}$ before applying the identical affine projection. Performance is assessed via ensemble mean error (MMSE) and standard-deviation error (SMSE) against the analytical ground truth, alongside the global directional alignment $\cos(\Theta)$ between the surplus displacement and the target shift.

F.2 Darcy Flow Inverse Benchmark

To audit distributional distortion in a realistic PDE inverse problem, we implement the steady-state Darcy flow benchmark following the formulation and numerical protocol established by Xu et al. (2026).

Governing PDE and Parameter Representation. The forward physical system describes single-phase fluid flow in a porous medium along a bounded spatial domain $\Omega \coloneqq [0,1]$, governed by the elliptic equation:

$$-\frac{d}{dx}\left(\kappa(x)\frac{du(x)}{dx}\right) = 1, \qquad x \in [0,1], \tag{151}$$

subject to homogeneous Dirichlet boundary conditions $u(0) = u(1) = 0$, where $u(x)$ is the pressure head and $\kappa(x) > 0$ is the hydraulic permeability field. The domain is discretized into 50 uniform subintervals with mesh spacing $\Delta x = 0.02$, yielding $n = 49$ interior spatial nodes, and is solved numerically using standard finite differences via the tridiagonal Thomas algorithm. The unknown log-permeability field is parameterized via a truncated Fourier-type cosine expansion $\log \kappa(x) = \sum_{k=1}^{D} \theta_k S_k \cos(k\pi x)$ with decay coefficients $S_k \coloneqq 1.2/k$, where $\boldsymbol{\theta} \coloneqq (\theta_1, \ldots, \theta_D)^\top \in \mathbb{R}^D$ denotes the unknown latent state vector. We evaluate two parameter dimensions: $D = 8$ (evaluated over three independently generated problem instances/seeds) and $D = 16$, with a standard Gaussian generative prior $\boldsymbol{\theta} \sim \pi = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_D)$.

Observation Operator and Feasible Manifold. Sparse measurement data $\boldsymbol{y} \in \mathbb{R}^m$ are acquired at $m = 3$ interior observation coordinates $x \in \{0.25, 0.50, 0.75\}$ via a linear point-evaluation operator H: $\boldsymbol{y} = H(G(\boldsymbol{\theta}^\star)) + \boldsymbol{\eta}$, where $G : \mathbb{R}^D \to \mathbb{R}^n$ denotes the forward PDE solution mapping, $\boldsymbol{\theta}^\star \sim \pi$ is the ground-truth coefficient profile, and $\boldsymbol{\eta}$ represents small measurement noise. The physical equation and observational matching jointly define the hard-constraint operator $c(\boldsymbol{\theta}) \coloneqq H(G(\boldsymbol{\theta})) - \boldsymbol{y} = \mathbf{0}$, defining a codimension-3 embedded manifold $\mathcal{M} \coloneqq \{\boldsymbol{\theta} \in \mathbb{R}^D \mid c(\boldsymbol{\theta}) = \mathbf{0}\}$. Because the mapping from permeability to interior pressure is highly nonlinear, the local constraint sensitivity metric $\sqrt{\det(J(\boldsymbol{\theta})J(\boldsymbol{\theta})^\top)}$ varies by more than an order of magnitude across the posterior support, providing an ideal testbed for examining geometric and measure-level distortions.

Gold-Standard Arbiter and Sampling Protocol. Following Xu et al. (2026), statistical ground-truth evaluations are anchored to an *i.i.d. residual-band rejection arbiter*, accepting prior samples satisfying $\|c(\boldsymbol{\theta})\|_2 < \varepsilon$ with $\varepsilon = 0.003$. Over 5.4×10^8 raw prior candidates are evaluated to harvest 30,000–60,000 independent accepted samples per instance. The intrinsic sampling-noise floor is computed from the two-sample $\overline{W}_1$ distance between two independent halves of the accepted ensemble (0.012 for $D = 8$ and 0.018 for $D = 16$). All generative samplers are initialized from identical prior ensembles of size $B = 8{,}000$ and integrate over $N = 100$ uniform Euler steps across $t \in [0,1]$ ($\Delta t = 0.01$). Distributional fidelity is quantified via the average per-coordinate 1-Wasserstein distance ($\overline{W}_1$) to the rejection arbiter and its ratio to the sampling-noise floor ($\times$floor), as formulated in Eq. 166.

F.3 1D Linear Heat Equation Benchmark

We follow the Heat Equation benchmark of PCFM (Utkarsh et al., 2026) across all physical parameters, numerical schemes, and constraint formulations, extending it across multiple task configurations.

Governing PDE and Analytical Solution. We consider the one-dimensional linear heat equation with periodic boundary conditions on the spatial domain $\Omega \coloneqq [0, 2\pi]$ over time $t \in [0,1]$:

$$u_t = \alpha u_{xx}, \qquad x \in [0, 2\pi], \quad t \in [0,1], \tag{152}$$

subject to periodic boundaries $u(0,t) = u(2\pi, t)$ and sinusoidal initial conditions $u(x,0) = \sin(x + \phi)$. The exact analytical solution is given by $u(x,t) = e^{-\alpha t}\sin(x+\phi)$. To train the unconstrained functional flow matching prior, analytical solutions are sampled on a uniform $(N_x \times N_t) = 100 \times 100$ spatiotemporal grid. Following the official PCFM setup, the pretraining dataset comprises 5000 trajectories generated by independently drawing the thermal diffusivity $\alpha \sim \mathcal{U}(1.0, 5.0)$ and initial phase $\phi \sim \mathcal{U}(0.0, \pi)$.

Physical Constraint Formulation. The constrained inverse problem enforces both the prescribed initial condition and global mass conservation over time. The composite hard-constraint operator is defined as:

$$H_{\text{Heat}}(\boldsymbol{u}) \coloneqq \begin{bmatrix} u(x,0) - u_{\text{IC}}(x) \\ \displaystyle\int_0^{2\pi} u(x,t)\,dx - \int_0^{2\pi} u(x,0)\,dx \end{bmatrix} = \mathbf{0}, \qquad \forall t \in [0,1]. \tag{153}$$

Since the sinusoidal initial profile satisfies $\int_0^{2\pi} \sin(x+\phi)\,dx \equiv 0$ for any phase ϕ, the total conserved mass is zero, and the second constraint simplifies to $\int_0^{2\pi} u(x,t)\,dx = 0$ throughout the trajectory.

Task Configurations and Evaluation Protocol. While PCFM originally evaluates only the single default phase $\phi = \pi/4$, we sweep across five representative initial configurations to evaluate stability under task variation:

$$\phi \in \left\{ \frac{\pi}{4},\ \frac{\pi}{2},\ \frac{3\pi}{4},\ \pi,\ \frac{5\pi}{4} \right\}, \tag{154}$$

where $\phi = 5\pi/4$ lies strictly outside the training support $[0, \pi)$ and serves as an out-of-distribution (OOD) testbed. For each phase configuration, all evaluated sampling methods perform $N = 100$ integration steps and generate an ensemble of 512 independent trajectories. Reconstruction accuracy (MMSE), predictive uncertainty calibration (SMSE), and distributional fidelity (FPD) are benchmarked against an equivalent ground-truth ensemble of 512 analytical solutions sampled over $\alpha \sim \mathcal{U}(1.0, 5.0)$.

F.4 4D Spherical-Constrained System

Manifold Constraint and Bingham Target Distribution. We construct a four-dimensional generative system constrained to the unit hypersphere $\mathbb{S}^3 \subset \mathbb{R}^4$, defined by the quadratic equality constraint $H_{\mathbb{S}^3}(\boldsymbol{u}) \coloneqq \|\boldsymbol{u}\|_2^2 - 1 = 0$, with the feasible manifold $\mathcal{M}_{\mathbb{S}^3} \coloneqq \{\boldsymbol{u} \in \mathbb{R}^4 \mid \|\boldsymbol{u}\|_2 = 1\}$. The metric projection onto $\mathbb{S}^3$ is given in closed form by radial normalization $\Pi_{\mathbb{S}^3}(\boldsymbol{u}) = \boldsymbol{u}/\|\boldsymbol{u}\|_2$. The target density is a centrally symmetric Bingham distribution on $\mathbb{S}^3$ with concentration matrix:

$$\boldsymbol{A} = \text{diag}(\kappa,\ \kappa,\ -\kappa,\ -\kappa), \qquad \text{with} \quad \kappa = 1.6. \tag{155}$$

Target reference samples are obtained via rejection sampling on the unit sphere. To strictly isolate second-order distributional deformation from first-moment drift, each drawn sample $\boldsymbol{x}$ is paired with its antipodal counterpart $-\boldsymbol{x}$, enforcing an exact zero empirical mean across all coordinates. The reference ensemble consists of 2048 samples generated with random seed 42.

Time-Varying Rotational Dynamics. The underlying reference dynamics are governed by an uncoupled, time-dependent rotational vector field:

$$\begin{aligned} \dot{x}_0 &= -\omega_1(t) x_1, \qquad \dot{x}_1 = \omega_1(t) x_0, \\ \dot{x}_2 &= -\omega_2(t) x_3, \qquad \dot{x}_3 = \omega_2(t) x_2, \end{aligned} \tag{156}$$

driven by angular frequencies $\omega_1(t) = 2\pi A_1 \cos(2\pi t)$ and $\omega_2(t) = 2\pi A_2 \cos(2\pi t)$, with amplitudes $A_1 = 1.5$ and $A_2 = 0.5$. The system consists of two independent harmonic rotations acting on the coordinate planes (x_0, x_1) and (x_2, x_3), respectively. Trajectories are evolved across $t \in [0, 1]$ over $N = 100$ uniform Euler steps ($\Delta t = 0.01$).

F.5 Stress-Test on Inviscid Burgers' Benchmark

Governing Equation and Godunov Finite-Volume Solver. We consider the one-dimensional inviscid Burgers' equation on the spatiotemporal domain $\Omega \times [0, 1] \coloneqq [0, 1] \times [0, 1]$:

$$u_t + \frac{1}{2}(u^2)_x = 0, \qquad x \in [0,1], \quad t \in [0,1], \tag{157}$$

with the initial condition defined by a smoothed step profile parameterized by the shock location p_{loc}:

$$u(x,0) = \frac{1}{1 + \exp\left(\frac{x - p_{\text{loc}}}{\epsilon}\right)}, \qquad \text{with} \quad \epsilon = 0.02. \tag{158}$$

We impose a constant Dirichlet boundary condition on the left, $u(0,t) = u_{\mathrm{bc}}$, and a zero-gradient Neumann condition on the right, $u_x(1,t) = 0$. Ground-truth solutions are computed on a uniform $(N_x+1) \times (N_t+1) = 101 \times 101$ spatiotemporal grid using an explicit Godunov finite-volume solver with the numerical flux:

$$\mathcal{F}(u_L, u_R) = \begin{cases} \min\left(\frac{1}{2}u_L^2,\ \frac{1}{2}u_R^2\right), & \text{if } u_L \le u_R, \\ \frac{1}{2}u_L^2, & \text{if } u_L > u_R \text{ and } \frac{u_L+u_R}{2} > 0, \\ \frac{1}{2}u_R^2, & \text{if } u_L > u_R \text{ and } \frac{u_L+u_R}{2} \le 0. \end{cases} \tag{159}$$

To pretrain the unconstrained functional flow matching prior, parameters are independently drawn from uniform distributions: $p_{\mathrm{loc}} \sim \mathcal{U}(0.2, 0.8)$ and $u_{\mathrm{bc}} \sim \mathcal{U}(0.0, 1.0)$. We sample $N_{\mathrm{ic}} = 80$ initial conditions and $N_{\mathrm{bc}} = 80$ boundary values, taking their Cartesian product to construct 6400 numerical solutions for training, alongside an independent validation set of 900 solutions ($N_{\mathrm{ic}} = 30, N_{\mathrm{bc}} = 30$) generated with a different random seed.

Composite Physical Constraint Formulation. For zero-shot constrained generation with a prescribed initial condition, the generated solution must satisfy the initial/boundary profile alongside the nonlinear mass-conservation relation $\frac{d}{dt}\int_0^1 u(x,t)\,dx = -\frac{1}{2}u(1,t)^2 + \frac{1}{2}u(0,t)^2 = 0$. Following PCFM (Utkarsh et al., 2026), we enforce the composite physical constraint operator:

$$\boldsymbol{H}_{\text{Burgers-IC}}(\boldsymbol{u}) \coloneqq \begin{bmatrix} u(x,0) - u_{\mathrm{IC}}(x) \\ \mathcal{R}^{(k)}_{\mathrm{Flux}}(\boldsymbol{u}) \\ \displaystyle\int_0^1 u(x,t)\,dx - \int_0^1 u(x,0)\,dx \end{bmatrix} = \mathbf{0}, \tag{160}$$

which unifies (i) the exact initial step profile $u_{\mathrm{IC}}(x)$, (ii) $k = 5$ unrolled discrete local Godunov flux updates $\mathcal{R}^{(k)}_{\mathrm{Flux}}$, and (iii) global temporal mass conservation.

Shock Location Parameter Sweep and Evaluation. To diagnose how the surplus velocity behaves across varying physical regimes, we conduct a controlled parameter sweep over five representative shock center locations:

$$p_{\mathrm{loc}} \in \{0.40,\ 0.50,\ 0.60,\ 0.70,\ 0.80\}. \tag{161}$$

For each fixed initial location p_{loc}, the unconstrained boundary condition is marginalized by evaluating an ensemble of 512 ground-truth solutions with randomly drawn boundary values $u_{\mathrm{bc}} \sim \mathcal{U}(0.0, 1.0)$. All evaluated flow matching samplers perform $N = 100$ integration steps to generate 512 independent trajectories. Reconstruction accuracy (MMSE), predictive uncertainty (SMSE), and distributional fidelity (FPD) are benchmarked against this ground-truth ensemble.

G Basic Information of Models and Metrics

G.1 Baseline and Hyperparameters

ECI. ECI sampling (Cheng et al., 2025) is a zero-shot, gradient-free framework for enforcing hard constraints during flow matching sampling. At each sampling step, ECI alternates between extrapolation, correction, and interpolation. Given the current state $\boldsymbol{u}_t$, the model first extrapolates a terminal prediction $\boldsymbol{u}_1 = \boldsymbol{u}_t + (1-t)\boldsymbol{v}_\theta(\boldsymbol{u}_t, t)$, then applies a constraint-specific correction operator $\mathcal{C}(\boldsymbol{u}_1, \mathcal{G})$ to obtain a constraint-satisfying terminal state, and finally interpolates the corrected state back to an intermediate time according to $\boldsymbol{u}_{t'} = (1-t')\boldsymbol{u}_0 + t'\boldsymbol{u}_1$. The correction operator is designed for different classes of constraints, including value constraints and conservation laws, and can be implemented without gradient backpropagation. In our experimental evaluations, following the standard setup recommended in the original paper, we set the number of recursive mixing iterations to $n_{\mathrm{mix}} = 3$ per Euler step to facilitate information propagation across constrained boundaries. ECI requires no retraining of the pretrained flow model and applies the constraint directly during sampling.

PCFM. PCFM (Utkarsh et al., 2026) is a zero-shot framework for enforcing general nonlinear equality constraints in pretrained flow matching models. At each sampling step, PCFM first performs a forward shooting operation from the current state to obtain a terminal prediction $\boldsymbol{u}_1$, and then applies a Gauss–Newton projection onto the constraint manifold. The projected terminal state is

subsequently propagated back to the next sampling time using the optimal-transport displacement interpolant. Because the pulled-back state may still exhibit constraint residuals due to finite-step discretization and nonlinear constraints, PCFM further introduces an optional relaxed constraint correction by solving a penalized optimization problem around the OT pullback state. Following the standard protocol in the original paper (where $\lambda = 0$ is adopted for standard benchmark comparisons to isolate intrinsic projection properties and preverse sampling efficiency), we set the number of inner-loop penalty optimization iterations to 0, relying on the standard Gauss–Newton manifold projection combined with a final terminal projection to ensure strict constraint satisfaction.

ProFlow. ProFlow (Yu et al., 2026) is a zero-shot proximal guidance framework for physics-consistent sampling with sparse observations. At each sampling step, ProFlow first uses the pretrained flow model to extrapolate a candidate terminal state $\hat{\boldsymbol{u}}_1 = \boldsymbol{u}_t + (1-t)\boldsymbol{v}_\theta(\boldsymbol{u}_t, t)$. It then refines the terminal prediction through a proximal optimization problem that balances proximity to $\hat{\boldsymbol{u}}_1$ with physical and observational consistency, producing a refined terminal state $\boldsymbol{u}_1$. The refined state is subsequently interpolated to the next sampling time using a freshly sampled reference $\boldsymbol{\varepsilon} \sim \mu_0$, i.e., $\boldsymbol{u}_{t_{n+1}} = (1-t_{n+1})\boldsymbol{\varepsilon} + t_{n+1}\boldsymbol{u}_1$. This two-step terminal optimization–interpolation procedure is repeated throughout sampling and does not require task-specific retraining of the generative prior. ProFlow further provides a Bayesian interpretation of the terminal proximal optimization as a local MAP update and is designed to jointly enforce physical constraints and sparse observations while maintaining compatibility with the pretrained flow probability path.

CoCoS. CoCoS (Xu et al., 2026) is a measure-aware constrained sampler that directly targets the co-area-corrected posterior on the constraint manifold. At each sampling step, CoCoS first constructs the tangent-space projector, proposes an isotropic tangent perturbation, and then projects the proposal back onto the constraint manifold along the normal directions using a Gauss–Newton solve. A Metropolis correction with the co-area (Fixman) factor and a reverse-projection check ensures the target posterior is invariant. For the 1D Darcy benchmark, we follow the original experimental configuration with $B = 8000$ walkers, tangent step size $s = 0.10$, 5000 sampling steps, 2000 burn-in steps, and thinning factor 10, using $(G + 10^{-6}I)^{-1}$ for the projection and a reverse-projection tolerance of 10^{-6}.

G.2 Common Evaluation Metrics

We use four standard metrics to evaluate reconstruction accuracy and distributional fidelity. All ensemble-based metrics are computed under the same physical condition by comparing the generated and ground-truth solution ensembles.

Mean Mean-Squared Error (MMSE). MMSE measures the error of the generated ensemble mean:

$$\mathrm{MMSE} = \frac{1}{D}\,\|\boldsymbol{\mu}_{\mathrm{gen}} - \boldsymbol{\mu}_{\mathrm{GT}}\|_2^2\,, \tag{162}$$

where $\boldsymbol{\mu}_{\mathrm{gen}}$ and $\boldsymbol{\mu}_{\mathrm{GT}}$ are the pointwise means of the generated and ground-truth ensembles, respectively, and D is the number of field degrees of freedom. MMSE evaluates first-moment reconstruction accuracy.

Standard-deviation Mean-Squared Error (SMSE). SMSE measures the discrepancy between the pointwise standard-deviation fields:

$$\mathrm{SMSE} = \frac{1}{D}\,\|\boldsymbol{\sigma}_{\mathrm{gen}} - \boldsymbol{\sigma}_{\mathrm{GT}}\|_2^2\,, \tag{163}$$

where $\boldsymbol{\sigma}_{\mathrm{gen}}$ and $\boldsymbol{\sigma}_{\mathrm{GT}}$ denote the corresponding standard-deviation fields. SMSE evaluates the accuracy of predictive uncertainty at each spatial or spatiotemporal location. The MMSE and SMSE definitions follow the ensemble evaluation protocol used by PCFM. (Utkarsh et al., 2026)

Fréchet Poseidon Distance (FPD). FPD measures distributional similarity in the feature space of a pretrained Poseidon model. Let $(\boldsymbol{\mu}^f_{\mathrm{gen}}, \boldsymbol{\Sigma}^f_{\mathrm{gen}})$ and $(\boldsymbol{\mu}^f_{\mathrm{GT}}, \boldsymbol{\Sigma}^f_{\mathrm{GT}})$ denote the empirical mean and covariance of the extracted feature distributions. We compute

$$\mathrm{FPD}^2 = \left\|\boldsymbol{\mu}^f_{\mathrm{gen}} - \boldsymbol{\mu}^f_{\mathrm{GT}}\right\|_2^2 + \mathrm{Tr}\left(\boldsymbol{\Sigma}^f_{\mathrm{gen}} + \boldsymbol{\Sigma}^f_{\mathrm{GT}} - 2\left(\boldsymbol{\Sigma}^f_{\mathrm{gen}}\boldsymbol{\Sigma}^f_{\mathrm{GT}}\right)^{1/2}\right). \tag{164}$$

FPD evaluates higher-order distributional similarity beyond pointwise mean and variance by comparing the generated and ground-truth distributions in a learned feature space.

Relative Covariance Frobenius Error (RCFE). RCFE directly measures the relative discrepancy between the generated and ground-truth covariance matrices:

$$\mathrm{RCFE} = \frac{\|\boldsymbol{\Sigma}_{\mathrm{gen}} - \boldsymbol{\Sigma}_{\mathrm{GT}}\|_F}{\|\boldsymbol{\Sigma}_{\mathrm{GT}}\|_F} \times 100\%. \tag{165}$$

RCFE provides a scale-normalized measure of second-order distributional distortion and is particularly useful for isolating covariance changes when mean errors are controlled.

Average Per-Coordinate Wasserstein-1 Distance ($\overline{W}_1$). For the Darcy inverse benchmark, we evaluate distributional fidelity in the parameter space using the Wasserstein-1 distance to the gold-standard i.i.d. rejection arbiter. Following the exact evaluation protocol of Xu et al. (2026), we compute the average per-coordinate 1-Wasserstein distance:

$$\overline{W}_1 \coloneqq \frac{1}{d}\sum_{j=1}^{d} W_1\left(\hat{p}^{(j)}_{\mathrm{gen}}, \hat{p}^{(j)}_{\mathrm{arb}}\right), \tag{166}$$

where $\hat{p}^{(j)}_{\mathrm{gen}}$ and $\hat{p}^{(j)}_{\mathrm{arb}}$ denote the 1D empirical marginal distributions along the j-th coordinate for the generated ensemble and the rejection arbiter, respectively. To account for finite-sample variability, we additionally report the normalized ratio to the arbiter's two-sample noise floor ($\times$floor):

$$\text{Ratio to Floor} \coloneqq \frac{\overline{W}_1}{\overline{W}_{1,\,\mathrm{floor}}}, \tag{167}$$

where $\overline{W}_{1,\,\mathrm{floor}}$ is the noise floor evaluated between two independent, identically distributed splits drawn from the same rejection arbiter. A ratio close to $1.0\times$ indicates that the generated posterior matches the ground truth up to intrinsic Monte Carlo sampling noise.

H More Visualization

H.1 Correlation Analysis of Surplus Alignment and Performance Gaps

To provide complete empirical documentation for the correlation analysis discussed in Section 6.1, we present the scatter plots and linear regression fits relating the global surplus alignment $\cos(\Theta)$ to the first- and second-order performance gaps (ΔMMSE and ΔSMSE) in Fig. 3.

The near-perfect negative correlation between $\cos(\Theta)$ and ΔMMSE is consistent with the leading inner-product term between the surplus-induced displacement and the ground-truth target shift. The weaker but still strong correlation with ΔSMSE reflects additional higher-order distributional deformation.

H.2 Surplus Velocity Orthogonal Decomposition

To examine whether the surplus velocity field $\boldsymbol{r}_t$ contributes to physical constraint enforcement discussed in Section 6.1, we orthogonally decompose it into manifold-tangential ($\boldsymbol{P}_T\boldsymbol{r}_t$) and constraint-normal ($\boldsymbol{P}_N\boldsymbol{r}_t$) components relative to the projected state $\boldsymbol{p}_t = \Pi_{\mathcal{M}}(\hat{\boldsymbol{u}}_1)$:

$$\boldsymbol{r}_t = \boldsymbol{P}_T(\hat{\boldsymbol{u}}_1)\boldsymbol{r}_t + \boldsymbol{P}_N(\hat{\boldsymbol{u}}_1)\boldsymbol{r}_t. \tag{168}$$

As shown in Figure 4, the normal component $\|\boldsymbol{P}_N\boldsymbol{r}_t\|_2$ is negligible ($< 10^{-3}$) throughout the trajectory, while the tangential component $\|\boldsymbol{P}_T\boldsymbol{r}_t\|_2$ accounts for over 99.9% of the total surplus magnitude across all tested task configurations $\alpha \in [0, 1]$.

This empirically confirms Proposition 5 and Corollary 3: the surplus velocity does not assist in constraint satisfaction, but instead acts almost purely as an unguided in-plane lateral drift along the manifold tangent space.

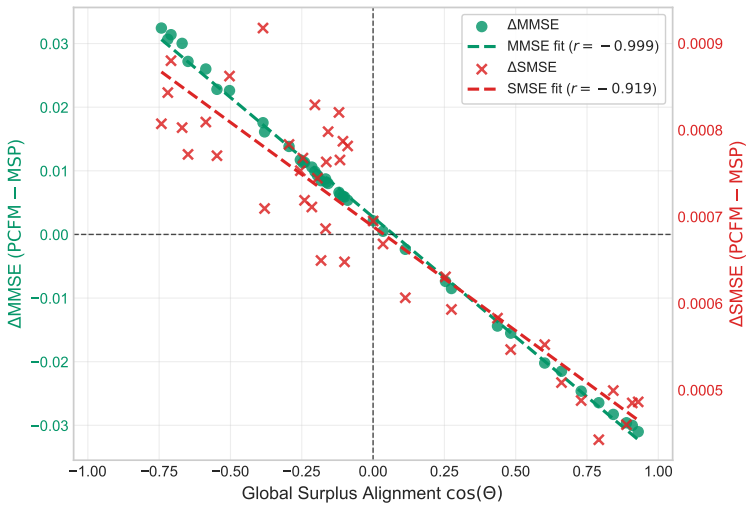


Figure 3: **Linear correlation between global surplus alignment** $\cos(\Theta)$ **and performance discrepancies.** The green circles (left axis) show the first-moment error gap $\Delta\mathrm{MMSE} := \mathrm{MMSE}_{\mathrm{PCFM}} - \mathrm{MMSE}_{\mathrm{MSP}}$ across $\alpha \in [-\pi, \pi)$, with $r = -0.999$ and $R^2 = 0.998$. The red crosses (right axis) show the corresponding second-moment error gap $\Delta\mathrm{SMSE} := \mathrm{SMSE}_{\mathrm{PCFM}} - \mathrm{SMSE}_{\mathrm{MSP}}$, with $r = -0.919$ and $R^2 = 0.845$. The horizontal and vertical dashed lines indicate zero alignment and zero performance discrepancy, respectively.

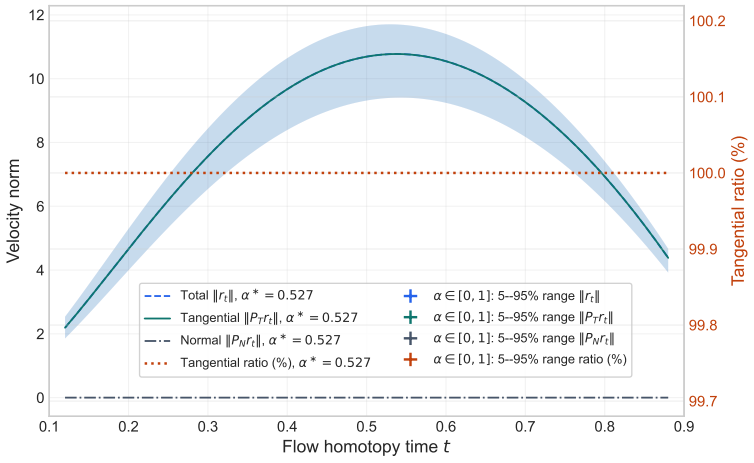


Figure 4: **Orthogonal decomposition of the surplus velocity field.** Across flow homotopy time $t \in (0, 1)$, the normal component $\|\boldsymbol{P}_N \boldsymbol{r}_t\|_2$ remains vanishingly small ($< 10^{-3}$), while the tangential component $\|\boldsymbol{P}_T \boldsymbol{r}_t\|_2$ accounts for nearly 100.0% of the total surplus magnitude. Shaded ribbons depict the 5–95% empirical range over $\alpha \in [0, 1]$, and solid curves highlight the isomean crossing root $\alpha^* = 0.527$.

H.3 Standard Deviation Comparison on the 4D Spherical Benchmark

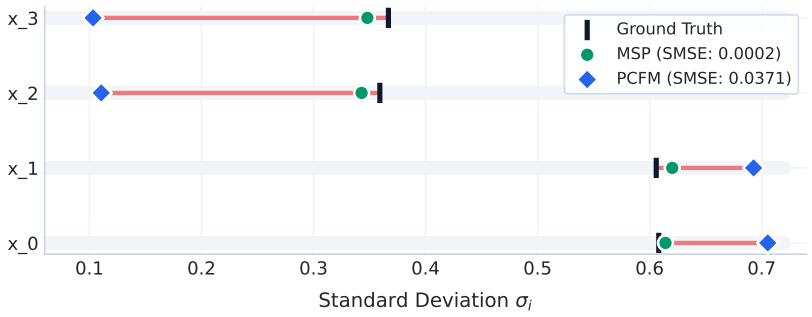


Figure 5: **Coordinate-wise standard deviation comparison on the 4D spherical benchmark.** The plot compares the standard deviation σ_i along each coordinate axis (x_0, x_1, x_2, x_3) for Ground Truth (black bars), MSP (green circles), and PCFM (blue diamonds).

To visually demonstrate the standard deviation comparison discussed in Section E.1, we plot Figure 5, which shows the coordinate-wise standard deviation (uncertainty/spread) comparison across the four dimensions (x_0, x_1, x_2, x_3) on the 4D spherical benchmark. The comparison shows that MSP matches the Ground Truth spread across all dimensions, whereas PCFM exhibits severe coordinate-wise variance distortion, expanding the spread in x_0, x_1 and contracting it in x_2, x_3.