Geometric Velocity Regularity for Flow Matching on Manifold-Concentrated Data

Shuntuo Xu[*]
KLATASDS-MOE, School of Statistics
East China Normal University
Shanghai, China 200062
oaksword98@gmail.com

Zhou Yu
KLATASDS-MOE, School of Statistics
East China Normal University
Shanghai, China 200062
zyu@stat.ecnu.edu.cn

Kenji Fukumizu[†]
The Institute of Statistical Mathematics
Tokyo, Japan 190-8562
fukumizu@ism.ac.jp

Abstract

Flow matching learns a velocity field whose ordinary differential equation transports a simple reference distribution to a data distribution. Existing theory typically controls this velocity through ambient, isotropic Lipschitz regularity, which becomes extremely pessimistic when data concentrate near a low-dimensional manifold. In this regime, standard error bounds can blow up drastically near the terminal time and fail to explain why flow matching should remain stable. We study flow matching for distributions obtained by smoothing a density on a compact manifold with small Gaussian noise. Our first result is a tube-local Lipschitz regularity theorem showing that, on the relevant neighborhood of the manifold, the population flow-matching velocity has a Jacobian bound that grows only at the inverse effective noise scale. This improves substantially over the stronger singular behavior suggested by existing ambient analyses. Our second result gives a directional description of the velocity Jacobian: tangential directions exhibit mild growth, normal directions are strongly contractive, and tangential-normal coupling is lower order. This reveals that the apparent singularity is highly structured rather than uniformly harmful. These results provide geometric foundations for sharper stability guarantees and suggest a path toward intrinsic-dimension generalization bounds for flow matching on manifold-concentrated data.

1 Introduction

Flow matching (FM) [2, 33, 35] generates samples by learning a velocity field whose ordinary differential equation (ODE) flow transports a Gaussian reference onto a target distribution. In many domains, such as images, molecular conformations, and protein structures, target data are often observed to be concentrated near a low-dimensional submanifold $\mathcal{M}$ of dimension k embedded in $\mathbb{R}^d$, with $k \leq d$ [13, 34, 45]. We consider an explicit model where the target distribution is a smooth density f on $\mathcal{M}$ convolved with isotropic Gaussian noise of scale $\sigma > 0$, written $\nu_\sigma = (f \operatorname{Vol}_{\mathcal{M}}) * \mathcal{N}(0, \sigma^2 I_d)$, where $\operatorname{Vol}_{\mathcal{M}}$ denotes the volume measure of $\mathcal{M}$. As σ shrinks toward zero, the target sharpens around $\mathcal{M}$, and the central question of this paper emerges: *how does the regularity of the FM velocity field depend on σ and on the geometry of $\mathcal{M}$?*

[*]This work was initiated and partially completed during a visit to the Institute of Statistical Mathematics, Tokyo.
[†]Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

This question is important because the foundational error bound for FM with respect to 2-Wasserstein distance [4],

$$\mathrm{W}_2(\widehat{\nu}_1, \nu_1) \le \varepsilon \, \exp\Big(\int_0^1 L_t \, \mathrm{d}t\Big),$$

converts the velocity estimation error ε into distributional error through an exponential amplification factor controlled by the integrated Lipschitz constant. Here L_t denotes a spatial Lipschitz envelope for the learned velocity field, which is often made comparable to the population envelope by restricting estimation to a suitably Lipschitz-controlled class [20]. Consequently, if this integral diverges rapidly as $\sigma \to 0$, the bound is vacuous.

The ambient-analysis bottleneck. Much of the available FM convergence theory tracks error through ambient-dimensional, isotropic Lipschitz controls, and these deteriorate sharply as the data concentrate on a submanifold. In the bounded-support and kernel-density-based analyses presently available, without uniform log-concavity and log-convexity assumptions, the velocity Lipschitz bounds scale as $O(\sigma_t^{-3})$, where σ_t denotes the effective noise scale and is comparable to $(1-t) \vee \sigma$ [27, 48]. In the FM setting, inserting such results into the standard stability upper bounds leads to severe terminal amplification [4]. A closely related phenomenon also appears in diffusion-model analyses in the manifold scenario, where control of the tangent process produces analogous terminal growth [6]. Analyses that bypass this difficulty either assume away the singular regime, or treat the needed regularity as a prerequisite. For instance, [18] require bounded-support densities, ruling out manifold-concentrated targets; [25] achieve the first intrinsic-dimension W_2 rates for FM on manifolds, but their analysis assumes a sub-linear Jacobian bound on the true velocity (their Assumption 3) and works at $\sigma = 0$ with early stopping, identifying the noisy-tube case $\sigma > 0$ as an open problem. On the structural side, ambient FM and Gaussian interpolation flows (GIFs) analyses already identify posterior conditioning and covariance as the key objects behind velocity regularity: Gaussian interpolation formulas express the true velocity field and its Jacobian, $v_t^\star$ and $\nabla_x v_t^\star$, through conditional moments [19], while [26] develops an exact Jacobian identity and a covariance-based Lipschitz regularity framework. What remains underexplored for manifold-concentrated targets is a geometric verification that the relevant posterior covariance collapses at the tube scale.

Qualitative anisotropy is known. Several recent works establish that manifold geometry imprints an anisotropic structure on score-based quantities: the score aligns with the normal bundle as noise vanishes [43], score singularity under the submanifold assumption has been analyzed via conditional-expectation and small-time asymptotics [37], score accuracy requirements separate sharply between geometric and distributional components [31], and recent diffusion analyses make the tangential–normal scale discrepancy explicit [36]. Related FM-side work further shows that the OT-CFM flow exhibits strictly negative terminal Lyapunov exponents in normal directions and zero tangentially [17], while tangent-space alignment structurally induces robustness despite perturbation [9]. However, these results are either asymptotic, qualitative without explicit rates, or concern the score rather than the population FM velocity field. None provides the pointwise, finite-time, quantitative rate envelopes needed for ODE error propagation.

Our contributions: tube-local regularity and anisotropic Jacobian geometry. Our analysis separates two questions that are often conflated. The first is an *isotropic* one: how large can the velocity Jacobian be in the operator norm? The second is a genuinely *anisotropic* one: how is that Jacobian distributed across tangential, normal, and mixed directions relative to the manifold?

Define $\sigma_t^2 := (1-t)^2 + \sigma^2 t^2$ with $t \in [0,1]$. At the isotropic level, recent manifold-score analyses indicate that score-based quantities become singular near thin tubes at the scale σ_t^{-2} [37, 36]. The score-form identity

$$\nabla_x v_t^\star = \frac{(1-t)\nabla^2 \log p_t + I_d}{t}$$

suggests a cancellation, since $1-t \le \sigma_t$, and therefore points to the milder scale $O(\sigma_t^{-1})$. We make this heuristic rigorous through the exact posterior-covariance formula of $\nabla_x v_t^\star$ and a tube-local posterior-covariance collapse guarantee, which yields the $O(\sigma_t^{-1})$ velocity Jacobian bound on the relevant tube.

The second step goes beyond operator-norm control. Using Fermi coordinates [21, 29] and small-noise geometric expansions, we analyze the tangential, normal, and mixed projections of the score

Hessian and transfer them through the Jacobian identity for the velocity. This yields a pointwise anisotropic decomposition of $\nabla_x v_t^\star$: tangential growth at scale $O(1 + \kappa r/\sigma_t)$, normal contraction at rate $\lambda_N^{\mathrm{flat}}(t)/2$ on the contractive window, and mixed tangential-normal coupling of lower order $O(\sigma_t)$. In particular, the small-noise geometric expansions recover the isotropic $O(\sigma_t^{-1})$ bound. It is this directional structure, rather than the scalar Lipschitz regularity alone, that feeds the stability theory where the tangential part carries genuine amplification.

In a nutshell, we develop the following regularity theory for FM on the tube-convolution target $\nu_\sigma = (f \operatorname{Vol}_{\mathcal{M}}) * \mathcal{N}(0, \sigma^2 I_d)$, where $\mathcal{M}$ is a compact C^3 submanifold of dimension k with positive reach τ, f is a C^2 density on $\mathcal{M}$, and $\sigma \in (0, \min(1, \tau/4))$.

1. **Tube-local Lipschitz control and anisotropic velocity geometry** (Sections 3–4). We first analyze the late-regime posterior covariance and prove the tube-local $O(\sigma_t^{-1})$ velocity Jacobian bound, thereby deriving from geometry a Jacobian control for the noisy-tube regime. We then establish late-regime score and Hessian projection bounds and use them to decompose the velocity Jacobian into tangential, normal, and mixed blocks, yielding a pointwise quadratic-form control on a contractive window $[t^\star, t^\dagger]$: tangential growth $O(1 + \kappa r/\sigma_t)$, normal contraction at rate $\lambda_N^{\mathrm{flat}}(t)/2$, and mixed coupling $O(\sigma_t)$. This anisotropic decomposition complements the isotropic $O(\sigma_t^{-1})$ bound with directional information, while making the asymptotic Lyapunov picture of [17] quantitative, pointwise, and finite-time.
2. **Downstream consequences** (Section 5). We sketch how the anisotropic block structure yields an ODE stability bound with controlled late-time amplification, and how tube-adapted function classes deliver estimation rates with intrinsic-dimension exponents. These results provide complementary transport and statistical components, while combining them for one estimator additionally requires conditions detailed in Appendix E.

Organization. Section 2 sets up the geometric model, FM objects, and the tube-coordinate toolkit. Section 3 proves the $O(\sigma_t^{-1})$ velocity Jacobian bound via posterior-covariance collapse. Section 4 develops the anisotropic tangential-normal block decomposition. Section 5 sketches the stability and estimation consequences. Section 6 discusses limitations and open problems. Detailed related work is in Appendix A; full proofs are in Appendices B–E, with numerical verification in Appendix F.

2 Setup and tube coordinates

2.1 Geometric model

Our target distribution is an ambient Gaussian thickening of a low-dimensional source measure. We make the following standing assumptions.

Assumption 1 (Manifold regularity)**.** $\mathcal{M} \subset \mathbb{R}^d$ is a compact, boundaryless, embedded C^3 submanifold of dimension k with reach $\tau > 0$ and second fundamental form uniformly bounded by $\kappa \le 1/\tau$.

By Federer's reach theory [14], every point x within the open tube $\mathcal{M}_r := \{x \in \mathbb{R}^d : \operatorname{dist}(x, \mathcal{M}) < r\}$, $r < \tau$, admits a unique nearest-point projection $\pi_{\mathcal{M}}(x) \in \mathcal{M}$ and a normal displacement $\rho_{\mathcal{M}}(x) := x - \pi_{\mathcal{M}}(x) \in N_{\pi_{\mathcal{M}}(x)}\mathcal{M}$. We write $P_T(y)$ and $P_N(y)$ for the orthogonal projections onto $T_y\mathcal{M}$ and $N_y\mathcal{M}$ in $\mathbb{R}^d$. For $y \in \mathcal{M}$ and $\rho \in N_y\mathcal{M}$, the normal-paired shape operator is

$$S_\rho(y) : T_y\mathcal{M} \to T_y\mathcal{M}, \qquad \langle u, S_\rho(y)v\rangle := \langle \rho, \mathrm{II}_y(u, v)\rangle\,, \tag{1}$$

where II_y denotes the second fundamental form at y; this curvature object enters the late-time Hessian expansion in Section 4. For a map g and a set S, we write $\operatorname{Lip}(g; S) := \sup_{x \neq x' \in S} \|g(x) - g(x')\|/\|x - x'\|$ for the restricted Lipschitz constant. For background on second fundamental form, tubes and Fermi coordinates, see [30, 21].

Assumption 2 (Intrinsic density)**.** $f : \mathcal{M} \to (0, \infty)$ is a density with respect to $\operatorname{Vol}_{\mathcal{M}}$ satisfying $0 < f_{\min} \le f \le f_{\max} < \infty$, with $\log f \in C^2(\mathcal{M})$ and $\|\nabla_{\mathcal{M}} \log f\|_\infty + \left\|\nabla^2_{\mathcal{M}} \log f\right\|_\infty \le L_f < \infty$.

Assumption 3 (Tube thickness)**.** Fix $\sigma \in (0, \min\{1, \tau/4\})$. Define the *tube-convolution target*

$$\nu_\sigma := (f \operatorname{Vol}_{\mathcal{M}}) * \mathcal{N}(0, \sigma^2 I_d), \tag{2}$$

with ambient density q_σ.

The tube-convolution model (2) is a natural Gaussian thickening of a manifold-supported distribution: $\sigma > 0$ ensures that ν_σ has a smooth ambient density, and the central question of this paper is how the regularity of the associated FM velocity field degrades as $\sigma \to 0$ and the target sharpens around $\mathcal{M}$.

2.2 Flow matching

We work with deterministic conditional flow matching (CFM) under the independent coupling [2, 33]. Let $\mu_0 = \mathcal{N}(0, I_d)$ and $\mu_1 = \nu_\sigma$. Draw $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$ independently and define the linear interpolant

$$X_t := (1-t)X_0 + tX_1, \qquad t \in [0,1]. \tag{3}$$

Its marginal law has density p_t, and the effective noise variance is

$$\sigma_t^2 = (1-t)^2 + t^2\sigma^2. \tag{4}$$

The population velocity field is

$$v_t^\star(x) := \mathbb{E}[X_1 - X_0 \mid X_t = x], \tag{5}$$

and the ODE $\dot{\widetilde{X}}_t = v_t^\star(\widetilde{X}_t)$ with $\widetilde{X}_0 \sim \mu_0$ generates marginals matching p_t under mild conditions; see, e.g., [2, 19].

The key structural identity for the velocity involves the posterior moments of the manifold source. Writing $X_1 = Y + \sigma\xi$ with $Y \sim f\,\mathrm{Vol}_\mathcal{M}$ and $\xi \sim \mathcal{N}(0, I_d)$ independent, define $m_t(x) := \mathbb{E}[Y \mid X_t = x]$ and $\Sigma_{Y|t}(x) := \mathrm{Cov}(Y \mid X_t = x)$.

Proposition 2.1. *Under Assumptions 1–3, for every $t \in [0,1]$ and $x \in \mathbb{R}^d$, with $\alpha_t := (t(1+\sigma^2) - 1)/\sigma_t^2$,*

$$v_t^\star(x) = \alpha_t\, x + \frac{1-t}{\sigma_t^2}\, m_t(x), \tag{6}$$

and the velocity Jacobian admits the exact decomposition

$$\nabla_x v_t^\star(x) = \alpha_t\, I_d + \frac{t(1-t)}{\sigma_t^4}\, \Sigma_{Y|t}(x). \tag{7}$$

The identities (6)–(7) are direct Gaussian-posterior formulas in our model and fit into a broader ambient FM or GIF toolkit: conditional-expectation and score-form characterizations appear in [19, 1], while covariance expressions of the Jacobian are also derived in ambient form in [19, 26]. Eq. (7) isolates the latent manifold source Y, which is the object needed for the tube-local geometric analysis in Section 3. While the velocity Jacobian can also be expressed through the score Hessian (see Section 4), the posterior covariance $\Sigma_{Y|t}$ is the more convenient route because its late-time collapse can be verified directly from the geometry of the Gaussian-smoothed manifold model.

Given n i.i.d. training pairs, the common CFM training objective is a quadratic regression loss over the interpolation time and a function class $\mathcal{V}$. To exploit the different geometric behavior of the early and late regimes, our theoretical estimator, denoted as $\widehat{v}$, is a time-split velocity field whose early component is obtained globally in space and whose late component is learned on the thin tube around $\mathcal{M}$; see Appendix E for the precise formulation. The learned ODE generates a terminal law $\widehat{\nu}_\sigma$, and our primary error criterion is $\mathrm{W}_2(\widehat{\nu}_\sigma, \nu_\sigma)$.

2.3 Tube localization

The FM marginal p_t admits a simple geometric description: after rescaling by $1/t$, it becomes a Gaussian thickening of the manifold-supported source measure with rescaled effective noise level $\varepsilon_t := \sigma_t/t$. Thus late-time analysis of the FM path can be reduced to a static small-noise problem on $\mathcal{M}$, and the corresponding marginals concentrate in thin tubes at the natural scale set by σ_t.

Let $R_\mathcal{M} := \sup_{y\in\mathcal{M}} \|y\| < \infty$ and write q_ε for the density of $(f\,\mathrm{Vol}_\mathcal{M}) * \mathcal{N}(0, \varepsilon^2 I_d)$.

Lemma 2.2. *For every $t \in (0,1]$,*

$$p_t(x) = t^{-d}\, q_{\varepsilon_t}(x/t). \tag{8}$$

Equivalently, if $Z_t = X_t/t$, then $\mathcal{L}(Z_t)$ has density q_{ε_t}, where $\mathcal{L}(\cdot)$ denotes the law of a random element. Hence, all late-regime characteristics for p_t may be obtained from small-noise properties for the static family q_ε, with time dependence entering only through $\varepsilon_t = \sigma_t/t$.

Lemma 2.3 (Tube localization of FM marginals)**.** *Under Assumptions 1–3, for every $t \in [0,1]$ and $\delta \in (0, 1/2)$,*

$$\mathbb{P}\left(X_t \in \mathcal{M}_{r(t,\delta)}\right) \geq 1-\delta, \qquad r(t,\delta) := (1-t)\, R_{\mathcal{M}} + \sigma_t\big(\sqrt{d} + \sqrt{2\log(1/\delta)}\big). \tag{9}$$

In particular, as $t \uparrow 1$, the localization radius reduces to $O(\sigma_t\sqrt{d})$, so the terminal and late-time marginals lie with high probability in tubes whose thickness is proportional to the effective noise scale. This is the localization property used in Sections 3 and 4. The proofs are deferred to Appendix B.

3 Velocity Lipschitz regularity

Recent manifold-score analyses indicate that score-based quantities near a thin tube around $\mathcal{M}$ are singular at the natural scale σ_t^{-2} [37, 36, 31]. Through the velocity–score relationship this points heuristically to a milder bound of order σ_t^{-1} for $\|\nabla_x v_t^\star\|_{\mathrm{op}}$, but making that intuition concrete would require delicate Hessian control in the tube regime, a task we defer to Section 4. Here we instead take the posterior-covariance route from (7). Ambient FM and GIF work already identifies posterior covariance as a natural control variable for $\nabla_x v_t^\star$ [19, 26]; our contribution in this section is to verify the required covariance collapse in the Gaussian-smoothed manifold scenario, which turns the σ_t^{-4} prefactor in (7) into the sharper $O(\sigma_t^{-1})$ Jacobian bound.

Proposition 3.1 (Posterior-covariance collapse)**.** *Under Assumptions 1–3, fix an observation radius $\varrho \in (0, \tau/2)$. There exist constants $\varepsilon_0 := \varepsilon_0(\varrho, \mathcal{M}, f) > 0$ and $C_\Sigma := C_\Sigma(\varrho, \mathcal{M}, f) < \infty$ such that for all $t \in (0,1]$ with $\varepsilon_t = \sigma_t/t \leq \varepsilon_0$, and for all $x \in \mathbb{R}^d$ satisfying $x/t \in \mathcal{M}_\varrho$,*

$$\big\|\Sigma_{Y|t}(x)\big\|_{\mathrm{op}} \leq C_\Sigma\, \frac{\sigma_t^2}{t^2}. \tag{10}$$

Consequently, fix $r \in (0, \varrho)$ and define $t_{\mathrm{tube}} := (R_{\mathcal{M}}+r)/(R_{\mathcal{M}}+\varrho) \in (0,1)$. Then (10) *holds for all $t \in [t_{\mathrm{tube}}, 1]$ with $\varepsilon_t \leq \varepsilon_0$ and all $x \in \mathcal{M}_r$.*

Proposition 3.1 supplies the geometric step missing from the ambient covariance-based FM regularity arguments: although formulas expressing $\nabla_x v_t^\star$ through posterior covariance are known [19, 26], those works do not verify the needed σ_t^2-scale covariance decay for manifold-concentrated targets. Here, the bound (10) shows that, when the rescaled observation x/t lies in an appropriately fixed tube around $\mathcal{M}$, the posterior of Y given $X_t{=}x$ concentrates near the nearest point $\pi_{\mathcal{M}}(x/t)$ within a ball of radius $O(\sigma_t/t)$. The mechanism is a local-global competition: the Gaussian likelihood $\exp(-\|x/t - y\|^2/(2\varepsilon_t^2))$ creates a sharply peaked local contribution from the nearest chart of $\mathcal{M}$, while the reach condition ensures a uniform gap in the squared distance to the far part, making the far contribution exponentially negligible. Outside the tube, however, the observation x/t may be equidistant from several patches of $\mathcal{M}$, the posterior remains multimodal, and the covariance retains an $O(1)$ between-mode contribution. While the bound is inherently tube-local, Lemma 2.3 shows that the FM marginals concentrate in $\mathcal{M}_r$ with high probability as $t \uparrow 1$, offering a tail control. In addition, $\varepsilon_0(\varrho, \mathcal{M}, f)$ depends only on $\varrho, \tau, k, \mathrm{diam}(\mathcal{M}), \mathrm{Vol}_{\mathcal{M}}(\mathcal{M}), f_{\min}, f_{\max}$, and $C_\Sigma(\varrho, \mathcal{M}, f)$ only on $\varrho, \tau, k, f_{\min}, f_{\max}$.

Proof sketch. Via the exact rescaling of Lemma 2.2, the posterior $\mathcal{L}(Y \mid X_t{=}x)$ is equivalent to the posterior in the static denoising model $Z = Y + \varepsilon_t\eta$, $\eta \sim \mathcal{N}(0, I_d)$, observed at $z = x/t$. It therefore suffices to prove $\|\mathrm{Cov}(Y \mid Z{=}z)\|_{\mathrm{op}} \leq C_\Sigma\varepsilon_t^2$ uniformly for $z \in \mathcal{M}_\varrho$ and $\varepsilon_t \leq \varepsilon_0$.

Since $\|\mathrm{Cov}(Y \mid Z{=}z)\|_{\mathrm{op}} \leq \mathbb{E}[\|Y - y_\star\|^2 \mid Z{=}z]$ with $y_\star = \pi_{\mathcal{M}}(z)$, we bound the second moment. The reach-controlled local chart $y(u) = y_\star + u + h(u)$ (where $h(0) = 0$, $Dh(0) = 0$, $\|h(u)\| \leq \|u\|^2/(2\tau)$) gives upper and lower quadratic bounds on $\|z - y(u)\|^2$, from which the local numerator contributes $O(\varepsilon_t^{k+2})$ while the normalizer is $\Omega(\varepsilon_t^k)$, yielding a local ratio of $O(\varepsilon_t^2)$. The far contribution is exponentially suppressed by the reach-induced gap, and choosing ε_0 small enough absorbs it. Details are in Appendix D. □

The posterior-covariance collapse feeds directly into the Jacobian identity (7) to produce the following regularity result: a quantitative Jacobian bound for the population FM velocity derived from the geometry of the Gaussian-smoothed manifold model, rather than assumed a priori or imported from an ambient distribution class.

Theorem 3.2 (Velocity Jacobian bound)**.** *Under Assumptions 1–3, fix $r \in (0, \tau/4)$, $\varrho \in (r, \tau/2)$, recall $t_{\mathrm{tube}} = (R_{\mathcal{M}}+r)/(R_{\mathcal{M}}+\varrho)$, and let ε_0, C_Σ be as in Proposition 3.1. For all $t \in [0,1]$ (with $x \in \mathcal{M}_r$ and $\varepsilon_t \le \varepsilon_0$ when $t \ge t_{\mathrm{tube}}$),*

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le \frac{C_{\mathcal{M}}(r)}{\sigma_t}, \tag{11}$$

where $C_{\mathcal{M}}(r)$ admits an upper bound depending only on $\varrho, \mathcal{M}, f$ for any fixed $\varrho < \tau/2$ and any $r < \varrho/2$. The bound applies globally for $t < t_{\mathrm{tube}}$ and tube-locally for $t \ge t_{\mathrm{tube}}$; the explicit three-regime definition of $C_{\mathcal{M}}(r)$ is given in Appendix D.

For the stability analysis in Section 5, we split the ODE at a switching time $t^\star := 1/(1+\sigma)$, where $1 - t^\star = t^\star\sigma$ and thus the source-noise and target-smoothing contributions to σ_t are equal. We assume the learned velocity admits a comparable global bound on $[0, t^\star]$ (Proposition E.10; yet $v_t^\star$ itself does not necessarily satisfy it when $t^\star \ge t_{\mathrm{tube}}$). Notice that

$$\int_0^{t^\star} \frac{1}{\sigma_t}\,\mathrm{d}t \lesssim 1 + \log\frac{1}{\sigma} \quad \text{while} \quad \int_0^{t^\star} \frac{1}{\sigma_t^3}\,\mathrm{d}t \asymp \frac{1}{\sigma^2}.$$

Under the assumed global bound on the learned velocity, standard Grönwall amplification [22] over this early window is therefore at most polynomial in $1/\sigma$; a cubic-scale envelope, of the kind appearing in prior analyses [27, Lemma 5.2][48, Lemma 1.3], would instead yield an $\exp(O(1/\sigma^2))$ factor, making the distributional error bound grow drastically as σ shrinks. The late regime $[t^\star, 1]$ requires a different argument: the next section shows that the velocity Jacobian there is strongly contractive in normal directions, so that the anisotropic structure controls the remaining amplification.

4 Anisotropic velocity structure

Section 3 establishes the scalar bound via posterior-covariance collapse, which treats all directions equally. We now refine it to a directional statement in the tangential–normal splitting. This anisotropy is consistent with existing small-noise manifold analyses of the score, which show that the normal component carries the singular behavior while tangential directions remain milder [31, 36]; related concurrent work also identifies the corresponding static Hessian splitting for Gaussian-smoothed manifold measures in the uniform-density setting [10]. Our goal here is different: we convert these geometric asymptotics into a pointwise, finite-time theorem for the *FM population velocity Jacobian* $\nabla_x v_t^\star$, with explicit envelopes along the FM schedule.

Proposition 4.1. *Under Assumptions 1–3, for every $t \in (0,1]$ and $x \in \mathbb{R}^d$ with $p_t(x) > 0$,*

$$\nabla_x v_t^\star(x) = \frac{(1-t)\,\nabla_x^2 \log p_t(x) + I_d}{t}. \tag{12}$$

Eq. (12) stems directly from Tweedie's formula [12]. Equivalent score-form identities also appear in FM and GIF analyses, and we record (12) here since it is the bridge from small-noise Hessian asymptotics for p_t to structural information about the FM velocity Jacobian. While the posterior-covariance form (7) was better suited to deriving the scalar Jacobian bound, the score form (12) is more natural for directional analysis. Indeed, in tube coordinates the singular part of the small-noise log-density is the squared-distance potential

$$\phi(x) := \frac{1}{2}\delta_{\mathcal{M}}(x)^2 = \frac{1}{2}\,\|\rho_{\mathcal{M}}(x)\|^2,$$

and the leading contribution to $\nabla^2 \log p_t$ after rescaling is $-\nabla^2\phi/\sigma_t^2$.

The next result gives the block expansion of the score Hessian in the late-time regime. Recall the normal-paired shape operator $S_\rho(y)$ from (1).

Proposition 4.2 (Late-regime Hessian block expansion)**.** *Under Assumptions 1–3, fix $0 < r \le r_0 < \tau/4$, where r_0 is a universal master radius. There exist $\varepsilon_0 > 0$ and $C_{\mathrm{sc}} < \infty$, depending only on r_0, $\mathcal{M}$ and f, such that for every $t \in [t^\star, 1]$ with $\sigma_t/t \le \varepsilon_0$ and every $x \in \mathcal{M}^{\mathrm{sc}}_{t,r} := \{z \in \mathbb{R}^d : z/t \in \mathcal{M}_r\}$, writing $(y_t, \rho_t) := (\pi_{\mathcal{M}}(x/t),\ \rho_{\mathcal{M}}(x/t))$, one has*

$$\nabla \log p_t(x) = -\frac{t}{\sigma_t^2}\,\rho_t + b_t(x), \qquad \|b_t(x)\| \le C_{\mathrm{sc}}, \tag{13}$$

$$\Big\| P_T(y_t)\,\nabla_x^2 \log p_t(x)\,P_T(y_t) - \sigma_t^{-2}\,S_{\rho_t}(y_t)\big(I - S_{\rho_t}(y_t)\big)^{-1} \Big\|_{\mathrm{op}} \le C_{\mathrm{sc}}, \tag{14}$$

$$\big\| P_T(y_t)\,\nabla_x^2 \log p_t(x)\,P_N(y_t) \big\|_{\mathrm{op}} \le C_{\mathrm{sc}}, \tag{15}$$

$$\big\| P_N(y_t)\,\nabla_x^2 \log p_t(x)\,P_N(y_t) + \sigma_t^{-2}\,P_N(y_t) \big\|_{\mathrm{op}} \le C_{\mathrm{sc}}. \tag{16}$$

Remark 4.3. The structural tangential–normal splitting in Proposition 4.2 is closely related to recent small-noise Hessian expansions for Gaussian-smoothed manifold measures, in particular [10], which presents the same static leading-order structure in the uniform-measure setting. Our formulation differs in three respects: it allows a general smooth density f on $\mathcal{M}$, it is stated for the FM time marginal via the exact rescaling (8), and it is used as a preliminary result to the velocity Jacobian theorem below.

Proof sketch. By exact rescaling of Lemma 2.2, $\nabla_x^2 \log p_t(x) = t^{-2}\nabla^2 \log q_{\varepsilon_t}(x/t)$ with $\varepsilon_t = \sigma_t/t$, reducing the problem to the tube-convolution density q_ε for a small $\varepsilon > 0$. In Fermi coordinates (y, ρ) near $\mathcal{M}$, the log-density admits the expansion

$$\log q_\varepsilon(y + \rho) = -\frac{\|\rho\|^2}{2\varepsilon^2} + \log f(y) - \frac{1}{2}\log\det(I - S_\rho(y)) + R_\varepsilon(y, \rho),$$

where R_ε is uniformly C^2-small on the master tube $\mathcal{M}_{r_0}$, established via a fixed-atlas argument. The singular part of the Hessian comes entirely from the squared-distance term, whose ambient Hessian has tangential block $S_\rho(I - S_\rho)^{-1}/\varepsilon^2$, zero mixed block, and normal block $-P_N/\varepsilon^2$. An analogous static Hessian expansion appears in [10, Theorem 3.5] for the uniform-measure setting; here we formulate the analysis in the FM-rescaled, non-uniform-density setting needed later. After the t^{-2} rescaling and absorbing the bounded remainder into C_{sc}, one obtains (13)–(16). The full proof, including the atlas construction and the Fermi-to-ambient transfer, is in Appendix C. □

These upper bounds admit a clear geometric reading. The *normal–normal* block (16) carries the dominant Gaussian-fiber curvature at rate σ_t^{-2}: points displaced from $\mathcal{M}$ are pulled back at a rate inversely proportional to the squared effective noise scale. The *tangential–tangential* block (14) is modulated by the shape operator: it vanishes on $\mathcal{M}$ itself (where $\rho_t = 0$) and grows as $O(\kappa r/\sigma_t^2)$ at tube radius r. The *tangential–normal* block (15) is bounded by a constant, independent of σ_t. The reason is that the singular term comes from $\phi(x) = \delta_{\mathcal{M}}(x)^2/2$: in Fermi coordinates, ϕ is quadratic in the normal displacement, so its ambient Hessian contributes the σ_t^{-2} singularity only to the pure tangential and normal blocks, while the mixed block has no singular contribution, resulting from the Weingarten formula [11] (Proposition C.6).

We now transfer these Hessian asymptotics to the velocity Jacobian. For $t \in [t^\star, 1]$ and x with $x/t \in \mathcal{M}_\tau$, define the *moving tangential–normal splitting*

$$P_T^{(t)}(x) := P_T(\pi_{\mathcal{M}}(x/t)), \qquad P_N^{(t)}(x) := P_N(\pi_{\mathcal{M}}(x/t)), \tag{17}$$

and the *flat normal eigenvalue*

$$\lambda_N^{\mathrm{flat}}(t) := \frac{(1-t) - t\sigma^2}{\sigma_t^2}. \tag{18}$$

Let

$$t^\dagger := \frac{1 - \sigma^2/2}{1 + \sigma^2}. \tag{19}$$

On the *contractive window* $[t^\star, t^\dagger]$, a direct computation gives $(1-t) - t\sigma^2 \ge \sigma^2/2$, and hence $\lambda_N^{\mathrm{flat}}(t) \ge 1/4$ throughout. Note that this interval is nonempty for $\sigma \le 1/2$.

Theorem 4.4 (Anisotropic velocity Jacobian)**.** *Under Assumptions 1–3, fix $0 < r \le r_0 < \tau/4$, where r_0 is a universal master radius. There exists $\sigma_0 = \sigma_0(r_0, \mathcal{M}, f) \in (0, 1/2]$ such that, whenever $\sigma \le \sigma_0$, the following holds for every $t \in [t^\star, t^\dagger]$ and every $x \in \mathcal{M}^{\mathrm{sc}}_{t,r}$. Writing $y_t := \pi_{\mathcal{M}}(x/t)$, the Jacobian of $v^\star_t$ satisfies the quadratic-form block bound*

$$\langle u + w, \nabla_x v^\star_t(x)(u + w)\rangle \le L_T(t)\ \|u\|^2 - \frac{1}{2}\lambda^{\mathrm{flat}}_N(t)\ \|w\|^2 + 2c_{TN}(t)\ \|u\|\,\|w\| \tag{20}$$

for all $u \in T_{y_t}\mathcal{M}$ and $w \in N_{y_t}\mathcal{M}$, with envelopes

$$L_T(t) := 2 + C_{\mathrm{sc}}\sigma + \frac{8\kappa\, r}{3\sigma_t}, \qquad c_{TN}(t) := 2C_{\mathrm{sc}}\sigma_t. \tag{21}$$

Theorem 4.4 turns the anisotropic geometry into an explicit statement about the FM velocity Jacobian. The tangential growth $L_T(t) = O(1+\kappa r/\sigma_t)$ is controlled by the curvature of $\mathcal{M}$ and the tube radius, in particular uniformly bounded when r is comparable to σ. The normal contraction rate $\lambda^{\mathrm{flat}}_N(t)/2$ is bounded below by an absolute constant, independent of σ and the geometry of $\mathcal{M}$. The mixed coupling $c_{TN}(t) = O(\sigma_t)$ is lower order on the contractive window. In this sense, the theorem turns the qualitative anisotropic picture suggested by prior score-side work and the asymptotic Lyapunov analysis of [17] into a pointwise, finite-time FM Jacobian statement with explicit envelopes.

As a corollary (Corollary D.5), the Hessian block analysis also recovers the scalar Jacobian bound from Section 3 via an alternative route, both reflecting the manifold geometry. In addition, the $O(1/\sigma_t)$ operator-norm rate is sharp; the normal block at $t^\star$ has magnitude $\Omega(1/\sigma_{t^\star})$ even on the manifold centerline (Corollary D.6). Appendix F provides numerical verification of these velocity regularity properties in a circle model.

5 Application: stability and estimation

We sketch how the velocity regularity structures from Sections 3–4 enable ODE stability and intrinsic-dimension-controlled estimation; extended statements and proofs are in Appendix E.

5.1 Anisotropic ODE stability

We now explain how the block-Jacobian bound of Theorem 4.4 enters an ODE stability argument. Let $\widetilde{X}_t$ solve the population ODE

$$\dot{\widetilde{X}}_t = v^\star_t(\widetilde{X}_t), \qquad \widetilde{X}_0 \sim \mathcal{N}(0, I_d),$$

and let $\widehat{X}_t$ solve the learned ODE

$$\dot{\widehat{X}}_t = \widehat{v}_t(\widehat{X}_t),$$

started from a coupling with $\widetilde{X}_t$ at the switching time $t^\star = 1/(1+\sigma)$. Under the regularity of $v^\star_t$ from Theorem 3.2, we assume that the population flow has marginals p_t, the density of X_t. This requires uniqueness of solutions to the corresponding continuity equation [8]. We write

$$e_t := \widehat{v}_t - v^\star_t, \qquad \Delta_t := \widehat{X}_t - \widetilde{X}_t, \qquad \mathcal{E}_t := \|\Delta_t\|^2,$$

and denote by $\widehat{p}_t$ the law of $\widehat{X}_t$, with terminal law $\widehat{\nu}_\sigma := \widehat{p}_1$; the target terminal law is $\nu_\sigma = p_1$. Our aim is to derive a well-controlled bound on $\mathrm{W}_2(\widehat{\nu}_\sigma, \nu_\sigma)$ from the velocity error and the anisotropic Jacobian structure.

The main difficulty is that Theorem 4.4 is tube-local, whereas a learned trajectory need not remain in the relevant tube a priori. We therefore use a stopped comparison: the two coupled trajectories are compared only until either the population trajectory leaves the inner scaled tube or the two trajectories separate by more than a prescribed threshold. Fix radii $0 < r < \bar{r} \le r_0 < \tau/4$ and define, for a threshold $\delta_* > 0$ specified in Appendix E.2,

$$\theta := \inf\Big\{t \in [t^\star, 1] :\ \widetilde{X}_t \notin \mathcal{M}^{\mathrm{sc}}_{t,r} \text{ or } \|\Delta_t\| > \delta_*\Big\},$$

together with the survival event

$$G := \{\theta = 1\} \cap \{\widetilde{X}_1 \in \mathcal{M}^{\mathrm{sc}}_{1,r}\} \cap \{\|\Delta_1\| \le \delta_*\}.$$

Before time θ, the true trajectory, the learned trajectory, and every mean-value segment $\widetilde{X}_t + s\Delta_t$, $s \in [0,1]$, lie in the enlarged scaled tube $\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$ given a suitable δ_*. Hence the population Jacobian along the segment inherits the tangential-growth, normal-contraction, and mixed-coupling bounds of Theorem 4.4. Applying the fundamental theorem of calculus,

$$\dot{\Delta}_t = \int_0^1 \nabla_x v_t^\star(\widetilde{X}_t + s\Delta_t)\,\mathrm{d}s\,\Delta_t + e_t(\widehat{X}_t),$$

and the negative normal block absorbs the mixed tangential–normal term in the resulting energy inequality. This gives the stopped energy bound

$$\mathbb{E}\sup_{t\in[t^\star,1]} \mathcal{E}_{t\wedge\theta} \le e^{\Psi_{\mathrm{eff}}}\Big(\mathbb{E}[\mathcal{E}_{t^\star}] + \Phi_\theta\Big), \qquad \Phi_\theta := \mathbb{E}\int_{t^\star}^{\theta} \Big\|e_s(\widehat{X}_s)\Big\|^2\,\mathrm{d}s.$$

Here Ψ_{eff} is the integrated expansive budget after the normal contraction has been used (Theorem E.7); for every fixed outer radius $\bar{r}$, $\Psi_{\mathrm{eff}} = O(1)$, rather than $O(\int \|\nabla v_t^\star\|_{\mathrm{op}}\,\mathrm{d}t)$.

It remains to compare the forcing term Φ_θ, evaluated along the learned flow, with the velocity risk measured under the population marginal. A convenient sufficient condition is a scaled-tube density-ratio bound: assume that $\widehat{p}_t \ll p_t$ on $\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$ for every $t \in [t^\star, 1]$ and that

$$\Lambda_{\mathrm{sc}}(\bar{r}) := \sup_{t\in[t^\star,1]} \operatorname*{ess\,sup}_{x\in\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}} \frac{\widehat{p}_t}{p_t}(x) < \infty.$$

Then

$$\Phi_\theta \le \Lambda_{\mathrm{sc}}(\bar{r})\,\mathcal{R}^{\mathrm{late}}_{\bar{r}_+}(\widehat{v}), \qquad \bar{r}_+ := \bar{r} + R_{\mathcal{M}}\sigma,$$

where $\mathcal{R}^{\mathrm{late}}_{\bar{r}_+}$ is the late-time tube risk on the slightly enlarged ambient tube (Definition E.1).

The stopped argument also controls the probability that the coupled trajectories fail to remain in the tube-local regime. Let

$$\delta_X(r,\sigma) := \mathbb{P}\big(\exists\, t \in [t^\star,1] :\ \widetilde{X}_t \notin \mathcal{M}^{\mathrm{sc}}_{t,r}\big)$$

be the population-flow tube-exit probability. Then

$$\mathbb{P}(G^c) \le \delta_X(r,\sigma) + \frac{e^{\Psi_{\mathrm{eff}}}}{\delta_*^2}\Big(\mathbb{E}[\mathcal{E}_{t^\star}] + \Phi_\theta\Big),$$

and the Gaussian normal perturbation in the tube-convolution target gives a Gaussian-type tail for $\delta_X(r,\sigma)$ as a function of r/σ. Combining the stopped interior bound with this exit probability and an ambient terminal $\mathfrak{p}$-moment bound, $\mathfrak{p} > 2$, yields the schematic transport bound

$$\mathrm{W}_2(\widehat{\nu}_\sigma, \nu_\sigma) \le \sqrt{\mathcal{S}} + 2\bar{M}^{1/\mathfrak{p}} q_*^{1/2-1/\mathfrak{p}}, \tag{22}$$

where

$$\mathcal{S} := e^{\Psi_{\mathrm{eff}}}\Big(B_{\mathrm{early}} + \Lambda_{\mathrm{sc}}(\bar{r})\,\mathcal{R}^{\mathrm{late}}_{\bar{r}_+}(\widehat{v})\Big), \qquad q_* := \delta_X(r,\sigma) + \delta_*^{-2}\mathcal{S}.$$

Here B_{early} is a switching-time coupling budget for $\mathbb{E}[\mathcal{E}_{t^\star}]$ and $\bar{M}$ is a terminal $\mathfrak{p}$-moment bound (Lemma E.18 and Corollary E.19). Eq. (22) demonstrates that the anisotropic Jacobian structure turns late-time propagation into a bounded effective amplification.

5.2 Curved-manifold estimation sketch

For curved manifolds, the estimator analyzed here uses the same early–late time splitting suggested by the geometry. In the late regime, when the FM marginal is confined to a thin tube around $\mathcal{M}$, we use a tube-local estimator in local charts, approximate the tangential coefficient fields by chartwise C^α functions, and retain the leading linear normal term from the tube expansion. Let h denote the chart resolution, equivalently the sieve mesh width. Each chart contributes $O(h^{-k})$ degrees of freedom, so the late estimator has intrinsic complexity $O(h^{-k})$. In the early regime, where tube localization is unavailable, we use a global-in-space source-side construction.

In the thin-tube low-Hölder scenario $r \lesssim \sigma$ and $0 < \alpha \le 1$, Appendix E.6 proves the concrete expected tube-risk bound

$$\mathbb{E}[\mathcal{R}_r(\widehat{v})] \le O(\sigma^{-k/(2\alpha+k)} n^{-2\alpha/(2\alpha+k)}\,\mathrm{polylog}(n,d,1/\sigma) + (1+\kappa^2)\sigma^2), \tag{23}$$

up to geometric prefactors. The dominant statistical term exhibits the exponent $2\alpha/(2\alpha + k)$, so the sample complexity is controlled by the intrinsic dimension k; curvature enters through geometric constants and the residual $O((1 + \kappa^2)\sigma^2)$ approximation term.

The stopped stability result (22) supplies the complementary tube-to-ambient mechanism, with bounded effective amplification and derived exit control. It still requires an early coupling budget, a scaled-tube density-ratio bound, and terminal moments to apply (22) to the stated estimator. Hence, the tube-risk rate and transport mechanism are proved separately, and establishing an end-to-end theorem remains future work. Furthermore, since $\mathrm{W}_2(\nu_\sigma, \nu_0) \le \sigma\sqrt{d}$, such a theorem could choose $\sigma = \sigma(n)$ to balance statistical error, stability amplification, and smoothing bias.

6 Discussion

We have established tube-local regularity for the FM velocity under the tube-convolution target. Posterior-covariance collapse yields the $1/\sigma_t$ Jacobian bound (Theorem 3.2), while Theorem 4.4 gives a quantitative finite-time tangential–normal decomposition, complementing prior qualitative and asymptotic descriptions [43, 36, 17]. The stopped stability analysis propagates this structure along coupled trajectories: normal contraction absorbs mixed interactions, yielding bounded effective late-time amplification and the ambient transport synthesis of Corollary E.19 without assuming learned-flow confinement in advance.

The remaining limitation is compatibility with the statistical construction. The transport synthesis requires an early coupling budget, a scaled-tube density-ratio bound, and terminal moments, none of which follows from tube-restricted L^2 risk alone. In particular, Remark E.15 motivates replacing the piecewise-defined late sieve by a refined tube class. We therefore prove the intrinsic-dimension tube-risk rate and the stopped transport mechanism separately, but not all required properties for one estimator. Our analysis also excludes the singular target $\sigma = 0$, non-Gaussian perturbations, and OT-path CFM.

For future work, a first direction is to integrate a compatible late class with the early source-side construction, promisingly rendering an end-to-end intrinsic-dimension transport theorem. A second is to transfer tools from stochastic localization and diffusion theory [41, 32, 24], potentially to obtain sharper results under weaker assumptions. Finally, tube-focused estimators that also enforce the flow regularity needed for density-ratio propagation deserve practical study, which may better exploit manifold geometry than ambient, globally uniform objectives.

Acknowledgments and Disclosure of Funding

The authors thank the Program Chairs and reviewers for their constructive and insightful advice to improve the manuscript.

This work was supported by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM904), the National Natural Science Foundation of China (Grant No. 12371289), the Shanghai Pilot Program for Basic Research (Grant No. TQ20220105), the Fundamental Research Funds for the Central Universities (Grant No. YBNLTS2025-047), JSPS KAKENHI (B) 26K02990, and JSPS Grant-in-Aid for Transformative Research Areas (A) 22H05106.

The authors declare that they have no competing interests.

References

[1] 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.

[2] Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

[3] Luigi Ambrosio and Gianluca Crippa. Continuity equations and ode flows with non-smooth velocity. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 144(6):1191–1244, 2014.

[4] Joe Benton, George Deligiannidis, and Arnaud Doucet. Error bounds for flow matching methods. *Transactions on Machine Learning Research*, 2024.

[5] Jean-Daniel Boissonnat, André Lieutier, and Mathijs Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of applied and computational topology*, 3(1):29–58, 2019.

[6] Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. Expert Certification.

[7] François Bouchut and François James. One-dimensional transport equations with discontinuous coefficients. *Nonlinear Analysis*, 32(7):891, 1998.

[8] C Le Bris and P-L Lions. Existence and uniqueness of solutions to fokker–planck type equations with irregular coefficients. *Communications in Partial Differential Equations*, 33(7):1272–1317, 2008.

[9] Nisha Chandramoorthy and Adriaan A. de Clercq. When and how can inexact generative models still sample from the data manifold? In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[10] Junhao Chen, Ruowei Li, and Zhigang Yao. Estimation of riemannian quantities from noisy data via density derivatives. *arXiv preprint arXiv:2603.27171*, 2026.

[11] Marcos Dajczer and Ruy Tojeiro. *Submanifold theory*. Springer, 2019.

[12] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

[13] Luke Evans, Octavian-Vlad Murad, Lars Dingeldein, Pilar Cossio, Roberto Covino, and Marina Meila. Cryo-em images are intrinsically low dimensional. *PRX Life*, 3(3):033025, 2025.

[14] Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.

[15] Nicholas I Fisher. *Statistical analysis of circular data*. cambridge university press, 1995.

[16] Robert L Foote. Regularity of the distance function. *Proceedings of the American Mathematical Society*, 92(1):153–155, 1984.

[17] Kenji Fukumizu, Wei Huang, Han Bao, Shuntuo Xu, and Nisha Chandramoothy. Flow matching from viewpoint of proximal operators. *arXiv preprint arXiv:2602.12683*, 2026.

[18] Kenji Fukumizu, Taiji Suzuki, Noboru Isobe, Kazusato Oko, and Masanori Koyama. Flow matching achieves almost minimax optimal convergence. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu, editors, *International Conference on Learning Representations*, volume 2025, pages 27608–27640, 2025.

[19] Yuan Gao, Jian Huang, and Yuling Jiao. Gaussian interpolation flows. *Journal of Machine Learning Research*, 25(253):1–52, 2024.

[20] Yuan Gao, Jian Huang, Yuling Jiao, and Shurong Zheng. Convergence of continuous normalizing flows for learning probability distributions. *arXiv preprint arXiv:2404.00551*, 2024.

[21] Alfred Gray. *Tubes*, volume 221. Springer Science & Business Media, 2003.

[22] Thomas Hakon Gronwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics*, 20(4):292–296, 1919.

[23] Morris W Hirsch. *Differential topology*. Springer Science & Business Media, 2012.

[24] Zhihan Huang, Yuting Wei, and Yuxin Chen. Denoising diffusion probabilistic models are optimally adaptive to unknown low dimensionality. *Mathematics of Operations Research*, 2026.

[25] Shivam Kumar, Yixin Wang, and Lizhen Lin. Flow matching is adaptive to manifold structures. *arXiv preprint arXiv:2602.22486*, 2026.

[26] Lea Kunkel. Distribution estimation via flow matching with lipschitz guarantees. *Journal of Multivariate Analysis*, page 105657, 2026.

[27] Lea Kunkel and Mathias Trabs. On the minimax optimality of flow matching through the connection to kernel density estimation. *arXiv preprint arXiv:2504.13336*, 2025.

[28] Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of statistics*, pages 1302–1338, 2000.

[29] John M Lee. Smooth manifolds. In *Introduction to smooth manifolds*, pages 1–29. Springer, 2003.

[30] John M Lee. *Introduction to Riemannian manifolds*, volume 2. Springer, Cham, 2018.

[31] Xiang Li, Zebang Shen, Ya-Ping Hsieh, and Niao He. When scores learn geometry: Rate separations under the manifold hypothesis. In *International Conference on Learning Representations*, volume 2026, pages 40627–40660, 2026.

[32] Jiadong Liang, Zhihan Huang, and Yuxin Chen. Low-dimensional adaptation of diffusion models: Convergence in total variation (extended abstract). In Nika Haghtalab and Ankur Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 3723–3729. PMLR, 30 Jun–04 Jul 2025.

[33] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[34] Bojun Liu, Jordan G Boysen, Ilona Christy Unarta, Xuefeng Du, Yixuan Li, and Xuhui Huang. Exploring transition states of protein conformational changes via out-of-distribution detection in the hyperspherical latent space. *Nature Communications*, 16(1):349, 2025.

[35] Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

[36] Zichen Liu, Wei Zhang, and Tiejun Li. Improving the euclidean diffusion generation of manifold data by mitigating score function singularity. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[37] Yubin Lu, Zhongjian Wang, and Guillaume Bal. Mathematical analysis of singularities in the diffusion model under the submanifold assumption. *East Asian Journal on Applied Mathematics*, 15(4):669–700, 2025.

[38] Ricardo AE Mendes. Diameter and focal radius of submanifolds. *Annals of Global Analysis and Geometry*, 68(2):4, 2025.

[39] Tiberiu Popoviciu. Sur les équations algébriques ayant toutes leurs racines réelles. *Mathematica*, 9(129-145):20, 1935.

[40] Frédéric Poupaud and Michel Rascle. Measure solutions to the linear multi-dimensional transport equation with non-smooth coefficients. *Communications in Partial Differential Equations*, 22(1-2):225–267, 1997.

[41] Saptarshi Roy, Alessandro Rinaldo, and Purnamrita Sarkar. Low-dimensional adaptation of rectified flow: A new perspective through the lens of diffusion and stochastic localization. *arXiv preprint arXiv:2601.15500*, 2026.

[42] Rajesh Sharma, Madhu Gupta, and Girish Kapoor. Some better bounds on the variance with applications. *Journal of Mathematical Inequalities*, 4(3):355–363, 2010.

[43] Jan Pawel Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Diffusion models encode the intrinsic dimension of data manifolds. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 46412–46440. PMLR, 21–27 Jul 2024.

[44] Arthur Stéphanovitch. Lipschitz regularity in flow matching and diffusion models: sharp sampling rates and functional inequalities. *arXiv preprint arXiv:2604.06065*, 2026.

[45] Wong Hiu Tung, Darrick Lee, and Hong Yan. Towards scalable topological regularizers. In *The Thirteenth International Conference on Learning Representations*, 2025.

[46] A. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, 1996.

[47] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

[48] Zhengyu Zhou and Weiwei Liu. An error analysis of flow matching for deep generative modeling. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 78903–78932. PMLR, 13–19 Jul 2025.

A Related work

Flow matching (FM) learns a time-dependent vector field whose ODE transports a simple reference distribution to a target distribution by regressing a conditional velocity along a chosen interpolation path [33]. Closely related formulations appear in stochastic interpolants, which unify deterministic flows and diffusion-style constructions through conditional expectation identities [2, 1]. Our paper focuses on deterministic conditional FM under linear interpolation, emphasizing how the *population* velocity field depends on the effective noise scale and on geometry when the target concentrates near a low-dimensional manifold.

A central element in end-to-end FM theory is stability of the sampling ODE with respect to errors in the learned velocity. In the deterministic setting, the resulting distributional error bounds typically depend on a time-integrated regularity quantity (often an ambient Lipschitz or Jacobian bound) and can therefore deteriorate sharply near terminal time when the target becomes highly concentrated [4]. Existing convergence analyses for FM in Wasserstein distance provide sharp ambient rates under distributional regularity assumptions, but do not directly resolve the thin-noise manifold-concentration regime in which naive ambient controls can become vacuous [18, 48].

On the structural side, recent FM and Gaussian-interpolation-flow analyses highlight that velocity regularity is governed by conditional moments: both $v_t^\star$ and its Jacobian admit exact forms in terms of posterior means and covariances [19, 26]. These identities suggest that bounding $\nabla_x v_t^\star$ in terms of a crude norm purely through score-Hessian operator norms can be overly complicated, since the potentially singular prefactors are coupled to posterior covariance terms that tend to collapse as the effective noise shrinks. However, prior works largely treat such covariance decay as an assumption or verify it under global distribution classes; they do not provide a tube-local geometric verification tailored to manifold-concentrated targets.

The closest FM-side results on intrinsic low-dimensional structure are either restricted to single-chart smooth manifolds or treat the exact manifold-supported case. [27] connect FM to kernel density estimation and obtain intrinsic-dimension improvements on single-chart manifolds, offering insightful precedents though in a comparatively narrow setting. [25] establish the first intrinsic-dimension W_2 rates for FM on smooth manifolds, but their analysis works in the singular $\sigma = 0$ regime with early stopping and assumes a sublinear Jacobian bound on the true velocity, which holds under a semi-log-convexity condition. Our setting instead targets the practically relevant noisy-tube model with $\sigma > 0$, where the population objects are smooth in the ambient space but exhibit highly anisotropic geometry at the tube scale.

In parallel, diffusion-model theory under the manifold scenario has developed precise descriptions of small-noise anisotropy for score-based quantities. Representative results show normal alignment and spectral gaps for the score function [43], quantify score singularities via conditional expectation and small-time expansions [37], and separate geometric versus distributional accuracy requirements [31]; recent work further makes the tangential–normal scale discrepancy explicit and proposes algorithmic mitigations [36]. These papers provide strong evidence that low-noise singularities are structured rather than isotropic, but they primarily analyze the score (or asymptotic regimes) and do not deliver the finite-time, tube-local, pointwise anisotropic envelopes for the *FM population velocity Jacobian* needed for ODE perturbation arguments. We note that a recent discussion also argues that, near terminal time, the relevant sampling dynamics for FM should be governed by a tangential–normal decomposition on the high-probability tube around the data manifold [44].

Finally, several recent works point to directional stability mechanisms in manifold settings, including asymptotic Lyapunov-exponent structure for OT-based FM and robustness induced by tangent-space alignment [17, 9]. Our results complement these perspectives by giving explicit tube-local rate control of $\nabla_x v_t^\star$ and a finite-time tangential–normal block decomposition with quantified envelopes along the FM schedule, which is the form required to replace isotropic terminal blow-up with an anisotropy-aware stability analysis.

B Tube localization proofs

The proofs of Lemmas 2.2 and 2.3 are straightforward.

Proof of Lemma 2.2. By the Gaussian mixture expression (3), conditionally on $X_1 = y + \sigma\xi$ with $Y{=}y \sim f\,\mathrm{Vol}_{\mathcal{M}}$ and $\xi \sim \mathcal{N}(0, I_d)$ independent, we have $X_t \mid Y{=}y \sim \mathcal{N}(ty, \sigma_t^2 I_d)$. Writing $\varphi_s(u) := (2\pi s)^{-d/2}\exp(-\|u\|^2/(2s))$,

$$p_t(x) = \int_{\mathcal{M}} f(y)\,\varphi_{\sigma_t^2}(x - ty)\,\mathrm{dVol}_{\mathcal{M}}(y).$$

The identity $\varphi_{\sigma_t^2}(x - ty) = t^{-d}\,\varphi_{\varepsilon_t^2}(x/t - y)$ with $\varepsilon_t^2 = \sigma_t^2/t^2$ gives

$$p_t(x) = t^{-d}\int_{\mathcal{M}} f(y)\,\varphi_{\varepsilon_t^2}(x/t - y)\,\mathrm{dVol}_{\mathcal{M}}(y) = t^{-d}\,q_{\varepsilon_t}(x/t). \qquad \square$$

Proof of Lemma 2.3. Write $X_1 = Y + \sigma\xi$ with $Y \sim f\,\mathrm{Vol}_{\mathcal{M}}$ and $\xi \sim \mathcal{N}(0, I_d)$ independent. Then $X_t = (1{-}t)X_0 + tX_1 = tY + W_t$ where $W_t := (1{-}t)X_0 + t\sigma\xi \sim \mathcal{N}(0, \sigma_t^2 I_d)$ is independent of Y. Since $Y \in \mathcal{M}$,

$$\mathrm{dist}(X_t, \mathcal{M}) \le \|X_t - Y\| = \|W_t - (1{-}t)Y\| \le \|W_t\| + (1{-}t)R_{\mathcal{M}}.$$

For $Z \sim \mathcal{N}(0, I_d)$, the chi-squared tail bound [28, Lemma 1] gives

$$\mathbb{P}(\|Z\|^2 \ge d + 2\sqrt{du} + 2u) \le e^{-u},$$

and $\sqrt{d + 2\sqrt{du} + 2u} \le \sqrt{d} + \sqrt{2u}$. Setting $u = \log(1/\delta)$ and using $W_t = \sigma_t Z$ yields

$$\mathbb{P}\big(\|W_t\| \ge \sigma_t(\sqrt{d} + \sqrt{2\log(1/\delta)})\big) \le \delta.$$

Hence $\mathbb{P}(X_t \notin \mathcal{M}_{r(t,\delta)}) \le \delta$ with $r(t,\delta) = (1{-}t)R_{\mathcal{M}} + \sigma_t(\sqrt{d} + \sqrt{2\log(1/\delta)})$. $\square$

C Small-σ calculus proofs

This appendix section proves the late-regime small-noise expansion in three steps. We first isolate the analytic Laplace ingredient in a parameter-uniform Gaussian integral lemma and then verify its conditions in a fixed finite atlas, yielding local rescaled amplitude bounds for the tube-convolution density. We next glue these chartwise bounds into a global Fermi-coordinate expansion on a master tube, where the singular term is exactly the Gaussian fiber factor $\exp(-\|\rho\|^2/(2\varepsilon^2))$ and the remaining geometry enters through $\det(I - S_\rho(y))$. Finally, we transfer the Fermi-coordinate expansion back to ambient derivatives using the tube map and the Hessian of the squared-distance function, and then apply the exact FM rescaling to obtain Proposition 4.2. In this section, C in upper bounds represents a constant which may vary from line to line.

Lemma C.1. *If $\|\rho\| < r < \tau/4$, then $I - S_\rho(y)$ is positive definite and*

$$\left\|(I - S_\rho(y))^{-1}\right\|_{\mathrm{op}} \le \frac{1}{1 - \kappa\|\rho\|} \le \frac{4}{3}, \tag{24}$$

as well as

$$\left\|S_\rho(y)(I - S_\rho(y))^{-1}\right\|_{\mathrm{op}} \le \frac{\kappa\|\rho\|}{1 - \kappa\|\rho\|} \le \frac{4}{3}\kappa\|\rho\|. \tag{25}$$

Proof. By $\|S_\rho(y)\|_{\mathrm{op}} \le \kappa\|\rho\|$ and $\kappa \le 1/\tau$,

$$\|S_\rho(y)\|_{\mathrm{op}} \le \kappa\|\rho\| < \frac{1}{4}.$$

Therefore every eigenvalue of $I - S_\rho(y)$ is at least $1 - \kappa\|\rho\|$, which gives (24). The bound (25) follows from

$$\left\|S_\rho(y)(I - S_\rho(y))^{-1}\right\|_{\mathrm{op}} \le \|S_\rho(y)\|_{\mathrm{op}} \left\|(I - S_\rho(y))^{-1}\right\|_{\mathrm{op}}.$$

This completes the proof. □

Lemma C.2 (Parameter-uniform Gaussian amplitude bound)**.** *Let $K \subset \mathbb{R}^m$ be compact, let $Q \in C^2(K; \mathrm{Sym}_k)$, and assume that*

$$Q(\theta) \succeq c_0 I_k, \qquad \theta \in K, \tag{26}$$

for some $c_0 > 0$. Fix $\eta < c_0/2$ and an integer $M \ge 0$. Here and below, for a multiindex $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^m$, we write $|\alpha| := \alpha_1 + \cdots + \alpha_m$. For $0 < \varepsilon \le \varepsilon_0$, let $B_\varepsilon : K \times \mathbb{R}^k \to \mathbb{R}$ be such that, for every multiindex α with $|\alpha| \le 2$,

$$\sup_{0<\varepsilon\le\varepsilon_0} \sup_{\theta\in K} |\partial_\theta^\alpha B_\varepsilon(\theta, w)| \le C_\alpha (1 + \|w\|)^M e^{\eta\|w\|^2}, \qquad w \in \mathbb{R}^k, \tag{27}$$

and $(\theta, w) \mapsto \partial_\theta^\alpha B_\varepsilon(\theta, w)$ is continuous. Define

$$A_\varepsilon(\theta) := \int_{\mathbb{R}^k} \exp\Big(-\frac{1}{2}\langle w, Q(\theta)w\rangle\Big) B_\varepsilon(\theta, w)\,\mathrm{d}w. \tag{28}$$

Then $A_\varepsilon \in C^2(K)$ for every ε, and for each multiindex $|\alpha| \le 2$,

$$\partial_\theta^\alpha A_\varepsilon(\theta) = \int_{\mathbb{R}^k} \partial_\theta^\alpha \left[\exp\Big(-\frac{1}{2}\langle w, Q(\theta)w\rangle\Big) B_\varepsilon(\theta, w)\right]\mathrm{d}w. \tag{29}$$

Moreover,

$$\sup_{0<\varepsilon\le\varepsilon_0} \sup_{\theta\in K} |\partial_\theta^\alpha A_\varepsilon(\theta)| \le C'_\alpha \int_{\mathbb{R}^k} (1 + \|w\|)^{M+2|\alpha|} e^{-(c_0/2-\eta)\|w\|^2}\,\mathrm{d}w < \infty. \tag{30}$$

If in addition

$$B_\varepsilon(\theta, w) = B_0(\theta, w) + \varepsilon R_\varepsilon(\theta, w) \tag{31}$$

and the family R_ε obeys the same bounds (27)*, then*

$$\|A_\varepsilon - A_0\|_{C^2(K)} \le C\varepsilon. \tag{32}$$

Proof. By (26),

$$\exp\Big(-\frac{1}{2}\langle w, Q(\theta)w\rangle\Big) \le e^{-c_0\|w\|^2/2}$$

uniformly in $\theta \in K$. Differentiating the Gaussian factor in θ brings down only polynomial factors in w: for example,

$$\partial_{\theta_i} \exp\Big(-\frac{1}{2}\langle w, Q(\theta)w\rangle\Big) = -\frac{1}{2}\langle w, (\partial_{\theta_i} Q(\theta))w\rangle \exp\Big(-\frac{1}{2}\langle w, Q(\theta)w\rangle\Big),$$

and second derivatives are finite sums of terms bounded by

$$C(1 + \|w\|)^4 e^{-c_0\|w\|^2/2}.$$

Combining this with (27), every θ-derivative of the integrand in (28) up to order 2 is bounded by

$$C(1+\|w\|)^{M+2|\alpha|}e^{-(c_0/2-\eta)\|w\|^2},$$

which is integrable because $\eta < c_0/2$. Dominated convergence therefore justifies (29), and the same domination gives the uniform bound (30) and continuity of the derivatives.

If (31) holds, then $A_\varepsilon - A_0$ is the integral of the same Gaussian factor against $\varepsilon R_\varepsilon$. Applying (30) to that difference yields (32). □

Remark C.3 (Master-tube convention for the small-noise calculus). Fix once and for all a master radius $r_0 \in (0, \tau/4)$. In the remainder of Section C, the constants and auxiliary objects produced by the local small-noise analysis, such as ε_0, C_{sc}, the fixed atlas, the chart radius δ, the cutoff χ, and the composition bounds, are chosen on the master tube $\mathcal{M}_{r_0}$ (equivalently on K_{r_0} in Fermi coordinates). They may depend on r_0, $\mathcal{M}$ and f, but they are uniform for every smaller analysis radius $r \in (0, r_0]$; later statements simply restrict these master-tube bounds from $\mathcal{M}_{r_0}$ to $\mathcal{M}_r$.

Proposition C.4 (Fixed-atlas rescaled amplitude bounds)**.** *Assume Assumptions 1 and 2, and fix a master radius $r_0 < \tau/4$. Then there exist:*

- *(i) finitely many chart maps $Y_j : U_j \to \mathcal{M}$, open sets $U_j' \Subset U_j$ (where $U' \Subset U$ denotes $\overline{U'} \subset U$ with $\overline{U'}$ compact) covering $\mathcal{M}$, and smooth orthonormal frame fields $E_j(a) : \mathbb{R}^k \to T_{Y_j(a)}\mathcal{M}$, $N_j(a) : \mathbb{R}^{d-k} \to N_{Y_j(a)}\mathcal{M}$ on U_j;*
- *(ii) a radius $\delta > 0$, a cutoff $\chi \in C_c^\infty(\mathbb{R}^k)$ with $\chi \equiv 1$ on $B_\delta(0)$ and $\mathrm{supp}\chi \subset B_{2\delta}(0)$, and $\varepsilon_0 > 0$;*
- *(iii) maps $h_j \in C^3(U_j \times B_{2\delta}(0); \mathbb{R}^{d-k})$*

such that, for every j, the local graph map

$$z_{j,a}(u) := Y_j(a) + E_j(a)u + N_j(a)h_j(a,u) \tag{33}$$

satisfies

$$h_j(a,0) = 0, \qquad D_u h_j(a,0) = 0, \qquad D_u^2 h_j(a,0) = \mathrm{II}_{Y_j(a)} \tag{34}$$

in the chosen frames, and all mixed (a,u)-derivatives of h_j up to total order 3 are uniformly bounded.

All of these objects, together with the constants $c, \eta, C_\alpha, M_\alpha$ in (42)–(44), may be chosen depending only on r_0, $\mathcal{M}$ and f. For every analysis radius $r \in (0, r_0]$, every $a \in U_j'$, every $\xi \in \mathbb{R}^{d-k}$ with $\|\xi\| < r$, and

$$x = \Phi_j(a,\xi) := Y_j(a) + N_j(a)\xi, \tag{35}$$

define $\rho_j(a,\xi) := N_j(a)\xi \in N_{Y_j(a)}\mathcal{M}$. Then for every $\varepsilon \in (0, \varepsilon_0]$,

$$q_\varepsilon(x) = (2\pi\varepsilon^2)^{-(d-k)/2} e^{-\|\xi\|^2/(2\varepsilon^2)} \Big(A_{\varepsilon,j}^{\mathrm{loc}}(a,\xi) + \widetilde{\mathcal{E}}_{\varepsilon,j}^{\mathrm{far}}(a,\xi) \Big), \tag{36}$$

where

$$A_{\varepsilon,j}^{\mathrm{loc}}(a,\xi) = (2\pi)^{-k/2} \int_{\mathbb{R}^k} \exp\Big(-\frac{1}{2} \big\langle w, (I - S_{\rho_j(a,\xi)}(Y_j(a)))w \big\rangle \Big) B_{\varepsilon,j}(a,\xi,w)\, \mathrm{d}w \tag{37}$$

with

$$B_{\varepsilon,j}(a,\xi,w) = \chi(\varepsilon w)\, f(z_{j,a}(\varepsilon w))\, J_j(a,\varepsilon w)\, e^{\Psi_{\varepsilon,j}(a,\xi,w)}. \tag{38}$$

Here

$$J_j(a,u) := \sqrt{\det(I_k + D_u h_j(a,u)^* D_u h_j(a,u))} \tag{39}$$

and

$$\begin{aligned} \Psi_{\varepsilon,j}(a,\xi,w) = \varepsilon \langle \xi, H_{3,j}(a,w,\varepsilon) \rangle - \frac{\varepsilon^2}{8} \big\| \mathrm{II}_{Y_j(a)}[w,w] \big\|^2 \\ - \frac{\varepsilon^3}{2} \big\langle \mathrm{II}_{Y_j(a)}[w,w], H_{3,j}(a,w,\varepsilon) \big\rangle - \frac{\varepsilon^4}{2} \|H_{3,j}(a,w,\varepsilon)\|^2, \end{aligned} \tag{40}$$

with

$$H_{3,j}(a,w,\varepsilon) := \frac{1}{2} \int_0^1 (1-t)^2 D_u^3 h_j(a, t\varepsilon w)[w,w,w]\, \mathrm{d}t. \tag{41}$$

Moreover:

$$\left|\partial^{\alpha}_{(a,\xi)}\widetilde{\mathcal{E}}^{\mathrm{far}}_{\varepsilon,j}(a,\xi)\right| \le C_\alpha e^{-c/\varepsilon^2}, \qquad |\alpha| \le 2, \tag{42}$$

$$\left|\partial^{\alpha}_{(a,\xi)}B_{\varepsilon,j}(a,\xi,w)\right| \le C_\alpha(1+\|w\|)^{M_\alpha}e^{\eta\|w\|^2}, \qquad |\alpha| \le 2, \tag{43}$$

$$B_{\varepsilon,j}(a,\xi,w) = f(Y_j(a)) + \varepsilon R_{\varepsilon,j}(a,\xi,w), \tag{44}$$

where the remainders $R_{\varepsilon,j}$ obey the same weighted bounds as in (43)*.*

Proof. By compactness of $\mathcal{M}$ and C^3 regularity, we may choose finitely many chart maps $Y_j : U_j \to \mathcal{M}$, smaller precompact sets $U_j' \Subset U_j$ covering $\mathcal{M}$, and C^2 orthonormal tangent and normal frames E_j, N_j on each U_j [29, Lemma 1.10]. After shrinking the charts and taking a minimum over finitely many patches, there exists $\delta > 0$ such that for every j and every $a \in U_j$, the manifold near $Y_j(a)$ is the graph (33) of a map $h_j(a,\cdot)$ on $B_{2\delta}(0)$ with the Taylor conditions (34). Uniform bounds on the mixed (a,u)-derivatives of h_j up to order 3 follow from compactness of $\overline{U_j'} \times \overline{B_{2\delta}(0)}$.

Fix an analysis radius $r \in (0, r_0]$, a chart index j, $a \in U_j'$, and $\|\xi\| < r$, and set $y := Y_j(a)$, $\rho := N_j(a)\xi$, $x := y + \rho$. By construction,

$$x - z_{j,a}(u) = \rho - E_j(a)u - N_j(a)h_j(a,u),$$

and the tangent and normal summands are orthogonal. Therefore

$$\|x - z_{j,a}(u)\|^2 = \|\rho\|^2 + \|u\|^2 - 2\langle \xi, h_j(a,u)\rangle + \|h_j(a,u)\|^2.$$

Using the Taylor expansion

$$h_j(a,u) = \frac{1}{2}\mathrm{II}_{Y_j(a)}[u,u] + R_{3,j}(a,u),$$

with

$$R_{3,j}(a,u) = \frac{1}{2}\int_0^1 (1-t)^2 D_u^3 h_j(a,tu)[u,u,u]\,\mathrm{d}t,$$

and then rescaling $u = \varepsilon w$, we obtain

$$\frac{\|x - z_{j,a}(\varepsilon w)\|^2}{2\varepsilon^2} = \frac{\|\xi\|^2}{2\varepsilon^2} + \frac{1}{2}\langle w, (I - S_\rho(y))w\rangle - \Psi_{\varepsilon,j}(a,\xi,w), \tag{45}$$

with $\Psi_{\varepsilon,j}$ as in (40). The graph Jacobian is exactly (39). On the local graph patch, the change of variables $u = \varepsilon w$ gives

$$\mathrm{dVol}_{\mathcal{M}}(z_{j,a}(u)) = J_j(a,u)\,\mathrm{d}u = J_j(a,\varepsilon w)\,\varepsilon^k\,\mathrm{d}w.$$

Therefore the Gaussian prefactor transforms as

$$(2\pi\varepsilon^2)^{-d/2}\varepsilon^k = (2\pi\varepsilon^2)^{-(d-k)/2}(2\pi)^{-k/2},$$

and substituting this together with (45) and (39) into the local part of the convolution integral gives (36)–(38).

To define the far term, split the manifold integral into the local graph patch $\{z_{j,a}(u) : \|u\| < 2\delta\}$ and its complement. For each chart index j, define the gap function

$$g_j(a,\xi,z) := \|Y_j(a) + N_j(a)\xi - z\|^2 - \|\xi\|^2,$$

and the parameter set

$$\Gamma_j := \Big\{(a,\xi,z) : a \in \overline{U_j'},\ \|\xi\| \le r_0,\ z \in \mathcal{M},\ z \notin z_{j,a}(B_{2\delta}(0))\Big\}.$$

Since $\mathcal{M}$ is compact and $(a,u) \mapsto z_{j,a}(u)$ is continuous, the set Γ_j is compact. Moreover, if $(a,\xi,z) \in \Gamma_j$ and $x := Y_j(a) + N_j(a)\xi$, then $\|\xi\| \le r_0 < \tau$, so x lies in the tubular neighborhood where $Y_j(a)$ is the unique nearest point on $\mathcal{M}$. Because $z \notin z_{j,a}(B_{2\delta}(0))$, in particular $z \ne Y_j(a)$, hence

$$g_j(a,\xi,z) = \|x - z\|^2 - \mathrm{dist}(x,\mathcal{M})^2 > 0.$$

Therefore g_j attains a strictly positive minimum on Γ_j; writing

$$2c_{0,j} := \min_{\Gamma_j} g_j > 0,$$

and then taking $c_0 := \min_j c_{0,j} > 0$ over the finite atlas, we obtain

$$\|x - z\|^2 \geq \|\xi\|^2 + 2c_0$$

uniformly in j, $a \in U_j'$, $\|\xi\| \leq r_0$, and every z outside the local graph patch. Shrinking the analysis radius from r_0 to r only restricts the admissible set of ξ values, so the same c_0 works for every $r \leq r_0$. Differentiating the factored far integral with respect to (a, ξ) up to order 2 differentiates only the phase $\|x - z\|^2/(2\varepsilon^2)$ and the smooth chart amplitudes, hence produces factors bounded by $C_\alpha \varepsilon^{-4}$ times the same Gaussian. Thus

$$\left|\partial^\alpha_{(a,\xi)} \widetilde{\mathcal{E}}^{\mathrm{far}}_{\varepsilon,j}(a, \xi)\right| \leq C_\alpha \varepsilon^{-4} e^{-c_0/\varepsilon^2}, \qquad |\alpha| \leq 2.$$

After reducing c_0 slightly, the factor ε^{-4} is absorbed into the exponential, yielding (42).

On the support of $\chi(\varepsilon w)$ one has $\varepsilon \|w\| \leq 2\delta$. The uniform C^3 bounds on h_j imply

$$\|H_{3,j}(a, w, \varepsilon)\| \leq C \|w\|^3,$$

and the same type of bound holds for (a, ξ)-derivatives up to order 2. Since also $\|\xi\| \leq r \leq r_0$, the first term in (40) is bounded by

$$|\varepsilon \langle \xi, H_{3,j}(a, w, \varepsilon)\rangle| \leq C r_0 \varepsilon \|w\|^3.$$

The remaining terms are bounded by $C\varepsilon^2\|w\|^4$, $C\varepsilon^3\|w\|^5$, and $C\varepsilon^4\|w\|^6$, respectively. Using $\varepsilon\|w\| \leq 2\delta$, each of these terms is at most $C_{r_0,\delta}\|w\|^2$, and hence

$$|\Psi_{\varepsilon,j}(a, \xi, w)| \leq C_{r_0}(\varepsilon \|w\|^3 + \varepsilon^2 \|w\|^4 + \varepsilon^3 \|w\|^5 + \varepsilon^4 \|w\|^6) \leq C_{r_0}(2\delta + 4\delta^2 + 8\delta^3 + 16\delta^4) \|w\|^2.$$

Choosing δ smaller if necessary makes the right-hand side at most $\eta\|w\|^2$. This choice depends only on the fixed atlas and the master radius r_0, not on the analysis radius. Differentiating $B_{\varepsilon,j}$ in (a, ξ) therefore produces only polynomial factors in w, times $\exp(\eta\|w\|^2)$, proving (43). The resulting admissible ε_0 is therefore determined by the master-tube data and remains valid for every smaller radius $r \leq r_0$.

It remains to prove (44). Write

$$B_{\varepsilon,j}(a, \xi, w) - f(Y_j(a)) = T_f + T_J + T_\Psi,$$

where

$$T_f := \chi(\varepsilon w)\big(f(z_{j,a}(\varepsilon w)) - f(Y_j(a))\big),$$
$$T_J := \chi(\varepsilon w) f(Y_j(a))\big(J_j(a, \varepsilon w) - 1\big),$$

and

$$T_\Psi := \chi(\varepsilon w) f(Y_j(a)) J_j(a, \varepsilon w)\big(e^{\Psi_{\varepsilon,j}(a,\xi,w)} - 1\big).$$

Since $z_{j,a}(\varepsilon w) - Y_j(a) = O(\varepsilon \|w\|)$ and $\log f \in C^2$, the mean-value theorem gives

$$T_f = \varepsilon R^{(f)}_{\varepsilon,j}(a, \xi, w),$$

with $R^{(f)}_{\varepsilon,j}$ obeying the weighted bounds (43). Likewise, $D_u h_j(a, 0) = 0$ and the uniform C^2 bound on h_j imply

$$D_u h_j(a, \varepsilon w) = O(\varepsilon \|w\|)$$

on the rescaled support, hence

$$J_j(a, \varepsilon w) = 1 + O(\varepsilon^2 \|w\|^2),$$

so $T_J = \varepsilon R^{(J)}_{\varepsilon,j}$ with the same weighted control. Finally,

$$\frac{e^{\Psi_{\varepsilon,j}} - 1}{\varepsilon} = \frac{e^{\Psi_{\varepsilon,j}} - 1}{\Psi_{\varepsilon,j}} \cdot \frac{\Psi_{\varepsilon,j}}{\varepsilon},$$

and the first factor is bounded by $\exp(\eta\|w\|^2)$ while the second is

$$O(\|w\|^3 + \varepsilon \|w\|^4)$$

on the rescaled support. Hence $T_\Psi = \varepsilon R^{(\Psi)}_{\varepsilon,j}$ with the same weighted bounds. Summing the three pieces proves (44). □

Proposition C.5 (Fermi-coordinate expansion from fixed-atlas weighted bounds)**.** *Assume Assumptions 1 and 2. Fix a master radius $r_0 < \tau/4$ and write*

$$K_{r_0} := \{(y, \rho) : y \in \mathcal{M},\ \rho \in N_y\mathcal{M},\ \|\rho\| < r_0\}.$$

Then there exist $\varepsilon_0 > 0$, a strictly positive function $A_0 \in C^2(K_{r_0})$, and remainders $R_\varepsilon \in C^2(K_{r_0})$ such that for every analysis radius $r \in (0, r_0]$, all $(y, \rho) \in K_r$, and all $\varepsilon \in (0, \varepsilon_0]$,

$$q_\varepsilon(y + \rho) = (2\pi\varepsilon^2)^{-(d-k)/2} e^{-\|\rho\|^2/(2\varepsilon^2)} \big(A_0(y, \rho) + R_\varepsilon(y, \rho)\big), \tag{46}$$

$$A_0(y, \rho) = f(y)\ \det(I - S_\rho(y))^{-1/2}, \tag{47}$$

$$\|R_\varepsilon\|_{C^2(K_{r_0})} \le C\varepsilon. \tag{48}$$

Consequently, with

$$\tilde{\ell}_\varepsilon(y, \rho) := \log q_\varepsilon(y + \rho), \tag{49}$$

one has

$$\tilde{\ell}_\varepsilon(y, \rho) = -\frac{\|\rho\|^2}{2\varepsilon^2} - \frac{d-k}{2}\log(2\pi\varepsilon^2) + \log f(y) - \frac{1}{2}\log\det(I - S_\rho(y)) + \widetilde{R}_\varepsilon(y, \rho), \tag{50}$$

with

$$\left\|\widetilde{R}_\varepsilon\right\|_{C^2(K_{r_0})} \le C\varepsilon. \tag{51}$$

Proof. Choose the finite atlas, frames, and chartwise decomposition from Proposition C.4. On each chart, apply Lemma C.2 with parameter $\theta = (a, \xi)$ and

$$Q(\theta) = I - S_{\rho_j(a,\xi)}(Y_j(a)).$$

By $\|S_\rho(y)\|_{\mathrm{op}} \le \kappa\|\rho\|$ and $\|\rho_j(a, \xi)\| = \|\xi\| < r_0 < \tau/4$, $Q(\theta) \succeq (1 - \kappa r_0)I_k \succeq \frac{3}{4}I_k$, so the ellipticity condition holds uniformly on the master tube. The weighted bound (43) and expansion (44) therefore imply

$$A^{\mathrm{loc}}_{\varepsilon,j}(a, \xi) = A_{0,j}(a, \xi) + O(\varepsilon) \quad \text{in } C^2,$$

where

$$A_{0,j}(a, \xi) = (2\pi)^{-k/2} f(Y_j(a)) \int_{\mathbb{R}^k} \exp\Big(-\frac{1}{2}\left\langle w, (I - S_{\rho_j(a,\xi)}(Y_j(a)))w\right\rangle\Big)\, \mathrm{d}w.$$

Evaluating the Gaussian integral gives

$$A_{0,j}(a, \xi) = f(Y_j(a))\det(I - S_{\rho_j(a,\xi)}(Y_j(a)))^{-1/2}.$$

This is exactly the Euclidean tube Jacobian factor identified by Gray [21, Lemma 3.14] in our present notation. Since the eigenvalues of $I - S_\rho(y)$ lie in $[3/4, 5/4]$ on K_{r_0}, one has

$$(3/4)^k \le \det(I - S_\rho(y)) \le (5/4)^k$$

uniformly on K_{r_0}. Since f is strictly positive, the function A_0 defined by (47) belongs to $C^2(K_{r_0})$ and satisfies the uniform lower bound

$$A_0(y, \rho) \ge f_{\min}(5/4)^{-k/2} > 0, \qquad (y, \rho) \in K_{r_0}.$$

The far term (42) is exponentially small in C^2, hence in particular $O(\varepsilon)$ in C^2 after reducing ε_0. Therefore the chartwise quantities glue to a global remainder R_ε satisfying (48), and (46) follows.

Finally, since A_0 is bounded below by a positive constant on the compact set K_{r_0}, the function $1/A_0$ is uniformly bounded in $C^2(K_{r_0})$. After reducing ε_0 if necessary, (48) gives

$$A_0(y, \rho) + R_\varepsilon(y, \rho) \ge \frac{1}{2}\inf_{K_{r_0}} A_0 > 0,$$

so

$$\widetilde{R}_\varepsilon = \log\Big(1 + \frac{R_\varepsilon}{A_0}\Big)$$

is well defined. The quotient rule and (48) imply

$$\|R_\varepsilon/A_0\|_{C^2(K_{r_0})} \le C\varepsilon,$$

and smooth composition with $\log(1 + s)$ on a fixed neighborhood of $s = 0$ yields (51). Restricting from K_{r_0} to any smaller $K_r \subset K_{r_0}$ gives the stated formulas on every analysis radius $r \in (0, r_0]$. This gives (50). □

Proposition C.6 (Tube-map differential and squared-distance Hessian)**.** *Fix $r < \tau/4$. The tube map*

$$F : N_r\mathcal{M} \to \mathcal{M}_r, \qquad F(y,\rho) = y + \rho,$$

is a C^2 diffeomorphism with inverse $G = (\pi_{\mathcal{M}}, \rho_{\mathcal{M}})$. Identify

$$T_{(y,\rho)}(N\mathcal{M}) \cong T_y\mathcal{M} \oplus N_y\mathcal{M}$$

via the splitting induced by the normal connection. Then for $x = y + \rho \in \mathcal{M}_r$, $u \in T_y\mathcal{M}$, and $\eta \in N_y\mathcal{M}$, one has

$$DF_{(y,\rho)}(u,\eta) = (I - S_\rho(y))u + \eta. \tag{52}$$

Consequently,

$$D(\rho_{\mathcal{M}})_x[v] = -S_\rho(y)(I - S_\rho(y))^{-1}P_T(y)v + P_N(y)v, \tag{53}$$

and if

$$\phi(x) := \frac{1}{2}\delta_{\mathcal{M}}(x)^2 = \frac{1}{2}\left\|\rho_{\mathcal{M}}(x)\right\|^2, \tag{54}$$

then

$$\nabla\phi(x) = \rho_{\mathcal{M}}(x) \tag{55}$$

and

$$P_T(y)\,\nabla^2\phi(x)\,P_T(y) = -S_\rho(y)(I - S_\rho(y))^{-1}, \tag{56}$$

$$P_T(y)\,\nabla^2\phi(x)\,P_N(y) = 0, \tag{57}$$

$$P_N(y)\,\nabla^2\phi(x)\,P_N(y) = P_N(y). \tag{58}$$

Proof. Foote's regularity theorem for the distance function near a C^3 submanifold [16, Theorem 1], together with the tubular-neighborhood theorem [23, Theorem 5.1][14, Theorem 4.8(13)], implies that for every $r < \tau$ the map F is a C^2 diffeomorphism from $N_r\mathcal{M}$ onto $\mathcal{M}_r$. Fix $(y,\rho) \in N_r\mathcal{M}$.

To compute the differential, choose a smooth curve $t \mapsto (y(t), \rho(t))$ in $N_r\mathcal{M}$ such that

$$y(0) = y, \qquad y'(0) = u, \qquad \nabla^\perp_t\rho(0) = \eta,$$

where $\nabla^\perp$ is the normal connection. The Weingarten formula [11, Chapter 1] gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t)\Big|_{t=0} = -S_\rho(y)u + \eta,$$

so for $x(t) := F(y(t),\rho(t)) = y(t) + \rho(t)$,

$$x'(0) = y'(0) + \rho'(0) = (I - S_\rho(y))u + \eta,$$

which is (52).

Since $T_y\mathcal{M}$ and $N_y\mathcal{M}$ are orthogonal and $I - S_\rho(y)$ maps $T_y\mathcal{M}$ to itself, (52) implies that every $v \in \mathbb{R}^d$ decomposes uniquely as

$$v = (I - S_\rho(y))u + \eta, \qquad u = (I - S_\rho(y))^{-1}P_T(y)v, \qquad \eta = P_N(y)v.$$

Differentiating the ambient normal vector field $\rho_{\mathcal{M}}(x) = \rho$ along the same curve therefore yields (53).

Federer [14, Theorem 4.8(5)] gives $\nabla\phi(x) = x - \pi_{\mathcal{M}}(x) = \rho_{\mathcal{M}}(x)$, proving (55). Differentiating once more and using (53) gives $\nabla^2\phi(x) = D(\rho_{\mathcal{M}})_x$. Projecting that operator onto the $T_y\mathcal{M} \oplus N_y\mathcal{M}$ splitting yields (56)–(58). □

Corollary C.7 (Fixed-atlas tubular composition control)**.** *Let $Y : U \to \mathcal{M}$ be a C^3 chart and choose a C^2 orthonormal normal frame $\upsilon_1, \dots, \upsilon_{d-k}$ along $Y(U)$. Writing*

$$N(a)\xi := \sum_{\alpha=1}^{d-k} \xi^\alpha \upsilon_\alpha(a), \qquad \Phi(a,\xi) := Y(a) + N(a)\xi,$$

there exist $U' \Subset U$ and $r' > 0$ such that

$$\Phi : U' \times B^{d-k}_{r'} \to V'$$

is a C^2 diffeomorphism onto an open tube patch $V' \subset \mathcal{M}_r$. If $G = \Phi^{-1} = (a,\xi)$, then DG and D^2G are bounded on V'. Consequently, for every $b \in C^2(U' \times B^{d-k}_{r'})$,

$$\|b \circ G\|_{C^2(V')} \le C_\Phi \|b\|_{C^2(U' \times B^{d-k}_{r'})}. \tag{59}$$

Proof. Follows from Proposition C.6 by the chain rule, which expresses the first and second derivatives of $b \circ G$ as sums of products of derivatives of b up to order 2 with DG and D^2G. □

Proposition C.8 (Ambient small-noise property from the Fermi-coordinate expansion)**.** *Assume the conditions of Proposition C.5 on a master tube K_{r_0} with $r_0 < \tau/4$. Then there exist $\varepsilon_0, C_{\mathrm{sc}} > 0$, depending only on r_0, $\mathcal{M}$ and f, such that for every analysis radius $r \in (0, r_0]$, every $\varepsilon \in (0, \varepsilon_0]$, and every $x \in \mathcal{M}_r$, writing $G(x) = (y, \rho)$,*

$$\left\|\nabla \log q_\varepsilon(x) + \rho/\varepsilon^2\right\| \le C_{\mathrm{sc}}, \tag{60}$$

$$\left\|P_T(y)\,\nabla^2 \log q_\varepsilon(x)\,P_T(y) - \varepsilon^{-2} S_\rho(y)(I - S_\rho(y))^{-1}\right\|_{\mathrm{op}} \le C_{\mathrm{sc}}, \tag{61}$$

$$\left\|P_T(y)\,\nabla^2 \log q_\varepsilon(x)\,P_N(y)\right\|_{\mathrm{op}} \le C_{\mathrm{sc}}, \tag{62}$$

$$\left\|P_N(y)\,\nabla^2 \log q_\varepsilon(x)\,P_N(y) + \varepsilon^{-2} P_N(y)\right\|_{\mathrm{op}} \le C_{\mathrm{sc}}. \tag{63}$$

Proof. By Proposition C.5,

$$\log q_\varepsilon(x) = -\frac{\phi(x)}{\varepsilon^2} - \frac{d-k}{2}\log(2\pi\varepsilon^2) + a(x) + b_\varepsilon(x),$$

where $\phi(x) = \delta_{\mathcal{M}}(x)^2/2$,

$$a(x) := \log f(\pi_{\mathcal{M}}(x)) - \frac{1}{2}\log\det(I - S_{\rho_{\mathcal{M}}(x)}(\pi_{\mathcal{M}}(x))),$$

and

$$b_\varepsilon(x) := \widetilde{R}_\varepsilon(\pi_{\mathcal{M}}(x), \rho_{\mathcal{M}}(x)).$$

Since f is positive with $\log f \in C^2(\mathcal{M})$, the map $(y, \rho) \mapsto \log f(y)$ is C^2 on K_r. The Jacobian factor is also C^2 on K_r, and Proposition C.6 gives that $G = (\pi_{\mathcal{M}}, \rho_{\mathcal{M}})$ is C^2 on $\mathcal{M}_r$. Hence $a \in C^2(\mathcal{M}_r)$ with $\|a\|_{C^2(\mathcal{M}_r)} \le C$.

Likewise, Proposition C.5 gives $\|\widetilde{R}_\varepsilon\|_{C^2(K_{r_0})} \le C\varepsilon$, and restricting to $K_r \subset K_{r_0}$ yields the same bound on K_r. The fixed-atlas composition control (59), with constants chosen on the master tube by Remark C.3, therefore yields

$$\|b_\varepsilon\|_{C^2(\mathcal{M}_r)} \le C\varepsilon.$$

Therefore

$$\nabla \log q_\varepsilon(x) = -\frac{\nabla\phi(x)}{\varepsilon^2} + O(1), \qquad \nabla^2 \log q_\varepsilon(x) = -\frac{\nabla^2\phi(x)}{\varepsilon^2} + O(1),$$

uniformly on $\mathcal{M}_r$. Applying Proposition C.6, specifically (55) and (56)–(58), gives (60)–(63). □

Proof of Proposition 4.2. Set $u := x/t$ and $\varepsilon := \varepsilon_t = \sigma_t/t$. By Lemma 2.2,

$$\nabla_x \log p_t(x) = \frac{1}{t}\,\nabla \log q_\varepsilon(u), \qquad \nabla_x^2 \log p_t(x) = \frac{1}{t^2}\,\nabla^2 \log q_\varepsilon(u).$$

Applying Proposition C.8 at the point u gives

$$\nabla \log q_\varepsilon(u) = -\frac{\rho_t}{\varepsilon^2} + \widetilde{b}_t(x), \qquad \left\|\widetilde{b}_t(x)\right\| \le C_{\mathrm{sc}}.$$

Since $t \ge t^\star = 1/(1+\sigma)$ and Assumption 3 gives $\sigma \le 1$, one has $t \ge t^\star \ge 1/2$. Enlarging C_{sc} if necessary, division by t yields (13). Likewise,

$$\left\|P_T(y_t)\,\nabla_x^2 \log p_t(x)\,P_T(y_t) - \sigma_t^{-2} S_{\rho_t}(y_t)(I - S_{\rho_t}(y_t))^{-1}\right\|_{\mathrm{op}} \le \frac{C_{\mathrm{sc}}}{t^2},$$

and $t \ge 1/2$ implies (14) after enlarging C_{sc} once more if needed. The mixed and normal-normal bounds are analogous. □

Proposition C.9. *Assume Assumptions 1–3, and fix $0 < r \le r_0 < \tau/4$. Then there exist constants $\varepsilon_0 > 0$ and $C < \infty$, depending only on r_0, $\mathcal{M}$ and f, such that for every $t \in [t^\star, 1]$ with $\varepsilon_t \le \varepsilon_0$ and every*

$$x \in \mathcal{M}_{t,r}^{\mathrm{sc}} := \{z \in \mathbb{R}^d : z/t \in \mathcal{M}_r\},$$

one has

$$\|\nabla \log p_t(x)\| \le C\Big(\frac{\|\rho_t\|}{\sigma_t^2} + 1\Big), \tag{64}$$

$$\big\|\nabla_x^2 \log p_t(x)\big\|_{\mathrm{op}} \le C\Big(\frac{1}{\sigma_t^2} + \frac{\kappa\,\|\rho_t\|}{\sigma_t^2} + 1\Big), \tag{65}$$

where $(y_t, \rho_t) = (\pi_{\mathcal{M}}(x/t), \rho_{\mathcal{M}}(x/t))$. In particular, on points with $\|\rho_t\| \le C_0\sigma_t$, the score is $O(1/\sigma_t)$ and the Hessian is $O(1/\sigma_t^2)$.

Proof. Eq. (64) follows immediately from (13).

For the Hessian, write $H_t(x) := \nabla_x^2 \log p_t(x)$ and decompose it into its TT, TN, NT, and NN blocks in the orthogonal splitting $T_{y_t}\mathcal{M} \oplus N_{y_t}\mathcal{M}$. Proposition 4.2 and inequality $\|S_\rho(y)\|_{\mathrm{op}} \le \kappa\,\|\rho\|$ give

$$\begin{aligned}\|H_{TT}\|_{\mathrm{op}} \le& \big\|H_{TT} - \sigma_t^{-2} S_{\rho_t}(y_t)(I - S_{\rho_t}(y_t))^{-1}\big\|_{\mathrm{op}} + \sigma_t^{-2}\,\big\|S_{\rho_t}(y_t)(I - S_{\rho_t}(y_t))^{-1}\big\|_{\mathrm{op}} \\ \le& C + \frac{4\kappa\,\|\rho_t\|}{3\sigma_t^2}\end{aligned}$$

by Lemma C.1. The mixed blocks satisfy

$$\|H_{TN}\|_{\mathrm{op}} + \|H_{NT}\|_{\mathrm{op}} \le 2C,$$

and the normal block obeys

$$\|H_{NN}\|_{\mathrm{op}} \le \big\|H_{NN} + \sigma_t^{-2} P_N(y_t)\big\|_{\mathrm{op}} + \frac{1}{\sigma_t^2} \le C + \frac{1}{\sigma_t^2}$$

by (16). Therefore

$$\|H_t(x)\|_{\mathrm{op}} \le \|H_{TT}\|_{\mathrm{op}} + \|H_{TN}\|_{\mathrm{op}} + \|H_{NT}\|_{\mathrm{op}} + \|H_{NN}\|_{\mathrm{op}} \le C\Big(\frac{1}{\sigma_t^2} + \frac{\kappa\,\|\rho_t\|}{\sigma_t^2} + 1\Big),$$

after enlarging C. This is (65). Indeed, for any unit vector $v = v_T + v_N$ in the orthogonal splitting $T_{y_t}\mathcal{M} \oplus N_{y_t}\mathcal{M}$,

$$\|H_t(x)v\| \le \|H_{TT}v_T\| + \|H_{TN}v_N\| + \|H_{NT}v_T\| + \|H_{NN}v_N\|,$$

and $\|v_T\|, \|v_N\| \le 1$, so taking the supremum over $\|v\| = 1$ gives the displayed block-sum bound. □

D Velocity Lipschitz and Jacobian proofs

This appendix section collects the proofs of the velocity Jacobian bounds in Sections 3 and 4. We proceed in the same order as the main text. First, we prove the posterior formula of $v_t^\star$ and the tube-local posterior-covariance collapse that drive the scalar $O(1/\sigma_t)$ control. We then state an explicit three-regime version of Theorem 3.2, which makes the constant $C_{\mathcal{M}}(r)$ transparent and immediately implies the main-text theorem. Finally, we prove the score-form Jacobian identity and the anisotropic block bound from Section 4, concluding with the late-regime and sharp-rate corollaries referenced there.

Proof of Proposition 2.1. Since $\mathcal{M}$ is compact, the density $f \in L^1(\mathcal{M})$ and the Gaussian factor are bounded on $\mathcal{M}$ for each fixed (x, t), we may differentiate under the integral freely.

Write $X_1 = Y + \sigma\xi$ with $Y \sim f\mathrm{Vol}_{\mathcal{M}}$, $\xi \sim \mathcal{N}(0, I_d)$, and $X_0 \sim \mathcal{N}(0, I_d)$, all independent. Then

$$X_t = (1-t)X_0 + tY + t\sigma\xi.$$

Conditioned on $Y = y$, the terms $(1-t)X_0$ and $t\sigma\xi$ are independent centered Gaussians with covariances $(1-t)^2 I_d$ and $t^2\sigma^2 I_d$, so $X_t \mid Y{=}y \sim \mathcal{N}(ty, \sigma_t^2 I_d)$.

Since X_0 and ξ are jointly Gaussian and isotropic, conditioned on $(X_t = x, Y = y)$, standard Gaussian conditioning gives

$$\mathbb{E}[\xi \mid X_t = x, Y = y] = \frac{t\sigma}{\sigma_t^2}(x - ty), \qquad \mathbb{E}[X_0 \mid X_t = x, Y = y] = \frac{1-t}{\sigma_t^2}(x - ty).$$

Hence

$$\begin{aligned}\mathbb{E}[X_1 - X_0 \mid X_t = x, Y = y] =& y + \sigma\,\mathbb{E}[\xi \mid X_t = x, Y = y] - \mathbb{E}[X_0 \mid X_t = x, Y = y] \\ =& y + \frac{t\sigma^2 - (1-t)}{\sigma_t^2}(x - ty).\end{aligned}$$

Averaging over the posterior $Y \mid X_t = x$ yields

$$v_t^\star(x) = m_t(x) + \frac{t\sigma^2 - (1-t)}{\sigma_t^2}\big(x - t\,m_t(x)\big) = \frac{t(1+\sigma^2) - 1}{\sigma_t^2}\,x + \frac{1-t}{\sigma_t^2}\,m_t(x),$$

which is (6) with $\alpha_t = (t(1+\sigma^2) - 1)/\sigma_t^2$ and $1 - \alpha_t t = (1-t)/\sigma_t^2$.

Write $\varphi_s(u) := (2\pi s)^{-d/2}\exp(-\|u\|^2/(2s))$ for the isotropic Gaussian density with variance s, and define the unnormalized posterior weight $w_t(x, y) := f(y)\varphi_{\sigma_t^2}(x - ty)$. Set

$$D_t(x) := \int_{\mathcal{M}} w_t(x, y)\,\mathrm{dVol}_{\mathcal{M}}(y), \qquad N_t(x) := \int_{\mathcal{M}} y\,w_t(x, y)\,\mathrm{dVol}_{\mathcal{M}}(y),$$

so $m_t(x) = N_t(x)/D_t(x)$. Differentiate using $\nabla_x\varphi_{\sigma_t^2}(x - ty) = -(x - ty)/\sigma_t^2 \cdot \varphi_{\sigma_t^2}(x - ty)$ and we have

$$\nabla_x D_t(x) = -\frac{1}{\sigma_t^2}\big(xD_t(x) - tN_t(x)\big), \qquad \nabla_x N_t(x) = -\frac{1}{\sigma_t^2}\big(N_t(x)x^\top - tM_t(x)\big),$$

where $M_t(x) := \int_{\mathcal{M}} yy^\top w_t(x, y)\,\mathrm{dVol}_{\mathcal{M}}(y)$. Applying the quotient rule and cancelling common terms yields

$$\nabla_x m_t(x) = \frac{t}{\sigma_t^2}\left(\frac{M_t(x)}{D_t(x)} - m_t(x)m_t(x)^\top\right) = \frac{t}{\sigma_t^2}\,\Sigma_{Y|t}(x).$$

Differentiating (6) and substituting the expression for $\nabla_x m_t$ gives (7). □

Proof of Proposition 3.1. *Setup: reduction to a denoising model.* Fix $t \in (0, 1]$. The conditional law of Y given $X_t = x$ has density, with respect to $\mathrm{Vol}_{\mathcal{M}}$, proportional to

$$f(y)\exp\left(-\frac{\|x - ty\|^2}{2\sigma_t^2}\right).$$

This coincides with the posterior distribution in the denoising model $Z = Y + \varepsilon\eta$, where $\eta \sim \mathcal{N}(0, I_d)$, observed at $z = x/t$ with noise level $\varepsilon = \varepsilon_t = \sigma_t/t$. Therefore it suffices to prove

$$\|\mathrm{Cov}(Y \mid Z = z)\|_{\mathrm{op}} \le C_\Sigma\,\varepsilon^2$$

uniformly for $z \in \mathcal{M}_\varrho$ and $\varepsilon \le \varepsilon_0$ with certain ε_0 determined below. This immediately yields the scaled-tube claim because $x/t \in \mathcal{M}_\varrho$.

Fixed-tube consequence. If $r \in (0, \varrho)$ and $x \in \mathcal{M}_r$, choose $y_0 \in \mathcal{M}$ with $\|x - y_0\| \le r$. Then

$$\mathrm{dist}\left(\frac{x}{t}, \mathcal{M}\right) \le \left\|\frac{x}{t} - y_0\right\| \le \frac{\|x - y_0\|}{t} + \frac{(1-t)}{t}\|y_0\| \le \frac{r + (1-t)R_{\mathcal{M}}}{t}.$$

The right-hand side is at most ϱ if and only if $t \ge (R_{\mathcal{M}} + r)/(R_{\mathcal{M}} + \varrho) = t_{\mathrm{tube}}$. Thus for $t \ge t_{\mathrm{tube}}$ and $x \in \mathcal{M}_r$ we have $x/t \in \mathcal{M}_\varrho$, and the denoising bound applies.

Step 1: Reduction to a second moment. Since

$$u^\top\mathrm{Cov}(Y \mid Z = z)\,u = \mathrm{Var}(u^\top Y \mid Z = z) \le \mathbb{E}[(u^\top(Y - y_\star))^2 \mid Z = z]$$

for every unit vector $u \in \mathbb{R}^d$ and $y_\star := \pi_{\mathcal{M}}(z)$, taking the supremum over u gives

$$\|\mathrm{Cov}(Y \mid Z = z)\|_{\mathrm{op}} \le \mathbb{E}\left[\|Y - y_\star\|^2 \mid Z = z\right].$$

It therefore suffices to show that $\mathbb{E}[\|Y - y_\star\|^2 \mid Z = z] \le C_\Sigma \varepsilon^2$ uniformly over $z \in \mathcal{M}_\varrho$ and $\varepsilon \le \varepsilon_0$.

Step 2: Local graph chart with reach-controlled geometry. Since $\mathrm{reach}(\mathcal{M}) = \tau$ and $\mathcal{M}$ is C^2, for each $y_\star \in \mathcal{M}$ there is a C^2 map $h : B_T(0, \tau/4) \to T^\perp$ (with $T := T_{y_\star}\mathcal{M}$) such that $\mathcal{M}$ near $y_\star$ is parametrized by

$$y(u) = y_\star + u + h(u), \qquad h(0) = 0, \quad Dh(0) = 0, \quad \|h(u)\| \le \frac{\|u\|^2}{2\tau}.$$

The volume element in these coordinates is $\mathrm{dVol}_{\mathcal{M}}(y(u)) = J(u)\,\mathrm{d}u$ where

$$J(u) = \sqrt{\det(I_k + (Dh(u))^\top Dh(u))}.$$

Since h is C^2 and $Dh(0) = 0$, the reach-controlled graph bound

$$\|Dh(u)\|_{\mathrm{op}} \le \frac{1}{3}$$

holds on $\|u\| \le \tau/4$. Indeed, since $T_{y(u)}\mathcal{M}$ is the graph of $Dh(u)$ over T,

$$\|Dh(u)\|_{\mathrm{op}} = \tan \angle(T, T_{y(u)}\mathcal{M}).$$

By the tangent-space variation bounds of [5, Lemma 6 and Corollary 3],

$$\sin\Big(\frac{\angle(T, T_{y(u)}\mathcal{M})}{2}\Big) \le \frac{\|y(u) - y_\star\|}{2\tau}.$$

Moreover,

$$\|y(u) - y_\star\|^2 = \|u\|^2 + \|h(u)\|^2 \le \|u\|^2\Big(1 + \frac{\|u\|^2}{4\tau^2}\Big) \le \frac{65}{64}\|u\|^2, \qquad \|u\| \le \tau/4.$$

Writing $\vartheta := \angle(T, T_{y(u)}\mathcal{M})$, we therefore get

$$\sin^2\vartheta \le 4\sin^2(\vartheta/2) \le \frac{65\|u\|^2}{64\tau^2}, \qquad \cos^2\vartheta = 1 - \sin^2\vartheta \ge 1 - \frac{65\|u\|^2}{64\tau^2}.$$

Hence

$$\|Dh(u)\|_{\mathrm{op}} = \tan\vartheta \le \sqrt{\frac{65\|u\|^2}{64\tau^2 - 65\|u\|^2}} \le \sqrt{\frac{65}{959}} \le \frac{1}{3}, \qquad \|u\| \le \tau/4.$$

Therefore, J is bounded above and below on $\{\|u\| \le \tau/4\}$: there exist constants $0 < J_{\min} \le J_{\max} < \infty$, depending only on k, such that

$$J_{\min} \le J(u) \le J_{\max}, \qquad \|u\| \le \tau/4.$$

Denote $\mathcal{M}_{\mathrm{loc}} := \{y(u) : \|u\| < \tau/4\}$, which is open in $\mathcal{M}$, and let $\mathcal{M}_{\mathrm{far}} := \mathcal{M} \setminus \mathcal{M}_{\mathrm{loc}}$, which is closed in $\mathcal{M}$ and therefore compact.

Step 3: Quadratic bounds on the likelihood along the local patch. Define $\delta := \|z - y_\star\| \le \varrho$ and set $n := z - y_\star \in T_{y_\star}^\perp\mathcal{M}$, so that $\|n\| = \delta$. For $y(u) \in \mathcal{M}_{\mathrm{loc}}$,

$$z - y(u) = n - u - h(u).$$

Since $u \in T$ and $n, h(u) \in T^\perp$, orthogonality gives $\|z - y(u)\|^2 = \|u\|^2 + \|n - h(u)\|^2$. Using $\|h(u)\| \le \|u\|^2/(2\tau)$ and $\delta \le \varrho$:

$$\delta^2 + c_-\|u\|^2 \le \|z - y(u)\|^2 \le \delta^2 + c_+\|u\|^2, \qquad c_- := 1 - \varrho/\tau > 0, \quad c_+ := 1 + \varrho/\tau + c_0,$$

where $c_0 > 0$ is an absolute constant such that the $O(\|u\|^4/\tau^2)$ remainder is bounded by $c_0\|u\|^2$ on $\|u\| \le \tau/4$.

Step 4: Lower bound on the normalizer. Define the normalizer

$$D(z) := \int_{\mathcal{M}} f(y) \exp\left(-\frac{\|z - y\|^2}{2\varepsilon^2}\right) \mathrm{dVol}_{\mathcal{M}}(y).$$

Restricting the integration to $\mathcal{M}_{\mathrm{loc}}$ and applying $f \geq f_{\min}$, $J \geq J_{\min}$, and the upper quadratic bound on $\|z - y(u)\|^2$:

$$D(z) \geq f_{\min} J_{\min} \int_{\|u\| \leq \varepsilon} \exp\left(-\frac{\delta^2 + c_+ \|u\|^2}{2\varepsilon^2}\right) du \geq f_{\min} J_{\min}\, e^{-\delta^2/(2\varepsilon^2)}\, e^{-c_+/2}\, v_k\, \varepsilon^k,$$

where $v_k = \mathrm{Vol}(B_{\mathbb{R}^k}(0,1))$, valid for $\varepsilon \leq \tau/4$.

Step 5: Upper bound on the second-moment numerator. Define

$$N_2(z) := \int_{\mathcal{M}} \|y - y_\star\|^2 f(y) \exp\left(-\frac{\|z-y\|^2}{2\varepsilon^2}\right) \mathrm{dVol}_{\mathcal{M}}(y),$$

and split the integral into contributions from $\mathcal{M}_{\mathrm{loc}}$ and $\mathcal{M}_{\mathrm{far}}$.

Local part. On $\mathcal{M}_{\mathrm{loc}}$, $\|y(u) - y_\star\|^2 = \|u + h(u)\|^2 \leq C_1 \|u\|^2$ for an absolute constant C_1. Using $f \leq f_{\max}$, $J \leq J_{\max}$, and the lower quadratic bound:

$$N_{2,\mathrm{loc}}(z) \leq f_{\max} J_{\max} C_1 \int_{\mathbb{R}^k} \|u\|^2\, e^{-(\delta^2 + c_- \|u\|^2)/(2\varepsilon^2)}\, \mathrm{d}u = C_{\mathrm{loc}}\, e^{-\delta^2/(2\varepsilon^2)}\, \varepsilon^{k+2},$$

where $C_{\mathrm{loc}} = C_{\mathrm{loc}}(k, c_-, f_{\max}, J_{\max})$ is explicit from the Gaussian integral.

Far part. Define $G(z,y) := \|z - y\|^2 - \delta^2$. We establish a uniform lower bound on $G(z,y)$ for $y \in \mathcal{M}_{\mathrm{far}}$ via two sub-steps.

(i) Global reach gap. If $\delta = 0$, then $G(z,y) = \|y - y_\star\|^2$ for every $y \in \mathcal{M}$, and the bound below holds trivially. Assume $\delta > 0$ and let $\hat{n} := n/\delta$, a unit vector in $T_{y_\star}^{\perp}\mathcal{M}$. By the rolling-ball characterisation of $\mathrm{reach}(\mathcal{M}) = \tau$, the open ball $B(y_\star + \tau\hat{n}, \tau)$ contains no point of $\mathcal{M}$. Hence for every $y \in \mathcal{M}$,

$$\|y - (y_\star + \tau\hat{n})\|^2 \geq \tau^2 \quad \text{implies} \quad \langle \hat{n},\, y - y_\star \rangle \leq \frac{\|y - y_\star\|^2}{2\tau}.$$

Multiplying both sides by δ and using $\delta\hat{n} = n = z - y_\star$ gives $\langle z - y_\star,\, y - y_\star \rangle \leq \delta \|y - y_\star\|^2/(2\tau)$. Substituting into the identity $\|z-y\|^2 = \delta^2 + \|y - y_\star\|^2 - 2\langle z - y_\star, y - y_\star\rangle$ yields, for every $y \in \mathcal{M}$,

$$G(z,y) \geq \Big(1 - \frac{\delta}{\tau}\Big)\|y - y_\star\|^2 \geq \Big(1 - \frac{\varrho}{\tau}\Big)\|y - y_\star\|^2.$$

(ii) Far-set separation. We claim that $\|y - y_\star\| \geq \tau/4$ for all $y \in \mathcal{M}_{\mathrm{far}}$. Indeed, suppose for contradiction that $y \in \mathcal{M}_{\mathrm{far}}$ and $\|y - y_\star\| < \tau/4$. Since $\mathrm{reach}(\mathcal{M}) = \tau$, the orthogonal projection Π_T is injective on $\mathcal{M} \cap B(y_\star, \tau)$, so $y = y_\star + u + h(u)$ for a unique $u \in T_{y_\star}\mathcal{M}$. Because $u \perp h(u)$,

$$\|u\|^2 \leq \|u\|^2 + \|h(u)\|^2 = \|y - y_\star\|^2 < (\tau/4)^2.$$

Therefore $\|u\| < \tau/4$, which implies $y \in \mathcal{M}_{\mathrm{loc}}$. This contradicts $y \in \mathcal{M}_{\mathrm{far}}$, and the claim follows.

Define $\gamma := \inf_{y \in \mathcal{M}_{\mathrm{far}}} G(z,y)$. Combining (i) and (ii) yields

$$\gamma \geq \Big(1 - \frac{\varrho}{\tau}\Big)\frac{\tau^2}{16},$$

a positive constant depending only on τ and ϱ. In particular, $\|z - y\|^2 \geq \delta^2 + \gamma$ for all $y \in \mathcal{M}_{\mathrm{far}}$, giving

$$N_{2,\mathrm{far}}(z) \leq \mathrm{diam}(\mathcal{M})^2\, f_{\max}\, \mathrm{Vol}_{\mathcal{M}}(\mathcal{M})\, e^{-(\delta^2+\gamma)/(2\varepsilon^2)} = C_{\mathrm{far}}\, e^{-\delta^2/(2\varepsilon^2)}\, e^{-\gamma/(2\varepsilon^2)}.$$

Step 6: Combination and choice of ε_0. Dividing $N_2 = N_{2,\mathrm{loc}} + N_{2,\mathrm{far}}$ by the normalizer lower bound from Step 4 (the common factor $\exp(-\delta^2/(2\varepsilon^2))$ cancels):

$$\mathbb{E}\left[\|Y - y_\star\|^2 \mid Z = z\right] \leq \underbrace{\frac{C_{\mathrm{loc}}}{c_D}}_{=:A_1}\varepsilon^2 + \underbrace{\frac{C_{\mathrm{far}}}{c_D}}_{=:A_2}\varepsilon^{-k}\, e^{-\gamma/(2\varepsilon^2)},$$

where $c_D := f_{\min} J_{\min} \exp(-c_+/2) v_k$. Since $\gamma > 0$, we have

$$\lim_{\varepsilon \to 0} \varepsilon^{-k-2} e^{-\gamma/(2\varepsilon^2)} = 0,$$

thus

$$\varepsilon^{-k} e^{-\gamma/(2\varepsilon^2)} = o(\varepsilon^2)$$

as $\varepsilon \to 0$. Therefore there exists $\varepsilon_0 \in (0, \tau/4]$ such that

$$A_2 \varepsilon^{-k} e^{-\gamma/(2\varepsilon^2)} \le A_1 \varepsilon^2$$

for all $\varepsilon \le \varepsilon_0$. For such ε,

$$\mathbb{E}\left[\|Y - y_\star\|^2 \mid Z = z\right] \le 2A_1\, \varepsilon^2,$$

uniformly over $z \in \mathcal{M}_\varrho$. Setting $C_\Sigma := 2A_1$ gives $\|\mathrm{Cov}(Y \mid Z = z)\|_{\mathrm{op}} \le C_\Sigma \varepsilon^2$. Translating back via $\varepsilon = \sigma_t/t$ yields (10). □

Lemma D.1. *Under Assumption 1, let $R_{\mathcal{M}} := \sup_{y \in \mathcal{M}} \|y\|$. Then*

$$R_{\mathcal{M}} \ge \tau.$$

Proof. For any $x, y \in \mathcal{M}$, $\|x - y\| \le \|x\| + \|y\| \le 2R_{\mathcal{M}}$, so

$$\mathrm{diam}(\mathcal{M}) \le 2R_{\mathcal{M}}, \quad \text{hence} \quad R_{\mathcal{M}} \ge \frac{1}{2}\mathrm{diam}(\mathcal{M}).$$

Let $\mathrm{foc}(\mathcal{M})$ denote the *focal radius* of $\mathcal{M}$, i.e., the supremum of $r > 0$ such that the normal exponential map $(y, v) \mapsto y + v$ is a local diffeomorphism on normal vectors of length less than r. By [38, Theorem A], for any compact, boundaryless, connected smooth submanifold of $\mathbb{R}^d$,

$$\mathrm{diam}(\mathcal{M}) \ge 2\,\mathrm{foc}(\mathcal{M}).$$

If $\mathcal{M}$ is disconnected, choose any connected component $\mathcal{M}_i$. Since uniqueness of nearest points for the full set implies uniqueness for each component, $\mathrm{reach}(\mathcal{M}_i) \ge \mathrm{reach}(\mathcal{M}) = \tau$. Applying the connected case to $\mathcal{M}_i$ gives

$$\mathrm{diam}(\mathcal{M}) \ge \mathrm{diam}(\mathcal{M}_i) \ge 2\,\mathrm{foc}(\mathcal{M}_i) \ge 2\tau.$$

Thus the same conclusion holds without connectedness.

By Federer's definition [14, Definition 4.1], $\tau = \mathrm{reach}(\mathcal{M})$ is the largest radius such that every point within distance less than τ of $\mathcal{M}$ has a unique nearest point on $\mathcal{M}$. For any $r < \tau$, uniqueness of nearest points on the open tube $\mathcal{M}_r$ implies that the normal exponential map is injective on normal vectors of length less than r; combined with the C^2 smoothness of $\mathcal{M}$, it is a local diffeomorphism there. Hence $r \le \mathrm{foc}(\mathcal{M})$ for every $r < \tau$, giving $\tau \le \mathrm{foc}(\mathcal{M})$.

As a consequence, $R_{\mathcal{M}} \ge \mathrm{diam}(\mathcal{M})/2 \ge \mathrm{foc}(\mathcal{M}) \ge \tau$. □

We now demonstrate the explicit three-regime form of Theorem 3.2. The statement below is in more detail than the main-text theorem: it isolates the early, middle, and tube-local contributions to the constant $C_{\mathcal{M}}(r)$.

Theorem D.2 (Explicit three-regime version of Theorem 3.2)**.** *Under Assumptions 1–3, the following holds for the population velocity Jacobian $\nabla_x v_t^\star(x)$ from Proposition 2.1. Fix $r \in (0, \tau/4)$, $\varrho \in (r, \tau/2)$, recall t_{tube} in Proposition 3.1, and let ε_0, C_Σ be as in Proposition 3.1 for the observation radius ϱ. By Lemma D.1, $R_{\mathcal{M}} \ge \tau > \varrho$, so $t_{\mathrm{tube}} = (R_{\mathcal{M}} + r)/(R_{\mathcal{M}} + \varrho) \ge \tau/(\tau + \varrho) > 2/3$ (using $\varrho < \tau/2$). In particular, $t_{\mathrm{tube}} > 1/2$ and the three sub-regimes below partition $[0, 1]$.*

(i) Early sub-regime $t \in [0, 1/2]$: global bound $O(1/\sigma_t)$. For all $t \in [0, 1/2]$ and all $x \in \mathbb{R}^d$,

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le \frac{A_{\mathrm{early}}}{\sigma_t}, \qquad A_{\mathrm{early}} := \sqrt{2} + \mathrm{diam}(\mathcal{M})^2. \tag{66}$$

(ii) Middle sub-regime $t \in (1/2, t_{\mathrm{tube}})$: global bound $O(1/\sigma_t)$. For all $t \in (1/2, t_{\mathrm{tube}})$ and all $x \in \mathbb{R}^d$,

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le \frac{A_{\mathrm{mid}}(r)}{\sigma_t}, \qquad A_{\mathrm{mid}}(r) := \sqrt{2} + \frac{\mathrm{diam}(\mathcal{M})^2}{4(1 - t_{\mathrm{tube}})^2} = \sqrt{2} + \frac{\mathrm{diam}(\mathcal{M})^2 (R_{\mathcal{M}} + \varrho)^2}{4(\varrho - r)^2}. \tag{67}$$

(iii) Tube sub-regime $t \in [t_{\mathrm{tube}}, 1]$: Jacobian bound $O(1/\sigma_t)$ on $\mathcal{M}_r$. For all $t \in [t_{\mathrm{tube}}, 1]$ satisfying $\sigma_t/t \le \varepsilon_0$ and for all $x \in \mathcal{M}_r$,

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le \frac{A_{\mathrm{tube}}(r)}{\sigma_t}, \qquad A_{\mathrm{tube}}(r) := \sqrt{2} + \frac{C_\Sigma}{t_{\mathrm{tube}}}. \tag{68}$$

Combining (i)–(iii), define

$$C_{\mathcal{M}}(r) := \max\Big(A_{\mathrm{early}}, A_{\mathrm{mid}}(r), A_{\mathrm{tube}}(r)\Big). \tag{69}$$

Then for all $t \in [0,1]$ (with $x \in \mathcal{M}_r$ and $\sigma_t/t \le \varepsilon_0$ when $t \ge t_{\mathrm{tube}}$),

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le \frac{C_{\mathcal{M}}(r)}{\sigma_t}. \tag{70}$$

In particular, Theorem 3.2 follows immediately from (70)*. The dependence of $C_{\mathcal{M}}(r)$ on r is explicit: $1/t_{\mathrm{tube}} = (R_{\mathcal{M}}+\varrho)/(R_{\mathcal{M}}+r)$ decreases in r, while $A_{\mathrm{mid}}(r) \propto (\varrho - r)^{-2}$ increases as r approaches ϱ. For any fixed ϱ and $r \in (0, \varrho/2]$, $C_{\mathcal{M}}(r) = O(1)$ with the implicit constant depending only on $\tau, \varrho, R_{\mathcal{M}}, k, f_{\min}, f_{\max}, \mathrm{diam}(\mathcal{M}), \mathrm{Vol}_{\mathcal{M}}(\mathcal{M})$.*

Proof. Start from the Jacobian identity (7):

$$\nabla_x v_t^\star(x) = a_t\, I_d + b_t\, \Sigma_{Y|t}(x), \quad a_t := \frac{t(1+\sigma^2)-1}{\sigma_t^2}, \quad b_t := \frac{t(1-t)}{\sigma_t^4} \ge 0.$$

Since $\Sigma_{Y|t}(x) \succeq 0$, the operator norm satisfies

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le |a_t| + b_t\, \|\Sigma_{Y|t}(x)\|_{\mathrm{op}}.$$

Universal bound $|a_t| \le \sqrt{2}/\sigma_t$. For every $t \in [0,1]$,

$$|t(1+\sigma^2) - 1| = |t\sigma^2 - (1-t)| \le t\sigma^2 + (1-t) \le t\sigma + (1-t).$$

Note that

$$t\sigma + (1-t) \le \sqrt{2}\,\sqrt{t^2\sigma^2 + (1-t)^2} = \sqrt{2}\,\sigma_t.$$

Hence

$$|a_t| \le \frac{\sqrt{2}}{\sigma_t}.$$

Part (i). Since $\mathcal{M}$ is compact, Popoviciu's inequality [39, 42] gives

$$\|\Sigma_{Y|t}(x)\|_{\mathrm{op}} \le \frac{\mathrm{diam}(\mathcal{M})^2}{4}, \qquad \forall\, x, t. \tag{71}$$

For $t \in [0, 1/2]$, $\sigma_t^2 \ge (1-t)^2$, so $b_t = t(1-t)/\sigma_t^4 \le t/(1-t)^3 \le 4$. Hence

$$b_t\, \|\Sigma_{Y|t}(x)\|_{\mathrm{op}} \le 4 \cdot \frac{\mathrm{diam}(\mathcal{M})^2}{4} = \mathrm{diam}(\mathcal{M})^2 \le \frac{\mathrm{diam}(\mathcal{M})^2}{\sigma_t},$$

because $\sigma_t \le 1$. Combining with $|a_t| \le \sqrt{2}/\sigma_t$ gives

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le \frac{\sqrt{2} + \mathrm{diam}(\mathcal{M})^2}{\sigma_t},$$

giving (66).

Part (ii). For $t \in (1/2, t_{\mathrm{tube}})$ we have $1 - t \ge 1 - t_{\mathrm{tube}} = (\varrho - r)/(R_{\mathcal{M}} + \varrho) > 0$. Using $\sigma_t \ge 1 - t$,

$$\frac{1-t}{\sigma_t^2} \le \frac{1}{\sigma_t}.$$

Also $\sigma_t \ge 1 - t \ge 1 - t_{\mathrm{tube}}$, so

$$\frac{t}{\sigma_t^2} \le \frac{1}{(1-t_{\mathrm{tube}})^2}.$$

Therefore, using (71),

$$b_t \, \|\Sigma_{Y|t}(x)\|_{\mathrm{op}} = \frac{t}{\sigma_t^2} \cdot \frac{1-t}{\sigma_t^2} \, \|\Sigma_{Y|t}(x)\|_{\mathrm{op}} \le \frac{1}{(1-t_{\mathrm{tube}})^2} \cdot \frac{1}{\sigma_t} \cdot \frac{\mathrm{diam}(\mathcal{M})^2}{4}.$$

Combined with the universal bound $|a_t| \le \sqrt{2}/\sigma_t$,

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le \frac{1}{\sigma_t}\Big(\sqrt{2} + \frac{\mathrm{diam}(\mathcal{M})^2}{4(1-t_{\mathrm{tube}})^2}\Big),$$

giving (67).

Part (iii). For $t \in [t_{\mathrm{tube}}, 1]$ and $x \in \mathcal{M}_r$, apply the fixed-tube consequence in Proposition 3.1: $\|\Sigma_{Y|t}(x)\|_{\mathrm{op}} \le C_\Sigma \sigma_t^2/t^2$. Then

$$b_t \, \|\Sigma_{Y|t}(x)\|_{\mathrm{op}} \le \frac{t(1-t)}{\sigma_t^4} \cdot C_\Sigma \frac{\sigma_t^2}{t^2} = \frac{C_\Sigma(1-t)}{t\,\sigma_t^2} \le \frac{C_\Sigma}{t_{\mathrm{tube}}\sigma_t},$$

using $\sigma_t \ge 1-t$ and $t \ge t_{\mathrm{tube}}$. Summing with $|a_t| \le \sqrt{2}/\sigma_t$:

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \le \frac{1}{\sigma_t}\Big(\sqrt{2} + \frac{C_\Sigma}{t_{\mathrm{tube}}}\Big),$$

giving (68). □

Proof of Proposition 4.1. We derive the score-form identity via Tweedie's formula [12] applied to the Gaussian conditional law of X_t given X_1. Fix $t \in (0,1)$; the boundary case $t = 1$ is treated at the end.

Since $X_0 \sim \mathcal{N}(0, I_d)$ is independent of X_1, the interpolant (3) gives, for every $y \in \mathbb{R}^d$,

$$X_t \,\big|\, X_1 = y \sim \mathcal{N}\big(ty,\, (1-t)^2 I_d\big).$$

Define the posterior mean of X_1 given X_t,

$$\bar{m}_t(x) := \mathbb{E}[X_1 \mid X_t = x],$$

which is well-defined since $\mathbb{E}\,\|X_1\| < \infty$ under Assumptions 1–3.

Given the independence of X_0 and X_1, Tweedie's formula [12, Eq. (1.4)] indicates

$$\bar{m}_t(x) = \frac{x + (1-t)^2 \, \nabla_x \log p_t(x)}{t}. \tag{72}$$

Solving (3) for X_0 yields $(1-t)X_0 = X_t - tX_1$. Taking conditional expectation given $X_t = x$,

$$\mathbb{E}[X_0 \mid X_t = x] = \frac{x - t\,\bar{m}_t(x)}{1-t}.$$

Substituting into the velocity definition (5),

$$v_t^\star(x) = \bar{m}_t(x) - \frac{x - t\,\bar{m}_t(x)}{1-t} = \frac{\bar{m}_t(x) - x}{1-t}. \tag{73}$$

Combining (72) and (73),

$$v_t^\star(x) = \frac{1}{1-t} \cdot \frac{x + (1-t)^2 \, \nabla_x \log p_t(x) - tx}{t} = \frac{x + (1-t)\, \nabla_x \log p_t(x)}{t}. \tag{74}$$

Differentiating the above in x yields

$$\nabla_x v_t^\star(x) = \frac{I_d + (1-t)\, \nabla_x^2 \log p_t(x)}{t},$$

which is exactly (12).

Boundary case $t = 1$. At $t = 1$, $X_t = X_1$, and by the independence of $X_0 \sim \mathcal{N}(0, I_d)$ and X_1,

$$v_1^\star(x) = \mathbb{E}[X_1 - X_0 \mid X_1 = x] = x - \mathbb{E}[X_0] = x,$$

so $\nabla_x v_1^\star(x) = I_d$. This matches the right-hand side of (12) evaluated at $t = 1$, which completes the proof. □

Lemma D.3. *Assume the uniform mixed Hessian bound* (15) *holds on* $\mathcal{M}_{t,r}^{\mathrm{sc}}$ *for some* $0 < r \le r_0 < \tau/4$. *Define the cross-block envelope*

$$c_{TN}(t) := \sup_{x \in \mathcal{M}_{t,r}^{\mathrm{sc}}} \left\| P_T^{(t)}(x)\, \nabla_x v_t^\star(x)\, P_N^{(t)}(x) \right\|_{\mathrm{op}}. \tag{75}$$

Then for all $t \in [t^\star, 1]$ *with* $\sigma_t/t \le \varepsilon_0$,

$$c_{TN}(t) \le \frac{1-t}{t}\, C_{\mathrm{sc}} \le 2C_{\mathrm{sc}}\sigma_t. \tag{76}$$

Proof. By the score-form Jacobian identity (12), the cross-block of $\nabla_x v_t^\star$ equals $(1-t)/t$ times the cross-block of $\nabla_x^2 \log p_t$ (the identity matrix contributes no cross terms). Applying (15) gives, for each $x \in \mathcal{M}_{t,r}^{\mathrm{sc}}$,

$$\left\| P_T^{(t)}(x)\, \nabla_x v_t^\star(x)\, P_N^{(t)}(x) \right\|_{\mathrm{op}} \le \frac{1-t}{t}\, C_{\mathrm{sc}}.$$

Taking the supremum over $x \in \mathcal{M}_{t,r}^{\mathrm{sc}}$ yields the first inequality in (76). For the second, note that $\sigma_t \ge 1 - t$ and $t \ge t^\star \ge 1/2$, hence $(1-t)/t \le 2\sigma_t$. □

Proposition D.4 (Pointwise diagonal Jacobian bounds on the contractive window)**.** *Fix* $0 < r \le r_0 < \tau/4$. *Let* $\varepsilon_0, C_{\mathrm{sc}}$ *be the constants from Proposition 4.2. There exists*

$$\sigma_{\mathrm{diag}} = \sigma_{\mathrm{diag}}(r_0, \mathcal{M}, f) \in \Big(0, \min\Big\{\frac{1}{2}, \frac{\tau}{4}, \frac{\varepsilon_0}{\sqrt{2}}\Big\}\Big] \tag{77}$$

such that, whenever $\sigma \le \sigma_{\mathrm{diag}}$, *the following holds for every* $t \in [t^\star, t^\dagger]$ *and every* $x \in \mathcal{M}_{t,r}^{\mathrm{sc}}$. *Writing* $(y_t, \rho_t) := (\pi_{\mathcal{M}}(x/t), \rho_{\mathcal{M}}(x/t))$, *one has*

$$\langle u, \nabla_x v_t^\star(x) u \rangle \le L_T(t)\, \|u\|^2, \qquad u \in T_{y_t}\mathcal{M}, \tag{78}$$

$$\langle w, \nabla_x v_t^\star(x) w \rangle \le -\frac{1}{2}\lambda_N^{\mathrm{flat}}(t)\, \|w\|^2, \qquad w \in N_{y_t}\mathcal{M}, \tag{79}$$

where

$$L_T(t) := 2 + C_{\mathrm{sc}}\sigma + \frac{8\kappa\, r}{3\sigma_t}. \tag{80}$$

Proof. Since $\sigma_t^2 = (1-t)^2 + t^2\sigma^2$, one has

$$\varepsilon_t^2 = \frac{\sigma_t^2}{t^2} = \sigma^2 + \Big(\frac{1-t}{t}\Big)^2.$$

Since $t \ge t^\star = 1/(1+\sigma)$, one has

$$\frac{1-t}{t} \le \sigma, \qquad \varepsilon_t^2 \le 2\sigma^2.$$

Hence $\varepsilon_t \le \sqrt{2}\,\sigma \le \varepsilon_0$, so Proposition 4.2 applies throughout $[t^\star, t^\dagger]$.

By the score-form identity (12),

$$\nabla_x v_t^\star(x) = \frac{(1-t)\,\nabla_x^2 \log p_t(x) + I_d}{t},$$

so $\nabla_x v_t^\star(x)$ is symmetric. Let $H_t(x) := \nabla_x^2 \log p_t(x)$.

For $u \in T_{y_t}\mathcal{M}$ with $\|u\| = 1$, Proposition 4.2 and Lemma C.1 give

$$\langle u, H_t(x) u \rangle \le C_{\mathrm{sc}} + \sigma_t^{-2} \left\| S_{\rho_t}(y_t)(I - S_{\rho_t}(y_t))^{-1} \right\|_{\mathrm{op}} \le C_{\mathrm{sc}} + \frac{4\kappa\, \|\rho_t\|}{3\sigma_t^2}.$$

Since $x \in \mathcal{M}_{t,r}^{\mathrm{sc}}$, one has $\|\rho_t\| \le r$. Therefore

$$\langle u, \nabla_x v_t^\star(x) u \rangle \le \frac{1}{t} + \frac{1-t}{t} C_{\mathrm{sc}} + \frac{1-t}{t}\frac{4\kappa r}{3\sigma_t^2}.$$

Using $t \geq 1/2$, $(1-t)/t \leq \sigma$, and $\sigma_t \geq t\sigma \geq \sigma/2$, we obtain

$$\frac{1}{t} \leq 2, \qquad \frac{1-t}{t} C_{\mathrm{sc}} \leq C_{\mathrm{sc}} \sigma, \qquad \frac{1-t}{t} \cdot \frac{4\kappa r}{3\sigma_t^2} \leq \frac{8\kappa r}{3\sigma_t},$$

which yields (78).

For $w \in N_{y_t}\mathcal{M}$ with $\|w\| = 1$, Proposition 4.2 gives

$$\langle w, H_t(x) w \rangle \leq -\sigma_t^{-2} + C_{\mathrm{sc}}.$$

Hence

$$\langle w, \nabla_x v_t^\star(x) w \rangle \leq \frac{1 - (1-t)\sigma_t^{-2}}{t} + \frac{1-t}{t} C_{\mathrm{sc}} = -\lambda_N^{\mathrm{flat}}(t) + \frac{1-t}{t} C_{\mathrm{sc}}.$$

On $[t^\star, t^\dagger]$, the definition of $t^\dagger$ gives $(1-t) - t\sigma^2 \geq \sigma^2/2$, while $t \geq t^\star$ implies $\sigma_t^2 \leq 2\sigma^2$. Therefore

$$\lambda_N^{\mathrm{flat}}(t) \geq \frac{1}{4}.$$

If σ_{diag} is chosen so that $\sigma_{\mathrm{diag}} \leq 1/2$ and $C_{\mathrm{sc}} \sigma_{\mathrm{diag}} \leq 1/8$, then

$$\frac{1-t}{t} C_{\mathrm{sc}} \leq C_{\mathrm{sc}} \sigma \leq \frac{1}{8} \leq \frac{1}{2} \lambda_N^{\mathrm{flat}}(t),$$

which yields (79). □

Proof of Theorem 4.4. By Proposition D.4, the diagonal quadratic forms of $\nabla_x v_t^\star(x)$ are bounded by $L_T(t)$ and $-\lambda_N^{\mathrm{flat}}(t)/2$, respectively. By Lemma D.3, the operator norm of the TN block is at most $c_{TN}(t)$. Therefore, for every $z = u + w$ with $u \in T_{y_t}\mathcal{M}$ and $w \in N_{y_t}\mathcal{M}$,

$$\langle z, \nabla_x v_t^\star(x) z \rangle \leq L_T(t) \|u\|^2 - \frac{1}{2} \lambda_N^{\mathrm{flat}}(t) \|w\|^2 + 2 c_{TN}(t) \|u\| \|w\|,$$

which is exactly (20). □

Corollary D.5. *Under the conditions of Proposition 4.2, for all $t \in [t^\star, 1]$ with $\sigma_t/t \leq \varepsilon_0$ and all $x \in \mathcal{M}_{t,r}^{\mathrm{sc}}$,*

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \leq \frac{C}{\sigma_t}, \tag{81}$$

for a constant C depending only on r_0, $\mathcal{M}$ and f.

Proof. For every $t \in [t^\star, 1]$ and every $x \in \mathcal{M}_{t,r}^{\mathrm{sc}}$, Proposition C.9 gives

$$\big\|\nabla_x^2 \log p_t(x)\big\|_{\mathrm{op}} \leq C_1\Big(\frac{1}{\sigma_t^2} + \frac{\kappa \|\rho_t\|}{\sigma_t^2} + 1\Big) \leq C_2\Big(\frac{1}{\sigma_t^2} + 1\Big),$$

since $\|\rho_t\| \leq r$ on $\mathcal{M}_{t,r}^{\mathrm{sc}}$. Here, C_1 and C_2 are constants depending only on $r_0, \mathcal{M}, f$. Using the score-form identity (12),

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \leq \frac{1}{t} + \frac{1-t}{t} \big\|\nabla_x^2 \log p_t(x)\big\|_{\mathrm{op}} \leq \frac{1}{t} + \frac{1-t}{t} C_2\Big(\frac{1}{\sigma_t^2} + 1\Big).$$

Now $t \geq t^\star \geq 1/2$, $\sigma_t \geq 1 - t$, and $\sigma_t \leq 1$ under Assumption 3. Therefore

$$\frac{1}{t} \leq 2 \leq \frac{2}{\sigma_t}, \qquad \frac{1-t}{t} \cdot \frac{1}{\sigma_t^2} \leq \frac{2}{\sigma_t}, \qquad \frac{1-t}{t} \leq 1 \leq \frac{1}{\sigma_t}.$$

Substituting these bounds into the previous display yields

$$\|\nabla_x v_t^\star(x)\|_{\mathrm{op}} \leq \frac{C}{\sigma_t}$$

uniformly on $\mathcal{M}_{t,r}^{\mathrm{sc}}$. □

Corollary D.6. *Under the conditions of Proposition 4.2, suppose $k < d$. For every $y \in \mathcal{M}$, let $x = t^\star y$. As $\sigma \downarrow 0$,*

$$P_N(y)\,\nabla_x v^\star_{t^\star}(x)\,P_N(y) = -\frac{1-\sigma^2}{2\sigma}\,P_N(y) + O(\sigma)$$

in operator norm, uniformly in y. Consequently, $\|\nabla_x v^\star_{t^\star}(t^\star y)\|_{\mathrm{op}} = \Omega(1/\sigma) = \Omega(1/\sigma_{t^\star})$ uniformly in y, so the rate in Theorem 3.2 cannot be improved in general.

Proof. At $x = ty$, the moving normal projection in Proposition 4.2 is $P_N(y)$. By (16),

$$P_N(y)\,\nabla_x^2 \log p_t(ty)\,P_N(y) = -\frac{1}{\sigma_t^2}P_N(y) + R_t(y), \qquad \|R_t(y)\|_{\mathrm{op}} \le C_{\mathrm{sc}}.$$

Projecting the score-form identity (12) thus gives

$$P_N(y)\,\nabla_x v_t^\star(ty)\,P_N(y) = \left(\frac{1}{t} - \frac{1-t}{t\sigma_t^2}\right)P_N(y) + \frac{1-t}{t}R_t(y).$$

Since $\sigma_t^2 = (1-t)^2 + t^2\sigma^2$, the coefficient simplifies to

$$\frac{1}{t} - \frac{1-t}{t\sigma_t^2} = \frac{t\sigma^2 - (1-t)}{\sigma_t^2} = -\lambda_N^{\mathrm{flat}}(t).$$

The remainder $E_t(y) := (1-t)R_t(y)/t$ satisfies $\|E_t(y)\|_{\mathrm{op}} \le C_{\mathrm{sc}}(1-t)/t$. For sufficiently small σ, the small-noise condition of Proposition 4.2 holds at $t = t^\star$. Direct calculation yields $\lambda_N^{\mathrm{flat}}(t^\star) = (1-\sigma^2)/(2\sigma)$, while $(1-t^\star)/t^\star = \sigma$. Since $P_N(y)$ is nonzero when $k < d$, the claimed operator-norm lower bound follows by the reverse triangle inequality. Finally, $\sigma_{t^\star} = \sqrt{2}\,\sigma/(1+\sigma)$. □

E Stability and estimation details

This appendix section develops the stability and statistical theory summarized in Section 5. The statistical part gives tube-restricted risk guarantees for the time-split estimator defined below. The stability part explains how the late-time anisotropic Jacobian bounds propagate velocity error along coupled ODE trajectories, and separates the remaining early-budget and density-ratio requirements needed for an ambient transport bound. Throughout, we work under Assumptions 1–3 and fix a master tube radius $r_0 < \tau/4$. The constant C may vary from line to line.

E.1 Estimation setup and tube-restricted risks

Given n i.i.d. training pairs $\{(X_0^{(i)}, X_1^{(i)})\}_{i=1}^n$ from the CFM coupling in Section 2, with $X_0^{(i)} \sim \mathcal{N}(0, I_d)$ and $X_1^{(i)} \sim \nu_\sigma$, define the interpolants $X_t^{(i)} := (1-t)X_0^{(i)} + tX_1^{(i)}$. As a reference point, the standard full-window CFM empirical risk minimizer over a class $\mathcal{V}$ of time-dependent vector fields is

$$\widehat{v}^{\mathrm{glob}} := \arg\min_{v\in\mathcal{V}} \frac{1}{n}\sum_{i=1}^n \int_0^1 \left\| X_1^{(i)} - X_0^{(i)} - v_t(X_t^{(i)}) \right\|^2 \mathrm{d}t. \tag{82}$$

We write the empirical objectives with exact time integrals in (82), (134), and (151) for presentation conciseness. The same arguments apply to the usual sampled-time implementation after replacing these integrals by the corresponding empirical averages over sampled times.

The construction analyzed below is a time-split version of this CFM procedure. To match the different geometry of the early and late regimes, define

$$\widehat{v}_t(x) := \mathbf{1}_{\{t<t^\star\}}\widehat{v}_t^{\mathrm{early}}(x) + \mathbf{1}_{\{t\ge t^\star\}}\widehat{v}_t^{\mathrm{late}}(x), \tag{83}$$

where the early component is obtained from the full-space early loss in Proposition E.33, and the late component from the tube-restricted late loss in Proposition E.25. Thus (82) records the usual CFM objective, while Theorem E.35 analyzes the explicitly time-split construction (83).

Let $\widehat{\nu}_\sigma$ denote the terminal law of the learned ODE $\dot{\widehat{X}}_t = \widehat{v}_t(\widehat{X}_t)$, $\widehat{X}_0 \sim \mathcal{N}(0, I_d)$. We define two types of velocity-level risks as follows.

Definition E.1. For a tube radius $r < \tau/2$, the *tube-restricted velocity risk* is

$$\mathcal{R}_r(v) := \int_0^1 \mathbb{E}\Big[\|v_t(X_t) - v_t^\star(X_t)\|^2 \mathbf{1}_{\{X_t \in \mathcal{M}_r\}} \Big] \, \mathrm{d}t, \tag{84}$$

and the *late tube-restricted velocity risk* is

$$\mathcal{R}_r^{\mathrm{late}}(v) := \int_{t^\star}^1 \mathbb{E}\Big[\|v_t(X_t) - v_t^\star(X_t)\|^2 \mathbf{1}_{\{X_t \in \mathcal{M}_r\}} \Big] \, \mathrm{d}t. \tag{85}$$

E.2 Anisotropic late-regime stability

This subsection turns the pointwise block-Jacobian structure from Section 4 into a trajectory-level late-window bound on $[t^\star, 1]$. The mechanism is anisotropic: tangential errors may expand mildly, whereas normal errors are damped by the strong inward contraction near $\mathcal{M}$. We first demonstrate algebraic tools for transferring block bounds between nearby tangent–normal splittings. We then formulate the stopped localization used in the Section 5 sketch: the coupled trajectories are propagated only until the population trajectory leaves the inner scaled tube or their separation exceeds a fixed threshold. On this stopped interval, the mean-value segment stays in an enlarged tube, so the block-Jacobian bounds apply along the whole segment. The result is a stopped perturbation bound whose amplification is governed by the effective anisotropic budget Ψ_{eff}.

E.2.1 Algebraic core

The following auxiliary lemmas isolate the algebraic preliminaries needed to compare tangent–normal splittings along nearby points in the tube.

Lemma E.2. *Let*

$$A = \begin{pmatrix} a & c \\ c & -\lambda \end{pmatrix},$$

with $\lambda > 0$ and $a \geq 0$. Then the largest eigenvalue of A, denoted by $\lambda_{\max}(A)$, satisfies

$$\lambda_{\max}(A) \leq a + \frac{c^2}{\lambda}. \tag{86}$$

Proof. The characteristic polynomial gives

$$\lambda_{\max}(A) = \frac{a - \lambda + \sqrt{(a+\lambda)^2 + 4c^2}}{2}.$$

Since

$$a + \lambda + \sqrt{(a+\lambda)^2 + 4c^2} \geq 2(a+\lambda) \geq 2\lambda,$$

we obtain $\lambda_{\max}(A) \leq a + c^2/\lambda$. □

Lemma E.3. *Let A be a real symmetric operator on $\mathbb{R}^d$, and let P, Q be orthogonal projections with $\varepsilon := \|P - Q\|_{\mathrm{op}}$. For $R \in \{P, Q\}$ and $\xi \in \mathbb{R}^d$, write*

$$\xi_R := R\xi, \qquad \xi_R^\perp := (I - R)\xi.$$

Suppose that, for some $L, c, \lambda \geq 0$ and every $\xi \in \mathbb{R}^d$, the Q-splitting satisfies

$$\langle \xi, A\xi \rangle \leq L \|\xi_Q\|^2 + 2c \|\xi_Q\| \, \|\xi_Q^\perp\| - \lambda \|\xi_Q^\perp\|^2.$$

Then, for every $\xi \in \mathbb{R}^d$, the P-splitting satisfies

$$\langle \xi, A\xi \rangle \leq \widetilde{L} \|\xi_P\|^2 + 2\widetilde{c} \|\xi_P\| \, \|\xi_P^\perp\| - \widetilde{\lambda} \|\xi_P^\perp\|^2,$$

where

$$\widetilde{L} := L + 2c\varepsilon, \qquad \widetilde{c} := c + (L + \lambda)\varepsilon + c\varepsilon^2, \qquad \widetilde{\lambda} := \lambda - 2c\varepsilon - (L + \lambda)\varepsilon^2. \tag{87}$$

Proof. Fix $\xi \in \mathbb{R}^d$. Since $\xi = \xi_P + \xi_P^\perp$, its Q-components satisfy

$$\xi_Q = Q\xi_P + Q\xi_P^\perp, \qquad \xi_Q^\perp = (I-Q)\xi_P + (I-Q)\xi_P^\perp.$$

Because $P\xi_P = \xi_P$ and $P\xi_P^\perp = 0$,

$$(I-Q)\xi_P = (P-Q)\xi_P, \qquad Q\xi_P^\perp = (Q-P)\xi_P^\perp.$$

Consequently,

$$\|\xi_Q\| \le \|\xi_P\| + \varepsilon \left\|\xi_P^\perp\right\|, \qquad \left\|\xi_Q^\perp\right\| \le \varepsilon \|\xi_P\| + \left\|\xi_P^\perp\right\|. \tag{88}$$

For the negative normal term, we derive a lower bound. Orthogonality of $Q\xi_P^\perp$ and $(I-Q)\xi_P^\perp$ gives

$$\left\|(I-Q)\xi_P^\perp\right\|^2 = \left\|\xi_P^\perp\right\|^2 - \left\|Q\xi_P^\perp\right\|^2 \ge (1-\varepsilon^2)\left\|\xi_P^\perp\right\|^2.$$

Expanding $\xi_Q^\perp = (I-Q)\xi_P + (I-Q)\xi_P^\perp$ and bounding its cross term from below, we obtain

$$\begin{aligned}\left\|\xi_Q^\perp\right\|^2 &\ge \left\|(I-Q)\xi_P^\perp\right\|^2 - 2\left\|(I-Q)\xi_P\right\| \left\|(I-Q)\xi_P^\perp\right\| \\ &\ge (1-\varepsilon^2)\left\|\xi_P^\perp\right\|^2 - 2\varepsilon \|\xi_P\| \left\|\xi_P^\perp\right\|,\end{aligned} \tag{89}$$

where $\|(I-Q)\xi_P^\perp\| \le \|\xi_P^\perp\|$ was used in the last step.

Applying the assumed quadratic-form bound and then using (88)–(89) yields

$$\begin{aligned}\langle \xi, A\xi\rangle \le\ & L(\|\xi_P\| + \varepsilon\left\|\xi_P^\perp\right\|)^2 \\ &+ 2c\left(\|\xi_P\| + \varepsilon\left\|\xi_P^\perp\right\|\right)\left(\varepsilon\|\xi_P\| + \left\|\xi_P^\perp\right\|\right) \\ &- \lambda\Big[(1-\varepsilon^2)\left\|\xi_P^\perp\right\|^2 - 2\varepsilon\|\xi_P\|\left\|\xi_P^\perp\right\|\Big].\end{aligned}$$

Collecting the $\|\xi_P\|^2$ terms gives $L + 2c\varepsilon = \widetilde{L}$. The coefficient of $2\|\xi_P\|\|\xi_P^\perp\|$ is

$$L\varepsilon + c(1+\varepsilon^2) + \lambda\varepsilon = c + (L+\lambda)\varepsilon + c\varepsilon^2 = \widetilde{c},$$

and the coefficient of $\|\xi_P^\perp\|^2$ is

$$L\varepsilon^2 + 2c\varepsilon - \lambda(1-\varepsilon^2) = -\big[\lambda - 2c\varepsilon - (L+\lambda)\varepsilon^2\big] = -\widetilde{\lambda}.$$

This proves the claimed bound. □

Lemma E.4. *Fix $\bar{r} < \tau$. There exists $C_{\mathrm{proj}}(\bar{r}, \mathcal{M}) < \infty$ such that, for every $t \in [t^\star, 1]$ and $x, z \in \mathbb{R}^d$ with $x/t, z/t \in \mathcal{M}_{\bar{r}}$,*

$$\left\|P_T^{(t)}(z) - P_T^{(t)}(x)\right\|_{\mathrm{op}} \le C_{\mathrm{proj}}(\bar{r}, \mathcal{M})\ \|z - x\|. \tag{90}$$

Proof. On the compact and smooth manifold $\mathcal{M}$, the projector field $y \mapsto P_T(y)$ is C^1, hence $L_{P,\mathcal{M}}$-Lipschitz for some constant $L_{P,\mathcal{M}} > 0$. By the Lipschitz bound on the nearest-point projection on $\mathcal{M}_{\bar{r}}$ [14, Theorem 4.8(8)],

$$\|\pi_{\mathcal{M}}(z/t) - \pi_{\mathcal{M}}(x/t)\| \le \frac{\tau}{\tau - \bar{r}}\|z/t - x/t\| \le \frac{2\tau}{\tau - \bar{r}}\|z - x\|$$

using $t \ge t^\star \ge 1/2$. Hence (90) follows with $C_{\mathrm{proj}} = 2\tau L_{P,\mathcal{M}}/(\tau - \bar{r})$. □

E.2.2 Stopped tubular localization

The late-time propagation argument is local in tube coordinates: it requires the mean-value segment between the population and learned trajectories to stay in a region where the tangent–normal splitting is comparable. We enforce this requirement by stopping the coupled dynamics before the tube-local assumptions can fail. Recall that $\widetilde{X}_t$ denotes the population ODE trajectory. Under the regularity of $v_t^\star$ and the associated uniqueness of solutions to the continuity equation (or the ODE flow), its marginal law agrees with that of the interpolant X_t for each t, as we have assumed in Section 5.

Definition E.5 (Late-time stopping rule)**.** Fix $t^\star = 1/(1+\sigma)$, tube radii $0 < r < \bar{r} < \tau/4$, and a threshold $\delta_* > 0$. For two late-time trajectories $\widetilde{X}_t$ and $\widehat{X}_t$, define $\Delta_t := \widehat{X}_t - \widetilde{X}_t$ and

$$\theta := \inf\Big\{t \in [t^\star, 1] : \ \widetilde{X}_t \notin \mathcal{M}^{\mathrm{sc}}_{t,r} \text{ or } \|\Delta_t\| > \delta_*\Big\}. \tag{91}$$

The associated survival event is

$$G := \{\theta = 1\} \cap \{\widetilde{X}_1 \in \mathcal{M}^{\mathrm{sc}}_{1,r}\} \cap \{\|\Delta_1\| \le \delta_*\}. \tag{92}$$

The stopping rule compares the two trajectories only while the population path remains in the inner scaled tube and the learned path remains within distance δ_* of it. Choose δ_* small enough that

$$\delta_* \le \frac{\bar{r} - r}{2} \wedge \frac{1}{4C_{\mathrm{proj}}(\bar{r}, \mathcal{M})}. \tag{93}$$

Then, for every $t \in [t^\star, \theta)$, both $\widehat{X}_t$ and the full segment $\widetilde{X}_t + s\Delta_t$, $s \in [0,1]$, lie in the enlarged scaled tube $\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$, and Lemma E.4 gives

$$\Big\|P_T^{(t)}(\widetilde{X}_t + s\Delta_t) - P_T^{(t)}(\widetilde{X}_t)\Big\|_{\mathrm{op}} \le C_{\mathrm{proj}}(\bar{r}, \mathcal{M})\,\|\Delta_t\| \le \frac{1}{4}.$$

Thus the stopped interval $[t^\star, \theta)$ is precisely the interval on which the tube-local block-Jacobian analysis can be applied along the entire mean-value segment.

Lemma E.6 (Stopped tube geometry and transferred block bounds)**.** *For the outer tube radius $\bar{r}$, write*

$$L_T(t;\bar{r}) := 2 + C_{\mathrm{sc}}\sigma + \frac{8\kappa\bar{r}}{3\sigma_t}, \qquad c_{TN}(t) := 2C_{\mathrm{sc}}\sigma_t, \tag{94}$$

and define

$$L_* := \sup_{t \in [t^\star, t^\dagger]} L_T(t;\bar{r}), \qquad c_* := \sup_{t \in [t^\star, t^\dagger]} c_{TN}(t). \tag{95}$$

Suppose that the threshold δ_* satisfies

$$C_{\mathrm{proj}}(\bar{r}, \mathcal{M})\delta_* \le \eta_* := \min\left\{\frac{1}{128(c_* \vee 1)}, \frac{1}{8\sqrt{L_* \vee 1}}, \frac{1}{2\sqrt{2}}\right\}. \tag{96}$$

Then, for every $t \in [t^\star, \theta)$, the following hold.

(a) $\widetilde{X}_t$, $\widehat{X}_t$, and $z_s := \widetilde{X}_t + s\Delta_t$ for all $s \in [0,1]$ lie in $\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$.

(b) For $t \le t^\dagger$, define

$$J_t := \int_0^1 \nabla_x v_t^\star(\widetilde{X}_t + s\Delta_t)\,\mathrm{d}s.$$

Then J_t satisfies the transferred block bound in the base splitting $P_T^{(t)}(\widetilde{X}_t) \oplus P_N^{(t)}(\widetilde{X}_t)$:

$$\langle \xi, J_t\xi\rangle \le \widetilde{L}_T(t)\,\|\xi_T\|^2 + 2\widetilde{c}_{TN}(t)\,\|\xi_T\|\,\|\xi_N\| - \widetilde{\lambda}_N(t)\,\|\xi_N\|^2, \tag{97}$$

where

$$\widetilde{L}_T(t) := L_T(t;\bar{r}) + 2c_{TN}(t)\eta_*, \tag{98}$$

$$\widetilde{c}_{TN}(t) := c_{TN}(t) + \Big(L_T(t;\bar{r}) + \tfrac{1}{2}\lambda_N^{\mathrm{flat}}(t)\Big)\eta_* + c_{TN}(t)\eta_*^2, \tag{99}$$

$$\widetilde{\lambda}_N(t) := \tfrac{1}{2}\lambda_N^{\mathrm{flat}}(t) - 2c_{TN}(t)\eta_* - \Big(L_T(t;\bar{r}) + \tfrac{1}{2}\lambda_N^{\mathrm{flat}}(t)\Big)\eta_*^2. \tag{100}$$

In particular, $\widetilde{\lambda}_N(t) \ge \lambda_N^{\mathrm{flat}}(t)/4$ on $[t^\star, t^\dagger]$.

(c) For $t \in [t^\dagger, 1]$, the same averaged Jacobian satisfies $\|J_t\|_{\mathrm{op}} \le C_J$, where C_J is independent of any negative power of σ.

Proof. We prove the three claims separately.

Step 1: tube containment before stopping. Fix $t \in [t^\star, \theta)$. By definition of θ, the population trajectory satisfies $\widetilde{X}_t \in \mathcal{M}^{\mathrm{sc}}_{t,r}$ and $\|\Delta_t\| \le \delta_*$. Hence there exists $y \in \mathcal{M}$ such that

$$\Big\|\widetilde{X}_t - ty\Big\| \le tr.$$

For the learned trajectory,

$$\left\|\widehat{X}_t - ty\right\| \le \left\|\widetilde{X}_t - ty\right\| + \|\Delta_t\| \le tr + \delta_*.$$

Since $t \in [t^\star, 1]$ and $t^\star = 1/(1+\sigma) \ge 1/2$, the choice $\delta_* \le (\bar{r} - r)/2$ implies $\delta_* \le t(\bar{r} - r)$. Therefore $\widehat{X}_t \in \mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$. The same calculation gives, for every $s \in [0,1]$,

$$\left\|\widetilde{X}_t + s\Delta_t - ty\right\| \le tr + s\,\|\Delta_t\| \le tr + \delta_* \le t\bar{r},$$

so $z_s := \widetilde{X}_t + s\Delta_t \in \mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$. This proves part (a).

Step 2: projector comparison along the segment. By Lemma E.4, uniformly in $s \in [0,1]$,

$$\left\|P_T^{(t)}(z_s) - P_T^{(t)}(\widetilde{X}_t)\right\|_{\mathrm{op}} \le C_{\mathrm{proj}}(\bar{r}, \mathcal{M}) \left\|z_s - \widetilde{X}_t\right\| \le C_{\mathrm{proj}}(\bar{r}, \mathcal{M})\delta_* \le \eta_*. \tag{101}$$

Here part (a) ensures that both points are in the domain where the projection field is Lipschitz.

Step 3: pointwise block transfer. Fix $s \in [0,1]$ and $t \in [t^\star, t^\dagger] \cap [t^\star, \theta)$. Since $z_s \in \mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$, Theorem 4.4 applies at z_s with tube radius $\bar{r}$. Thus, in the moving splitting at z_s, the quadratic form of $\nabla_x v_t^\star(z_s)$ is bounded with tangential envelope $L_T(t;\bar{r})$, mixed envelope $c_{TN}(t)$, and normal contraction $\lambda_N^{\mathrm{flat}}(t)/2$. The projector distance between this splitting and the base splitting at $\widetilde{X}_t$ is bounded by (101). Applying Lemma E.3 with

$$L = L_T(t;\bar{r}), \qquad c = c_{TN}(t), \qquad \lambda = \tfrac{1}{2}\lambda_N^{\mathrm{flat}}(t), \qquad \varepsilon \le \eta_*,$$

gives, in the base splitting $P_T^{(t)}(\widetilde{X}_t) \oplus P_N^{(t)}(\widetilde{X}_t)$,

$$\langle \xi, \nabla_x v_t^\star(z_s)\xi\rangle \le \widetilde{L}_T(t)\,\|\xi_T\|^2 + 2\widetilde{c}_{TN}(t)\,\|\xi_T\|\,\|\xi_N\| - \widetilde{\lambda}_N(t)\,\|\xi_N\|^2, \tag{102}$$

with the envelopes in (98)–(100).

Step 4: averaging and the normal-contraction lower bound. Integrating (102) over $s \in [0,1]$ gives (97), because the right-hand side is independent of s and J_t is the s-average of $\nabla_x v_t^\star(z_s)$. It remains to justify that the transferred normal block is still contractive. On $[t^\star, t^\dagger]$, $\lambda_N^{\mathrm{flat}}(t) \ge 1/4$. Also $c_{TN}(t) \le c_*$ and $L_T(t;\bar{r}) \le L_*$. By (96),

$$2c_{TN}(t)\eta_* \le 2c_*\eta_* \le \frac{1}{64},$$

and

$$\Big(L_T(t;\bar{r}) + \tfrac{1}{2}\lambda_N^{\mathrm{flat}}(t)\Big)\eta_*^2 \le L_*\eta_*^2 + \frac{1}{2}\lambda_N^{\mathrm{flat}}(t)\eta_*^2 \le \frac{1}{64} + \frac{1}{16}\lambda_N^{\mathrm{flat}}(t),$$

where the last inequality uses $\eta_* \le 1/(8\sqrt{L_* \vee 1})$ and $\eta_* \le 1/(2\sqrt{2})$. Since $\lambda_N^{\mathrm{flat}}(t) \ge 1/4$, the total contraction loss is bounded by

$$\frac{1}{32} + \frac{1}{16}\lambda_N^{\mathrm{flat}}(t) \le \frac{1}{8}\lambda_N^{\mathrm{flat}}(t) + \frac{1}{16}\lambda_N^{\mathrm{flat}}(t) \le \frac{1}{4}\lambda_N^{\mathrm{flat}}(t).$$

Hence $\widetilde{\lambda}_N(t) \ge \lambda_N^{\mathrm{flat}}(t)/4$, proving part (b).

Step 5: terminal-window operator norm. For $t \in [t^\dagger, 1]$, part (a) still gives $z_s \in \mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$. On this terminal window, $1 - t = O(\sigma^2)$ and $\sigma_t \asymp \sigma$. The Hessian block bounds in Proposition 4.2, together with the score-form identity (12), imply a uniform operator-norm bound for $\nabla_x v_t^\star(z_s)$ on the enlarged tube: the prefactor $1 - t$ in (12) cancels the possible σ_t^{-2} normal Hessian singularity. Thus $\|\nabla_x v_t^\star(z_s)\|_{\mathrm{op}} \le C_J$ uniformly in s, with C_J independent of any negative power of σ. Averaging over s gives $\|J_t\|_{\mathrm{op}} \le C_J$, proving part (c). □

E.2.3 Late-time anisotropic propagation

Theorem E.7 (Stopped anisotropic flow perturbation)**.** *Fix $0 < r < \bar{r} < \tau/4$ and suppose $\sigma \le \sigma_0(\bar{r}, \mathcal{M}, f)$. Consider two ODE flows on the late interval $[t^\star, 1]$:*

$$\dot{\widetilde{X}}_t = v_t^\star(\widetilde{X}_t), \qquad \dot{\widehat{X}}_t = \widehat{v}_t(\widehat{X}_t),$$

with initial conditions $(\widetilde{X}_{t^\star}, \widehat{X}_{t^\star})$ coupled arbitrarily. Recall $\Delta_t = \widehat{X}_t - \widetilde{X}_t$, define $\mathcal{E}_t := \|\Delta_t\|^2$, and let θ be the stopping time in (91)*. Suppose that δ_* satisfies* (93) *and* (96)*. Then*

$$\sup_{t\in[t^\star,1]} \mathcal{E}_{t\wedge\theta} \le e^{\Psi_{\mathrm{eff}}}\left(\mathcal{E}_{t^\star} + \int_{t^\star}^{\theta} \left\|\widehat{v}_s(\widehat{X}_s) - v_s^\star(\widehat{X}_s)\right\|^2 \mathrm{d}s\right), \tag{103}$$

where the effective exponent is defined by

$$a(t) := \begin{cases} 1 + 2\widetilde{L}_T(t) + \dfrac{16\widetilde{c}_{TN}(t)^2}{\widetilde{\lambda}_N(t)}, & t \in [t^\star, t^\dagger], \\ 1 + 2C_J, & t \in [t^\dagger, 1], \end{cases} \qquad \Psi_{\mathrm{eff}} := \int_{t^\star}^{1} a(t)\,\mathrm{d}t. \tag{104}$$

Here $\widetilde{L}_T, \widetilde{c}_{TN}, \widetilde{\lambda}_N$ and C_J are defined in Lemma E.6; in particular, C_J is independent of any negative power of σ. Consequently,

$$\mathbb{E}\sup_{t\in[t^\star,1]} \mathcal{E}_{t\wedge\theta} \le e^{\Psi_{\mathrm{eff}}}\Big(\mathbb{E}[\mathcal{E}_{t^\star}] + \Phi_\theta\Big), \qquad \Phi_\theta := \mathbb{E}\int_{t^\star}^{\theta} \left\|\widehat{v}_s(\widehat{X}_s) - v_s^\star(\widehat{X}_s)\right\|^2 \mathrm{d}s. \tag{105}$$

Proof. Write $e_t := \widehat{v}_t - v_t^\star$. For $t < \theta$, define

$$J_t := \int_0^1 \nabla_x v_t^\star(\widetilde{X}_t + s\Delta_t)\,\mathrm{d}s.$$

Then

$$\dot{\Delta}_t = v_t^\star(\widehat{X}_t) - v_t^\star(\widetilde{X}_t) + e_t(\widehat{X}_t) = J_t\Delta_t + e_t(\widehat{X}_t),$$

where the second equality follows from the fundamental theorem of calculus along the segment $\widetilde{X}_t + s\Delta_t$. Set $P_t := P_T^{(t)}(\widetilde{X}_t)$, $Q_t := I - P_t$, $\Delta_T := P_t\Delta_t$, and $\Delta_N := Q_t\Delta_t$. Differentiating $\mathcal{E}_t$ on $[t^\star, \theta)$ gives

$$\dot{\mathcal{E}}_t = 2\langle \Delta_t, J_t\Delta_t\rangle + 2\left\langle \Delta_t, e_t(\widehat{X}_t)\right\rangle.$$

On $[t^\star, t^\dagger] \cap [t^\star, \theta)$, Lemma E.6 gives

$$2\langle \Delta_t, J_t\Delta_t\rangle \le 2\widetilde{L}_T(t)\|\Delta_T\|^2 - 2\widetilde{\lambda}_N(t)\|\Delta_N\|^2 + 4\widetilde{c}_{TN}(t)\|\Delta_T\|\,\|\Delta_N\|.$$

Young's inequality at the normal-contraction scale yields

$$4\widetilde{c}_{TN}(t)\|\Delta_T\|\,\|\Delta_N\| \le \frac{16\widetilde{c}_{TN}(t)^2}{\widetilde{\lambda}_N(t)}\|\Delta_T\|^2 + \widetilde{\lambda}_N(t)\|\Delta_N\|^2.$$

Thus

$$\dot{\mathcal{E}}_t \le \left(2\widetilde{L}_T(t) + \frac{16\widetilde{c}_{TN}(t)^2}{\widetilde{\lambda}_N(t)}\right)\|\Delta_T\|^2 - \widetilde{\lambda}_N(t)\|\Delta_N\|^2 + 2\left\langle \Delta_t, e_t(\widehat{X}_t)\right\rangle. \tag{106}$$

Moreover,

$$2\left\langle \Delta_t, e_t(\widehat{X}_t)\right\rangle \le \mathcal{E}_t + \left\|e_t(\widehat{X}_t)\right\|^2. \tag{107}$$

Dropping the negative normal term and using $\|\Delta_T\|^2 \le \mathcal{E}_t$ gives

$$\dot{\mathcal{E}}_t \le a(t)\mathcal{E}_t + \left\|e_t(\widehat{X}_t)\right\|^2, \qquad a(t) := 1 + 2\widetilde{L}_T(t) + \frac{16\widetilde{c}_{TN}(t)^2}{\widetilde{\lambda}_N(t)}, \qquad t \in [t^\star, t^\dagger] \cap [t^\star, \theta). \tag{108}$$

This is the anisotropic step: the normal block remains dissipative before scalarization, while the mixed block contributes only the lower-order ratio $\widetilde{c}_{TN}^2/\widetilde{\lambda}_N$ to the tangential envelope.

On $[t^\dagger, 1] \cap [t^\star, \theta)$, Lemma E.6 gives $\|J_t\|_{\mathrm{op}} \le C_J$. Therefore,

$$\dot{\mathcal{E}}_t \le (1 + 2C_J)\mathcal{E}_t + \left\|e_t(\widehat{X}_t)\right\|^2, \qquad t \in [t^\dagger, 1] \cap [t^\star, \theta). \tag{109}$$

Combining (108) and (109), we have $\dot{\mathcal{E}}_t \le a(t)\mathcal{E}_t + \|e_t(\widehat{X}_t)\|^2$ for $t < \theta$, with $a(t)$ defined in (104). Grönwall's inequality gives, for every $u \le \theta$,

$$\mathcal{E}_u \le e^{\Psi_{\mathrm{eff}}}\left(\mathcal{E}_{t^\star} + \int_{t^\star}^{u} \left\|e_s(\widehat{X}_s)\right\|^2 \mathrm{d}s\right).$$

Taking the supremum over $u = t \wedge \theta$, $t \in [t^\star, 1]$, yields (103). Taking expectations gives (105). □

Proposition E.8. *Under the conditions of Theorem E.7,*

$$\Psi_{\mathrm{eff}} \le C\bigg(1 + \int_{t^\star}^{t^\dagger} L_T(t;\bar{r})\,\mathrm{d}t + \sigma\bigg), \tag{110}$$

where C depends on $\bar{r}, \mathcal{M}, f$ but not on σ.

Proof. By (104),

$$\Psi_{\mathrm{eff}} = \int_{t^\star}^{t^\dagger} \left(1 + 2\widetilde{L}_T(t) + \frac{16\widetilde{c}_{TN}(t)^2}{\widetilde{\lambda}_N(t)}\right)\mathrm{d}t + (1+2C_J)(1-t^\dagger).$$

We bound the terms beyond the principal tangential budget. First, $t^\dagger - t^\star = O(\sigma)$ and $1-t^\dagger = O(\sigma^2)$, while C_J is independent of any negative power of σ. Hence the unit contribution and the terminal-window contribution are $O(\sigma)$.

Next, by Lemma E.6,

$$\widetilde{L}_T(t) = L_T(t;\bar{r}) + 2c_{TN}(t)\eta_*.$$

Since $c_{TN}(t) = 2C_{\mathrm{sc}}\sigma_t$ and $\eta_* \le 1$, the transferred tangential correction satisfies

$$\int_{t^\star}^{t^\dagger} 4c_{TN}(t)\eta_*\,\mathrm{d}t \le C\int_{t^\star}^{t^\dagger} \sigma_t\,\mathrm{d}t \le C\sigma^2.$$

It remains to control the mixed term. Lemma E.6 gives $\widetilde{\lambda}_N(t) \ge \lambda_N^{\mathrm{flat}}(t)/4$, and

$$\widetilde{c}_{TN}(t) \le c_{TN}(t) + \Big(L_T(t;\bar{r}) + \tfrac{1}{2}\lambda_N^{\mathrm{flat}}(t)\Big)\eta_* + c_{TN}(t)\eta_*^2.$$

Using $(a+b+c)^2 \le 3(a^2+b^2+c^2)$, it is enough to bound the integrated contributions

$$\int_{t^\star}^{t^\dagger} \frac{c_{TN}(t)^2}{\lambda_N^{\mathrm{flat}}(t)}\,\mathrm{d}t, \quad \int_{t^\star}^{t^\dagger} \frac{L_T(t;\bar{r})^2\eta_*^2}{\lambda_N^{\mathrm{flat}}(t)}\,\mathrm{d}t, \quad \int_{t^\star}^{t^\dagger} \lambda_N^{\mathrm{flat}}(t)\eta_*^2\,\mathrm{d}t,$$

together with the smaller duplicate term involving $c_{TN}(t)^2\eta_*^4/\lambda_N^{\mathrm{flat}}(t)$.

For the terms involving c_{TN}, use $\lambda_N^{\mathrm{flat}}(t) \ge 1/4$ and $c_{TN}(t) = O(\sigma_t)$ to get

$$\int_{t^\star}^{t^\dagger} \frac{c_{TN}(t)^2}{\lambda_N^{\mathrm{flat}}(t)}\,\mathrm{d}t + \int_{t^\star}^{t^\dagger} \frac{c_{TN}(t)^2\eta_*^4}{\lambda_N^{\mathrm{flat}}(t)}\,\mathrm{d}t \le C\int_{t^\star}^{t^\dagger} \sigma_t^2\,\mathrm{d}t \le C\sigma^3.$$

For the transferred tangential part, $L_T(t;\bar{r}) \le L_*$ and $L_*\eta_*^2 \le C$, so

$$\int_{t^\star}^{t^\dagger} \frac{L_T(t;\bar{r})^2\eta_*^2}{\lambda_N^{\mathrm{flat}}(t)}\,\mathrm{d}t \le C\int_{t^\star}^{t^\dagger} L_T(t;\bar{r})\,\mathrm{d}t.$$

Finally, although $\lambda_N^{\mathrm{flat}}(t)$ can be as large as $O(\sigma^{-1})$ near $t^\star$, this singularity is harmless on the contractive window. Indeed,

$$\lambda_N^{\mathrm{flat}}(t) = \frac{(1-t) - t\sigma^2}{\sigma_t^2} \le \frac{1-t}{\sigma_t^2} \le \frac{1}{\sigma_t},$$

and $\int_{t^\star}^{t^\dagger} \sigma_t^{-1}\,\mathrm{d}t \le C$. Therefore

$$\int_{t^\star}^{t^\dagger} \lambda_N^{\mathrm{flat}}(t)\eta_*^2\,\mathrm{d}t \le C.$$

Combining the above bounds proves (110), after increasing C. □

Remark E.9. A naive scalar argument that inserts an ambient Lipschitz envelope $\mathrm{Lip}(v_t^\star) = O(\sigma_t^{-3})$ over $[0,1]$ into Grönwall yields $\int_0^1 \sigma_t^{-3}\,\mathrm{d}t = O(\sigma^{-2})$ and hence $\exp(O(\sigma^{-2}))$ amplification. By contrast, the anisotropic functional (104) separates the expansive and contractive directions. The tangential growth contributes through the integrable budget $\int_{t^\star}^{t^\dagger} L_T(t;\bar{r})\,\mathrm{d}t$, the normal block is dissipative, and the mixed tangential–normal block is absorbed at the normal-contraction scale before scalarization. Proposition E.8 therefore gives a bounded $\Psi_{\mathrm{eff}} = O(1 + \int_{t^\star}^{t^\dagger} L_T(t;\bar{r})\,\mathrm{d}t + \sigma)$, rather than a negative power of σ. Note that the explicit constant accounts for the integrated effect of transferring nearby tangent–normal splittings through the normal-contraction scale.

E.3 Stopped tube-to-ambient synthesis

This subsection turns the stopped late-time energy bound into an ambient W_2 statement. Theorem E.7 starts at $t^\star$ and controls the forcing along the stopped learned trajectory. We first formulate an early coupling budget and a density-ratio transfer from population tube risk to this forcing. We then derive the exit probability from the population-flow confinement bound and the stopped perturbation energy, and complete the transport synthesis under a terminal-moment assumption.

E.3.1 The early budget at the switching time

Proposition E.10. *Over $[0, t^\star]$ with $t^\star = 1/(1+\sigma)$, suppose $\mathrm{Lip}(\widehat{v}_t) \le C/(1-t)$ for some constant $C > 0$. Then there exists a constant C_0 such that*

$$\mathrm{W}_2(\mathcal{L}(\widehat{X}_{t^\star}), \mathcal{L}(X_{t^\star})) \le (1+\sigma)^{C_0/2} \sigma^{-C_0/2} \cdot \left(\int_0^{t^\star} \|\widehat{v}_t - v_t^\star\|^2_{L^2(p_t)} \, \mathrm{d}t \right)^{1/2}. \tag{111}$$

One may take $C_0 = 2C$.

Proof. By the Alekseev–Gröbner stability bound [18, Lemma 16 and Theorem 3],

$$\mathrm{W}_2^2(\mathcal{L}(\widehat{X}_t), \mathcal{L}(X_t)) \le t \int_0^t e^{2\int_s^t L_u \, \mathrm{d}u} \, \|\widehat{v}_s - v_s^\star\|^2_{L^2(p_s)} \, \mathrm{d}s.$$

With $L_u \le C/(1-u)$, $\int_s^{t^\star} L_u \, \mathrm{d}u \le C \log((1+\sigma)/\sigma)$, so $e^{2\int_s^{t^\star} L_u \, \mathrm{d}u} \le (1+\sigma)^{2C} \sigma^{-2C}$. □

Corollary E.11. *Under the conditions of Proposition E.10, suppose the global early risk satisfies, for some constant C_{glob},*

$$\int_0^{t^\star} \|\widehat{v}_t - v_t^\star\|^2_{L^2(p_t)} \, \mathrm{d}t \le C_{\mathrm{glob}} \, \mathcal{R}_r(\widehat{v}),$$

and the coupling at the switching time satisfies, for some constant C_{coup},

$$\mathbb{E}[\mathcal{E}_{t^\star}] \le C_{\mathrm{coup}} \, \mathrm{W}_2^2(\mathcal{L}(\widehat{X}_{t^\star}), \mathcal{L}(X_{t^\star})).$$

Then

$$B_{\mathrm{early}} := C \, C_{\mathrm{coup}} \, C_{\mathrm{glob}} \, \sigma^{-C_0} \, \mathcal{R}_r(\widehat{v}) \tag{112}$$

is an admissible early budget, in the sense that $\mathbb{E}[\mathcal{E}_{t^\star}] \le B_{\mathrm{early}}$.

E.3.2 Density-ratio transfer of the stopped forcing

The stopped energy bound depends on the forcing term evaluated along the learned trajectory,

$$\Phi_\theta = \mathbb{E} \int_{t^\star}^{\theta} \left\| \widehat{v}_t(\widehat{X}_t) - v_t^\star(\widehat{X}_t) \right\|^2 \, \mathrm{d}t. \tag{113}$$

This quantity is a distribution-shift version of the late tube risk: the integrand is the usual squared velocity error, but the evaluation distribution is the learned flow rather than the population marginal. The following condition gives a direct transfer.

Definition E.12. For an outer scaled tube radius $\bar{r}$, define

$$\Lambda_{\mathrm{sc}}(\bar{r}) := \sup_{t \in [t^\star, 1]} \operatorname*{ess\,sup}_{x \in \mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}} \frac{\widehat{p}_t}{p_t}(x), \tag{114}$$

whenever $\widehat{p}_t \ll p_t$ on $\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$ for every $t \in [t^\star, 1]$.

Lemma E.13. *Assume that $\widehat{p}_t \ll p_t$ on $\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$ for every $t \in [t^\star, 1]$ and that $\Lambda_{\mathrm{sc}}(\bar{r}) < \infty$. Then*

$$\Phi_\theta \le \Lambda_{\mathrm{sc}}(\bar{r}) \, \mathcal{R}^{\mathrm{late}}_{\bar{r}_+}(\widehat{v}), \qquad \bar{r}_+ := \bar{r} + R_{\mathcal{M}} \sigma. \tag{115}$$

Proof. For $t \in [t^\star, \theta)$, Lemma E.6 gives $\widehat{X}_t \in \mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$. Therefore

$$\begin{aligned}\Phi_\theta &\le \int_{t^\star}^1 \int_{\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}} \|\widehat{v}_t(x) - v_t^\star(x)\|^2 \, \widehat{p}_t(x) \, \mathrm{d}x \, \mathrm{d}t \\ &\le \Lambda_{\mathrm{sc}}(\bar{r}) \int_{t^\star}^1 \int_{\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}} \|\widehat{v}_t(x) - v_t^\star(x)\|^2 \, p_t(x) \, \mathrm{d}x \, \mathrm{d}t.\end{aligned}$$

It remains only to compare the scaled tube with the fixed ambient tube used in Definition E.1. If $x \in \mathcal{M}^{\mathrm{sc}}_{t,\bar{r}}$, then $x = tz$ for some $z \in \mathcal{M}_{\bar{r}}$. Choose $y \in \mathcal{M}$ with $\|z - y\| \le \bar{r}$. Then

$$\mathrm{dist}(x, \mathcal{M}) \le \|x - y\| = \|tz - y\| \le t \, \|z - y\| + (1 - t) \, \|y\| \le \bar{r} + R_{\mathcal{M}}\sigma,$$

where we used $t \in [t^\star, 1]$ and $1 - t^\star \le \sigma$. Thus $\mathcal{M}^{\mathrm{sc}}_{t,\bar{r}} \subset \mathcal{M}_{\bar{r}_+}$, which gives (115). □

Remark E.14. The transfer in Lemma E.13 is continuity-free in the velocity error: it does not require a Lipschitz or Hölder modulus for $\widehat{v}_t - v_t^\star$. This is useful to a certain degree for sieve constructions in which the late-time estimator is only piecewise smooth or chartwise defined. The statistical condition is instead the density-ratio bound (114), which quantifies how much the learned flow can reweight the population marginal inside the tube-local region.

Remark E.15. Although Lemma E.13 does not require continuity of the velocity error as an integrand, establishing (114) imposes regularity on the learned flow. In particular, a cellwise-constant late velocity may exhibit compressive jumps across cell boundaries, causing characteristics to coalesce and potentially producing singular components in the transported marginal, even when the distribution at the switching time is absolutely continuous [7, 40]. A natural refinement is a continuous, slope-controlled tube-adapted class whose normal component is matched, or matched to the required late-time order, to the population normal velocity on a tube plateau [3]. Together with a positive source-density construction at the switching time, this structure permits density ratios to be propagated through the continuity equation while retaining intrinsic complexity of order h^{-k} and the same approximation order as the late sieve (133) below. Thus (114) should be viewed as a criterion for designing the late hypothesis class, not as a consequence of tube-risk control alone. Developing this refined class within the present statistical construction is left for future work.

E.3.3 Exit probability of the stopped coupling

The stopped energy bound also controls the probability that the coupled trajectories exit the tube-local comparison regime. Define the population-flow tube-exit probability

$$\delta_X(r, \sigma) := \mathbb{P}\big(\exists \, t \in [t^\star, 1] : \ \widetilde{X}_t \notin \mathcal{M}^{\mathrm{sc}}_{t,r}\big). \tag{116}$$

Because $\widetilde{X}_t$ is assumed to have the same marginal law as the interpolant X_t, this is a pathwise strengthening of the tube localization in Lemma 2.3.

Lemma E.16 (Stopped-coupling exit probability)**.** *Under the conditions of Theorem E.7,*

$$\mathbb{P}(G^c) \le \delta_X(r, \sigma) + \delta_*^{-2} \, \mathbb{E} \sup_{t \in [t^\star, 1]} \mathcal{E}_{t \wedge \theta}. \tag{117}$$

Consequently, if $\mathbb{E}[\mathcal{E}_{t^\star}] \le B_{\mathrm{early}}$ *and the assumptions of Lemma E.13 hold, then*

$$\mathbb{P}(G^c) \le q_* := \delta_X(r, \sigma) + \delta_*^{-2} \mathcal{S}, \tag{118}$$

where

$$\mathcal{S} := e^{\Psi_{\mathrm{eff}}} \Big(B_{\mathrm{early}} + \Lambda_{\mathrm{sc}}(\bar{r}) \, \mathcal{R}^{\mathrm{late}}_{\bar{r}_+}(\widehat{v}) \Big). \tag{119}$$

Proof. If G fails, then either the population path exits the inner scaled tube during $[t^\star, 1]$, or the displacement threshold is exceeded at some time before the terminal survival event holds. Thus

$$G^c \subseteq \Big\{ \exists \, t \in [t^\star, 1] : \ \widetilde{X}_t \notin \mathcal{M}^{\mathrm{sc}}_{t,r} \Big\} \cup \Big\{ \sup_{t \in [t^\star, 1]} \mathcal{E}_{t \wedge \theta} \ge \delta_*^2 \Big\}.$$

Taking probabilities and applying Markov's inequality gives (117). Theorem E.7 gives

$$\mathbb{E} \sup_{t \in [t^\star, 1]} \mathcal{E}_{t \wedge \theta} \le e^{\Psi_{\mathrm{eff}}} \big(\mathbb{E}[\mathcal{E}_{t^\star}] + \Phi_\theta \big),$$

and Lemma E.13 gives

$$\Phi_\theta \le \Lambda_{\mathrm{sc}}(\bar{r})\, \mathcal{R}^{\mathrm{late}}_{\bar{r}_+}(\widehat{v}).$$

Combining these bounds with $\mathbb{E}[\mathcal{E}_{t^\star}] \le B_{\mathrm{early}}$ proves (118). □

Proposition E.17 (Pathwise confinement of the population late flow)**.** *Let $r_0 < \tau/4$ be the master tube radius in Theorem 4.4, and let $\sigma_0(r_0, \mathcal{M}, f)$ denote the corresponding geometric small-noise threshold. For every fixed $\vartheta \in (0, 1/2)$, there exists a threshold $\sigma_{0,\vartheta} \in (0, \sigma_0(r_0, \mathcal{M}, f)]$ such that, for every $\sigma \le \sigma_{0,\vartheta}$ and every $0 < r \le r_0$,*

$$\delta_X(r, \sigma) \le e(1 - 2\vartheta)^{-d/2} \exp\left(-\frac{\vartheta}{2}\frac{r^2}{\sigma^2}\right). \tag{120}$$

Proof. Set $Z_t := \widetilde{X}_t/t$ and $R_t := \mathrm{dist}(Z_t, \mathcal{M}) = \|\rho_{\mathcal{M}}(Z_t)\|$. Since we assume that $\widetilde{X}_t$ has marginal law p_t, Lemma 2.2 gives

$$Z_{t^\star} \sim (f \,\mathrm{Vol}_{\mathcal{M}}) * \mathcal{N}(0, 2\sigma^2 I_d).$$

Moreover, by the velocity–score identity (74),

$$\dot{Z}_t = \frac{t\dot{\widetilde{X}}_t - \widetilde{X}_t}{t^2} = \frac{1-t}{t^2}\nabla \log p_t(\widetilde{X}_t). \tag{121}$$

We first show that, up to an $O(\sigma^2)$ equilibrium radius, leaving the scaled tube is controlled by the initial position at $t^\star$.

While $Z_t \in \mathcal{M}_{r_0}$, write $n_t := \rho_{\mathcal{M}}(Z_t)$. The score expansion (13) gives

$$\nabla \log p_t(\widetilde{X}_t) = -\frac{t}{\sigma_t^2} n_t + b_t(\widetilde{X}_t), \qquad \left\| b_t(\widetilde{X}_t) \right\| \le C_{\mathrm{sc}}.$$

Inside the reach tube, $\phi(z) := \mathrm{dist}(z, \mathcal{M})^2/2$ is smooth and $\nabla\phi(z) = \rho_{\mathcal{M}}(z)$. Hence $V_t := \phi(Z_t) = R_t^2/2$ satisfies, as long as $Z_t \in \mathcal{M}_{r_0}$,

$$\begin{aligned}\dot{V}_t = \left\langle n_t, \dot{Z}_t \right\rangle &= -\frac{1-t}{t\sigma_t^2} R_t^2 + \frac{1-t}{t^2}\left\langle n_t, b_t(\widetilde{X}_t)\right\rangle \\ &\le -\lambda_t R_t\big(R_t - h_\sigma\big), \qquad \lambda_t := \frac{1-t}{t\sigma_t^2},\end{aligned}$$

where $h_\sigma := 4C_{\mathrm{sc}}\sigma^2$, using $t \ge 1/2$ and $\sigma_t^2 \le 2\sigma^2$ on $[t^\star, 1]$. We choose $\sigma_{0,\vartheta}$ small enough that $h_\sigma < r_0$ for all $\sigma \le \sigma_{0,\vartheta}$.

Fix $r \in (h_\sigma, r_0]$. On the event $\{R_{t^\star} < r\}$, the trajectory cannot hit level r during $[t^\star, 1]$: if t_1 were the first hitting time, then $R_{t_1} = r$ and, just before t_1, $R_t > h_\sigma$ on a short interval. On that interval $\dot{V}_t \le 0$, contradicting the first increase to level r. Therefore

$$\big\{\exists\, t \in [t^\star, 1] :\ R_t \ge r\big\} \subseteq \{R_{t^\star} \ge r\}, \qquad r \in (h_\sigma, r_0]. \tag{122}$$

Since $\widetilde{X}_t \notin \mathcal{M}^{\mathrm{sc}}_{t,r}$ if and only if $R_t \ge r$, this gives $\delta_X(r, \sigma) \le \mathbb{P}(R_{t^\star} \ge r)$ for $r \in (h_\sigma, r_0]$.

It remains to bound the initial tail. We can write $Z_{t^\star} = Y + \sqrt{2}\sigma\xi$, with $Y \sim f\,\mathrm{Vol}_{\mathcal{M}}$, $Y \in \mathcal{M}$, and $\xi \sim \mathcal{N}(0, I_d)$ independent. Hence $R_{t^\star} \le \sqrt{2}\sigma\,\|\xi\|$. For $\vartheta \in (0, 1/2)$, Markov's inequality and the chi-square moment generating function give

$$\begin{aligned}\mathbb{P}(R_{t^\star} \ge r) &\le \mathbb{P}\left(\|\xi\|^2 \ge \frac{r^2}{2\sigma^2}\right) \\ &\le (1 - 2\vartheta)^{-d/2} \exp\left(-\frac{\vartheta}{2}\frac{r^2}{\sigma^2}\right).\end{aligned}$$

This proves (120) for $r \in (h_\sigma, r_0]$, without the prefactor e.

For $0 < r \le h_\sigma$, choose $\sigma_{0,\vartheta}$ small enough that $(\vartheta/2)h_\sigma^2/\sigma^2 \le 1$ for all $\sigma \le \sigma_{0,\vartheta}$. Then

$$e \exp\left(-\frac{\vartheta}{2}\frac{r^2}{\sigma^2}\right) \ge 1,$$

and the trivial bound $\delta_X(r, \sigma) \le 1$ proves (120) in the small-radius range as well. □

E.3.4 Terminal tail control and ambient synthesis

The last step converts the stopped coupling into an ambient transport bound. We use a moment-tail inequality to charge the complement of the survival event.

Lemma E.18. *Let (U, V) be a coupling of two random elements on $\mathbb{R}^d$, and let A be any event. Suppose that for some $\mathfrak{p} > 2$, some point $z_0 \in \mathbb{R}^d$, and some finite constant $\bar{M} > 0$,*

$$\mathbb{E}\,\|U - z_0\|^{\mathfrak{p}} + \mathbb{E}\,\|V - z_0\|^{\mathfrak{p}} \le \bar{M}. \tag{123}$$

Then

$$\mathbb{E}\big[\,\|U - V\|^2\,\mathbf{1}_{A^c}\big] \le 2^{2-2/\mathfrak{p}}\bar{M}^{2/\mathfrak{p}}\mathbb{P}(A^c)^{1-2/\mathfrak{p}}. \tag{124}$$

Proof. By the triangle inequality and $(a+b)^2 \le 2a^2 + 2b^2$,

$$\|U - V\|^2 \le 2\,\|U - z_0\|^2 + 2\,\|V - z_0\|^2.$$

For the first term, Hölder's inequality with exponents $\mathfrak{p}/2$ and $\mathfrak{p}/(\mathfrak{p}-2)$ gives

$$\mathbb{E}\big[\,\|U - z_0\|^2\,\mathbf{1}_{A^c}\big] \le \big(\mathbb{E}\,\|U - z_0\|^{\mathfrak{p}}\,\big)^{2/\mathfrak{p}}\mathbb{P}(A^c)^{1-2/\mathfrak{p}}.$$

The same bound applies to V. Since $a^{2/\mathfrak{p}} + b^{2/\mathfrak{p}} \le 2^{1-2/\mathfrak{p}}(a+b)^{2/\mathfrak{p}}$ for $a, b \ge 0$, the two terms are bounded by $2^{2-2/\mathfrak{p}}\bar{M}^{2/\mathfrak{p}}\mathbb{P}(A^c)^{1-2/\mathfrak{p}}$. □

Corollary E.19 (Stopped tube-to-ambient synthesis)**.** *Assume the conditions of Theorem E.7, Lemma E.13, and Lemma E.16. Suppose $\mathbb{E}[\mathcal{E}_{t^\star}] \le B_{\mathrm{early}}$, and assume that for some $\mathfrak{p} > 2$, $z_0 \in \mathbb{R}^d$, and finite $\bar{M} > 0$,*

$$\mathbb{E}\left\|\widehat{X}_1 - z_0\right\|^{\mathfrak{p}} + \mathbb{E}\left\|\widetilde{X}_1 - z_0\right\|^{\mathfrak{p}} \le \bar{M}. \tag{125}$$

Let $\mathcal{S}$ and q_* be defined by (119) *and* (118)*. Then*

$$\mathrm{W}_2(\widehat{\nu}_\sigma, \nu_\sigma) \le \sqrt{\mathcal{S}} + 2\bar{M}^{1/\mathfrak{p}} q_*^{1/2-1/\mathfrak{p}}. \tag{126}$$

Combining with Proposition E.17, for every fixed $\vartheta \in (0, 1/2)$ and sufficiently small σ,

$$q_* \le e(1-2\vartheta)^{-d/2}\exp\left(-\frac{\vartheta}{2}\frac{r^2}{\sigma^2}\right) + \delta_*^{-2}\mathcal{S}. \tag{127}$$

Proof. The synchronous terminal pair $(\widehat{X}_1, \widetilde{X}_1)$ is a coupling of $(\widehat{\nu}_\sigma, \nu_\sigma)$. Hence

$$\mathrm{W}_2^2(\widehat{\nu}_\sigma, \nu_\sigma) \le \mathbb{E}\left\|\widehat{X}_1 - \widetilde{X}_1\right\|^2.$$

Decompose the expectation over G and G^c. On G, $\theta = 1$, so

$$\left\|\widehat{X}_1 - \widetilde{X}_1\right\|^2 \mathbf{1}_G = \mathcal{E}_1 \mathbf{1}_G \le \sup_{t \in [t^\star, 1]} \mathcal{E}_{t\wedge\theta}.$$

Theorem E.7, the early budget, and Lemma E.13 therefore give

$$\mathbb{E}\Big[\,\left\|\widehat{X}_1 - \widetilde{X}_1\right\|^2 \mathbf{1}_G\Big] \le \mathcal{S}.$$

On G^c, Lemma E.18 with $U = \widehat{X}_1$, $V = \widetilde{X}_1$, and $A = G$ gives

$$\mathbb{E}\Big[\,\left\|\widehat{X}_1 - \widetilde{X}_1\right\|^2 \mathbf{1}_{G^c}\Big] \le 2^{2-2/\mathfrak{p}}\bar{M}^{2/\mathfrak{p}}\mathbb{P}(G^c)^{1-2/\mathfrak{p}} \le 2^{2-2/\mathfrak{p}}\bar{M}^{2/\mathfrak{p}} q_*^{1-2/\mathfrak{p}},$$

where the last step uses Lemma E.16. Thus

$$\mathrm{W}_2^2(\widehat{\nu}_\sigma, \nu_\sigma) \le \mathcal{S} + 2^{2-2/\mathfrak{p}}\bar{M}^{2/\mathfrak{p}} q_*^{1-2/\mathfrak{p}}.$$

Taking square roots and using $2^{1-1/\mathfrak{p}} \le 2$ gives (126). The explicit bound (127) follows by substituting Proposition E.17 into (118). □

E.4 Tube-adapted velocity class and approximation

This subsection identifies the approximation structure behind the late-regime estimator. Once the FM marginal is restricted to a thin tube, the population velocity admits a low-complexity tube-coordinate description: an intrinsic tangential coefficient on $\mathcal{M}$, the geometric normal offset $P_N(y)y$, and an approximately linear normal contraction in the tube coordinate. We encode this structure in a tube-adapted class with finite-resolution chartwise coefficients, then verify that the class has sufficient continuity and achieves small late tube-restricted risk on σ-scale tubes. The resulting approximation bound is the preliminary of the sieve estimation result in the next subsection.

Definition E.20 (Tube-adapted velocity class)**.** Fix $\alpha \in (0,1]$, a resolution $\varepsilon_{\mathcal{M}} > 0$, a tube radius $r < \tau/2$, and a finite smooth atlas $\{(U_j, \phi_j, \psi_j, E_j)\}_{j=1}^{N_{\mathrm{chart}}}$ on $\mathcal{M}$ with smooth orthonormal tangent frames E_j and subordinate partition of unity $\{\psi_j\}$. Define

$$\mathcal{V}_{\varepsilon_{\mathcal{M}}} := \Big\{ v_t(y,\rho) = \sum_{j=1}^{N_{\mathrm{chart}}} \psi_j(y)\, [E_j(y)\, g_j(t,\phi_j(y)) + P_N(y)y + a_j(t,\phi_j(y))\,\rho] : g_j \in \mathcal{F}_T, a_j \in \mathcal{F}_a \Big\}, \tag{128}$$

where $\rho \in N_y\mathcal{M}$, $E_j(y)g_j(t,\phi_j(y)) \in T_y\mathcal{M}$, and $P_N(y)y + a_j(t,\phi_j(y))\rho \in N_y\mathcal{M}$. Here $\mathcal{F}_T = \mathcal{F}_T(\varepsilon_{\mathcal{M}})$ and $\mathcal{F}_a = \mathcal{F}_a(\varepsilon_{\mathcal{M}})$ are finite-resolution coefficient families in the fixed chart coordinates, with coefficient resolution $\varepsilon_{\mathcal{M}}$. We assume their elements obey the uniform-in-time spatial bounds $\sup_t \|g(t,\cdot)\|_{C^\alpha} \le R_T$ and $\sup_t \|a(t,\cdot)\|_{C^\alpha} \le R_a$, with $R_T = O(\mathrm{diam}(\mathcal{M}))$ and $R_a = O(1/\sigma)$. We also assume the corresponding approximation property on the late window $[t^\star, 1]$: every chartwise tangential coefficient and scalar normal coefficient with the same uniform spatial C^α radius admits an element of $\mathcal{F}_T$ or $\mathcal{F}_a$, respectively, with sup-norm error $O(\varepsilon_{\mathcal{M}}^\alpha)$. In addition, $\mathcal{F}_a$ contains, or approximates to this accuracy, the spatially constant scalar coefficient $a_{\mathrm{flat}}(t) := 1 - (1-t)/\sigma_t^2$. The tangential component depends only on the manifold point y; the normal component captures the geometric offset $P_N(y)y$ and a term linear in ρ. Outside the tube, v_t is extended by zero.

Lemma E.21. *Fix $\alpha \in (0,1]$ and $0 < r \le r_0 < \tau/4$. Every $v \in \mathcal{V}_{\varepsilon_{\mathcal{M}}}$ satisfies, for all $t \in [0,1]$ and $x, z \in \mathcal{M}_r$,*

$$\|v_t(z) - v_t(x)\| \le C_{\mathrm{mod}} \Big((1 + R_T + (1+r)R_a)\, \|z - x\| + (R_T + rR_a)\, \|z-x\|^\alpha \Big). \tag{129}$$

Here, C_{mod} is a constant depending only on r_0, $\mathcal{M}$, and the fixed atlas.

Proof. Write $x = y + \rho$ and $z = y' + \rho'$ in tube coordinates, with $y = \pi_{\mathcal{M}}(x)$, $y' = \pi_{\mathcal{M}}(z)$, $\rho = \rho_{\mathcal{M}}(x)$, and $\rho' = \rho_{\mathcal{M}}(z)$. For a fixed chart term set

$$\Xi_j(t; y, \rho) := E_j(y) g_j(t, \phi_j(y)) + P_N(y)y + a_j(t,\phi_j(y))\rho \in \mathbb{R}^d.$$

Then

$$\psi_j(y)\Xi_j(t; y, \rho) - \psi_j(y')\Xi_j(t; y', \rho')$$

is split into the partition-of-unity difference and the ambient coefficient difference, tracking tangent and normal contributions separately. Since the atlas, frames, projectors, and partition functions are smooth on the compact manifold, their Lipschitz constants are uniformly bounded. Hence the tangential part contributes at most

$$C\{R_T \|y - y'\| + R_T \|y - y'\|^\alpha\},$$

while the geometric normal offset $P_N(y)y$ contributes $C\|y - y'\|$. For the linear normal coefficient,

$$a_j(t,\phi_j(y))\rho - a_j(t,\phi_j(y'))\rho' = a_j(t,\phi_j(y))(\rho - \rho') + \big(a_j(t,\phi_j(y)) - a_j(t,\phi_j(y'))\big)\rho',$$

which is bounded by

$$CR_a \|\rho - \rho'\| + CrR_a \|y - y'\|^\alpha,$$

because $\|\rho'\|, \|\rho\| \le r$ and $\|a_j\|_{C^\alpha} \le R_a$. The partition terms are bounded by $C(1 + R_T + rR_a)\|y - y'\|$. Summing over the finite atlas gives

$$\|v_t(z) - v_t(x)\| \le C\Big((1 + R_T + rR_a)\|y - y'\| + (R_T + rR_a)\|y - y'\|^\alpha + R_a \|\rho - \rho'\| \Big).$$

Finally, the projection and normal-coordinate maps are Lipschitz on the fixed tube $\mathcal{M}_{r_0}$, so $\|y - y'\| + \|\rho - \rho'\| \le C\|z - x\|$. Absorbing constants and using $r \le r_0$ yields (129). □

Proposition E.22. *Under Assumptions 1–3, fix $0 < r \le r_0 < \tau/4$. There exists $\sigma_0 = \sigma_0(r_0, \mathcal{M}, f) > 0$ such that, whenever $\sigma \le \sigma_0$, the following holds. For $t \in [t^\star, 1]$ and $x = t(y+\rho) \in \mathcal{M}^{\mathrm{sc}}_{t,r}$, the population velocity has the coordinate decomposition*

$$v_t^{\star,T}(y,\rho) = g_\star(t,y) + r_T(t,y,\rho), \tag{130}$$

$$v_t^{\star,N}(y,\rho) = P_N(y)y + \beta_t h_{\mathcal{M}}(y) + a_{\mathrm{flat}}(t)\rho + r_N(t,y,\rho), \tag{131}$$

where $\beta_t := (1-t)/t^2$, $a_{\mathrm{flat}}(t) := 1-(1-t)/\sigma_t^2$, $g_\star(t,y) := P_T(y)y + \beta_t \nabla_{\mathcal{M}} \log f(y)$, $h_{\mathcal{M}}$ is a C^1 normal field with $\|h_{\mathcal{M}}\|_{L^\infty} \le C\kappa$, and the remainders satisfy $\|r_T\| + \|r_N\| \le C\beta_t(\|\rho\| + \varepsilon_t)$. In particular, on a σ-scale tube $r \le c_{\mathrm{thin}}\sigma$, $\|r_T\| + \|r_N\| \le C\sigma^2$. Here, C is a constant. Moreover, the coefficient fields $g_\star(t,\cdot)$ and $\beta_t h_{\mathcal{M}}(\cdot)$ are uniformly C^α in chart coordinates for every $\alpha \in (0,1]$ on the late window $t \in [t^\star, 1]$.

Proof. Set $u := x/t = y + \rho$ and $\varepsilon := \varepsilon_t = \sigma_t/t$. Combining the score formula (74) with the exact rescaling in Lemma 2.2 gives

$$v_t^\star(x) = u + \beta_t \nabla \log q_\varepsilon(u), \qquad \beta_t = \frac{1-t}{t^2}.$$

From the Fermi-coordinate expansion used in Proposition C.8, on the master tube one may write

$$\log q_\varepsilon(w) = -\frac{\phi(w)}{\varepsilon^2} - \frac{d-k}{2}\log(2\pi\varepsilon^2) + a(w) + b_\varepsilon(w),$$

where $\phi(w) = \delta_{\mathcal{M}}(w)^2/2$, $\|b_\varepsilon\|_{C^1(\mathcal{M}_{r_0})} \le C\varepsilon$, and

$$a(w) := \log f(\pi_{\mathcal{M}}(w)) - \frac{1}{2}\log\det(I - S_{\rho_{\mathcal{M}}(w)}(\pi_{\mathcal{M}}(w))).$$

Since $\nabla\phi(y+\rho) = \rho$ on the tube, we obtain

$$v_t^\star(x) = y + \rho - \frac{\beta_t}{\varepsilon^2}\rho + \beta_t \nabla a(y+\rho) + \beta_t \nabla b_\varepsilon(y+\rho).$$

Because $\beta_t/\varepsilon^2 = (1-t)/\sigma_t^2$, the coefficient of ρ is exactly

$$1 - \frac{1-t}{\sigma_t^2} = a_{\mathrm{flat}}(t).$$

Thus

$$v_t^\star(x) = P_T(y)y + P_N(y)y + a_{\mathrm{flat}}(t)\rho + \beta_t \nabla a(y+\rho) + \beta_t \nabla b_\varepsilon(y+\rho).$$

Since $a \in C^2$ on the fixed tube, $\|\nabla a(y+\rho) - \nabla a(y)\| \le C\|\rho\|$, while $\|\nabla b_\varepsilon\|_{L^\infty} \le C\varepsilon$. Projecting onto $T_y\mathcal{M}$ and $N_y\mathcal{M}$ therefore gives remainders bounded by

$$C\beta_t(\|\rho\| + \varepsilon).$$

At $\rho = 0$, the determinant term in a equals one, so $a(y) = \log f(y)$ and hence

$$P_T(y)\nabla a(y) = \nabla_{\mathcal{M}} \log f(y).$$

Define the normal field

$$h_{\mathcal{M}}(y) := P_N(y)\nabla a(y).$$

For a unit normal direction $\eta \in N_y\mathcal{M}$, differentiating

$$-\frac{1}{2}\log\det(I - sS_\eta(y))$$

at $s = 0$ gives $\langle h_{\mathcal{M}}(y), \eta\rangle = \mathrm{tr}S_\eta(y)/2$; hence $\|h_{\mathcal{M}}\|_{L^\infty} \le C\kappa$. This proves the decompositions (130)–(131) and the stated remainder bound.

It remains to tackle the regularity and thin-tube consequences. Since $\mathcal{M}$ is C^3 and $\log f \in C^2$, the maps $y \mapsto P_T(y)y$, $y \mapsto P_N(y)y$, $y \mapsto \nabla_{\mathcal{M}} \log f(y)$, and $y \mapsto h_{\mathcal{M}}(y)$ are C^1 in each fixed chart. Since β_t is uniformly bounded on the late window, $g_\star(t,\cdot)$ and $\beta_t h_{\mathcal{M}}(\cdot)$ are uniformly C^α in chart coordinates for every $\alpha \in (0,1]$ on $t \in [t^\star, 1]$. Finally, $t \in [t^\star, 1]$ implies $t \ge 1/2$, $1-t \le \sigma$, $\beta_t \le 4\sigma$, and $\varepsilon_t \le C\sigma$. If $r \le c_{\mathrm{thin}}\sigma$, then $\|\rho\| \le r$ gives

$$\|r_T\| + \|r_N\| \le C\beta_t(\|\rho\| + \varepsilon_t) \le C\sigma^2,$$

which completes the proof. □

Corollary E.23 (Thin-tube approximation)**.** *Fix $\alpha \in (0,1]$, $0 < r \le r_0 < \tau/4$, and suppose $\sigma \le \sigma_0$ and $r \le c_{\mathrm{thin}}\sigma$, where σ_0 is as in Proposition E.22. Then the best-in-class element $v^\dagger \in \mathcal{V}_{\varepsilon_{\mathcal{M}}}$ satisfies*

$$\mathcal{R}_r^{\mathrm{late}}(v^\dagger) \le C\,\varepsilon_{\mathcal{M}}^{2\alpha} + C(1+\kappa^2)\sigma^2, \tag{132}$$

for some constant C.

Proof. Apply Proposition E.22 and construct a comparison element in $\mathcal{V}_{\varepsilon_{\mathcal{M}}}$. By the finite-resolution approximation property in Definition E.20, the chartwise C^α coefficient field $g_\star(t,\cdot)$ is approximated uniformly on the late window with error $O(\varepsilon_{\mathcal{M}}^\alpha)$. Its contribution to the squared tube risk is therefore $O(\varepsilon_{\mathcal{M}}^{2\alpha})$.

For the normal linear term, the scalar coefficient $a_{\mathrm{flat}}(t)$ is independent of y and is contained in, or approximated by, the corresponding coefficient family with error $O(\varepsilon_{\mathcal{M}}^\alpha)$. Since this coefficient multiplies ρ and $\|\rho\| \le r$ on $\mathcal{M}_r$, the associated contribution is at most $Cr^2\varepsilon_{\mathcal{M}}^{2\alpha}$, which is absorbed into $C\varepsilon_{\mathcal{M}}^{2\alpha}$ for $r \le c_{\mathrm{thin}}\sigma$ and $\sigma \le 1$.

It remains to account for the terms in the structural expansion that are not represented by the class. The curvature-induced offset satisfies $\|\beta_t h_{\mathcal{M}}(y)\| \le C\kappa\sigma$ on the late thin tube, and hence contributes at most $C\kappa^2\sigma^2$ to $\mathcal{R}_r^{\mathrm{late}}$. The residual terms satisfy $\|r_T\| + \|r_N\| \le C\sigma^2$, contributing $O(\sigma^4)$. The constructed comparison element thus has risk bounded by the right-hand side of (132); since $v^\dagger$ is best in class, the same bound holds for $v^\dagger$. Using $\sigma \le 1$ completes the proof. □

E.5 Intrinsic regression on thin tubes

This subsection provides the statistical risk controls for the two pieces of the time-split estimator. In the late regime, tube localization reduces the effective regression problem to estimating chartwise coefficients on the k-dimensional manifold, so the relevant sieve complexity scales as h^{-k} rather than with the ambient dimension. In the early regime, the data are not localized near $\mathcal{M}$; we instead use the posterior-mean denoiser form to transfer approximation and entropy control from a source-density sieve on $\mathcal{M}$ to the early velocity field. Together these two arguments give intrinsic-dimension rates for the late and early components that are assembled in the tube-risk theorem.

E.5.1 Late-regime sieve and ERM

The class $\mathcal{V}_{\varepsilon_{\mathcal{M}}}$ above isolates the tube-coordinate approximation structure without committing to a particular finite-dimensional parametrization. For statistical estimation, we now instantiate the same template by a concrete piecewise-constant sieve with mesh size h; this makes the metric entropy and parameter count explicitly intrinsic, of order h^{-k}.

Definition E.24 (Thin-tube time-factored late sieve)**.** Fix $\alpha \in (0,1]$ and mesh size $h \in (0,1)$. Use the fixed atlas $\{(U_j,\phi_j,\psi_j,E_j)\}_{j=1}^{N_{\mathrm{chart}}}$ from Definition E.20, independent of h. For each chart, choose a measurable partition $\phi_j(U_j) = \bigcup_{\ell=1}^{L_j} Q_{j,\ell}$ with $\mathrm{diam}(Q_{j,\ell}) \le h$ and $L_j \le Ch^{-k}$. Let $R_0, R_1 < \infty$ be fixed coefficient radii, chosen large enough to contain the chartwise coefficients used to approximate $P_T(y)y$ and $\nabla_{\mathcal{M}} \log f(y)$ on the fixed atlas. With $\beta_t := (1-t)/t^2$ and $a_{\mathrm{flat}}(t) := 1 - (1-t)/\sigma_t^2$, define the late sieve by

$$\mathcal{V}_h^{\mathrm{late}} := \left\{ v_t(y,\rho) = \sum_{j=1}^{N_{\mathrm{chart}}} \sum_{\ell=1}^{L_j} \psi_j(y)\mathbf{1}_{\{\phi_j(y)\in Q_{j,\ell}\}} \left[E_j(y)(c_{j,\ell}^{(0)} + \beta_t c_{j,\ell}^{(1)}) + P_N(y)y + a_{\mathrm{flat}}(t)\rho \right] \right\}, \tag{133}$$

where $\rho \in N_y\mathcal{M}$ and the coefficients satisfy $c_{j,\ell}^{(0)}, c_{j,\ell}^{(1)} \in \mathbb{R}^k$, $\|c_{j,\ell}^{(0)}\| \le R_0$, and $\|c_{j,\ell}^{(1)}\| \le R_1$. As in Definition E.20, we regard each element as an ambient vector field on the tube by setting $v_t(x) = v_t(\pi_{\mathcal{M}}(x), \rho_{\mathcal{M}}(x))$ for $x \in \mathcal{M}_r$, and extend it by zero outside the tube.

Proposition E.25 (Late-regime ERM on thin tubes)**.** *Under Assumptions 1–3, fix $\alpha \in (0,1]$, $0 < r \le r_0 < \tau/4$, and suppose $\sigma \le \sigma_0$ and $r \le c_{\mathrm{thin}}\sigma$, where σ_0 is as in Proposition E.22. Let $\widehat{v}_h^{\mathrm{late}}$ minimize the late empirical regression loss*

$$\widehat{L}_{\mathrm{late}}(v) := \frac{1}{n}\sum_{i=1}^n \int_{t^\star}^1 \left\| v_t(X_t^{(i)}) - (X_1^{(i)} - X_0^{(i)}) \right\|^2 \mathbf{1}_{\{X_t^{(i)} \in \mathcal{M}_r\}}\,\mathrm{d}t \tag{134}$$

over $\mathcal{V}_h^{\mathrm{late}}$. Then

$$\mathbb{E}\Big[\mathcal{R}_r^{\mathrm{late}}(\widehat{v}_h^{\mathrm{late}}) - \inf_{v\in\mathcal{V}_h^{\mathrm{late}}} \mathcal{R}_r^{\mathrm{late}}(v)\Big] \le C\frac{h^{-k}}{n}\mathrm{polylog}(n,d,1/\sigma). \tag{135}$$

Consequently, with $h = n^{-1/(2\alpha+k)}$,

$$\mathbb{E}[\mathcal{R}_r^{\mathrm{late}}(\widehat{v}_h^{\mathrm{late}})] \le C\, n^{-2\alpha/(2\alpha+k)}\mathrm{polylog}(n,d,1/\sigma) + C(1+\kappa^2)\sigma^2. \tag{136}$$

Proof. Write the free coefficient array as $c = (c_{j,\ell}^{(0)}, c_{j,\ell}^{(1)})_{j,\ell} \in \mathbb{R}^{M_h}$, where

$$M_h = 2k\sum_{j=1}^{N_{\mathrm{chart}}} L_j = O(h^{-k}).$$

The normal part $P_N(y)y + a_{\mathrm{flat}}(t)\rho$ is fixed across the sieve, while the tangential part is linear in c through the bounded features $\psi_j(y)E_j(y)\mathbf{1}_{\{\phi_j(y)\in Q_{j,\ell}\}}$ and the bounded time factor β_t on $[t^\star,1]$. Thus the sieve has metric entropy (in the sense of log-covering number under an ε-net; see, e.g., [47, Chapter 5.1][46, Chapter 2.7]) $O(h^{-k}\log(C/\varepsilon))$ in the relevant space–time L^2 norm and a bounded envelope on the tube.

The population version of (134) is a least-squares regression criterion whose conditional mean is $v_t^\star(X_t)$. Since the tube indicator is measurable with respect to X_t, the standard regression identity gives, for every candidate v,

$$\mathbb{E}[\widehat{L}_{\mathrm{late}}(v)] = \mathcal{R}_r^{\mathrm{late}}(v) + \int_{t^\star}^1 \mathbb{E}\Big[\,\|X_1 - X_0 - v_t^\star(X_t)\|^2\,\mathbf{1}_{\{X_t\in\mathcal{M}_r\}}\Big]\mathrm{d}t.$$

The second term is independent of v, so excess population loss over the sieve is exactly excess late tube risk. Since the normal component $P_N(y)y + a_{\mathrm{flat}}(t)\rho$ is fixed across the sieve, the stochastic optimization is effectively over the M_h tangential coefficients. Applying the standard finite-dimensional least-squares sieve bound (e.g., [46]) to (134) gives (135); the response $X_1 - X_0$ is sub-Gaussian under the CFM sampling model, and the tube restriction controls the sieve envelope.

For approximation, the same chartwise Hölder argument as in Corollary E.23, now with mesh size h and with the time-factored tangential form $c^{(0)} + \beta_t c^{(1)}$, gives

$$\inf_{v\in\mathcal{V}_h^{\mathrm{late}}} \mathcal{R}_r^{\mathrm{late}}(v) \le Ch^{2\alpha} + C(1+\kappa^2)\sigma^2.$$

Combining this approximation bound with (135) and choosing $h = n^{-1/(2\alpha+k)}$ yields (136). □

E.5.2 Early-regime intrinsic route via the posterior-mean denoiser

The early velocity $v_t^\star$ is not tube-localized: p_t spreads over all of $\mathbb{R}^d$ for small t. Nevertheless, the exact posterior-mean formula connects the early velocity to a family of Gaussian denoisers indexed by the manifold source measure, enabling an intrinsic early construction.

Proposition E.26. *For $\mu := f\,\mathrm{Vol}_{\mathcal{M}}$, define*

$$q_\lambda(z) := \int_{\mathcal{M}} \varphi_{\lambda^2}(z-y)\,\mathrm{d}\mu(y), \qquad J_\lambda(z) := \int_{\mathcal{M}} y\,\varphi_{\lambda^2}(z-y)\,\mathrm{d}\mu(y), \qquad \eta_\lambda(z) := \frac{J_\lambda(z)}{q_\lambda(z)}. \tag{137}$$

For $t\in(0,1]$, set $\lambda_t := \sigma_t/t$. Then the posterior mean in Proposition 2.1 satisfies

$$m_t(x) = \eta_{\lambda_t}(x/t), \tag{138}$$

and therefore

$$v_t^\star(x) = \alpha_t x + \frac{1-t}{\sigma_t^2}\eta_{\lambda_t}(x/t). \tag{139}$$

Moreover, for any comparison family $\widetilde{\eta}$ and $\widetilde{m}_t(x) := \widetilde{\eta}_{\lambda_t}(x/t)$,

$$\|\widetilde{m}_t - m_t\|_{L^2(p_t)}^2 = \|\widetilde{\eta}_{\lambda_t} - \eta_{\lambda_t}\|_{L^2(q_{\lambda_t})}^2. \tag{140}$$

Proof. Eq. (138) is the posterior-mean formula from Proposition 2.1 after the rescaling $z = x/t$ and $\lambda_t = \sigma_t/t$; equivalently, the common Gaussian factor gives the ratio $J_{\lambda_t}(x/t)/q_{\lambda_t}(x/t)$. Eq. (139) then follows from (6). Finally, Lemma 2.2 gives $p_t(x) = t^{-d} q_{\lambda_t}(x/t)$, so the change of variables $x = tz$ proves (140). □

Proposition E.27. *Let $\widetilde{\mu} := \widetilde{f}\,\mathrm{Vol}_{\mathcal{M}}$ with*

$$\int_{\mathcal{M}} \widetilde{f}\,\mathrm{dVol}_{\mathcal{M}} = 1, \qquad \widetilde{f} \geq 0,$$

and define $\chi^2(\widetilde{\mu} \mid \mu) := \int_{\mathcal{M}} (\widetilde{f} - f)^2/f\,\mathrm{dVol}_{\mathcal{M}}$. For $\widetilde{\mu}$, write

$$q_\lambda^{\widetilde{\mu}}(z) := \int_{\mathcal{M}} \varphi_{\lambda^2}(z-y)\,\mathrm{d}\widetilde{\mu}(y), \qquad J_\lambda^{\widetilde{\mu}}(z) := \int_{\mathcal{M}} y\,\varphi_{\lambda^2}(z-y)\,\mathrm{d}\widetilde{\mu}(y).$$

Then, for every $\lambda > 0$,

$$\int_{\mathbb{R}^d} \frac{|q_\lambda^{\widetilde{\mu}}(z) - q_\lambda(z)|^2}{q_\lambda(z)}\,\mathrm{d}z \leq \chi^2(\widetilde{\mu} \mid \mu), \tag{141}$$

$$\int_{\mathbb{R}^d} \frac{\left\| J_\lambda^{\widetilde{\mu}}(z) - J_\lambda(z) \right\|^2}{q_\lambda(z)}\,\mathrm{d}z \leq R_{\mathcal{M}}^2\,\chi^2(\widetilde{\mu} \mid \mu), \tag{142}$$

where $R_{\mathcal{M}} = \sup_{y \in \mathcal{M}} \|y\|$.

Proof. Write $a(y) := (\widetilde{f}(y) - f(y))/f(y)$. Then

$$q_\lambda^{\widetilde{\mu}}(z) - q_\lambda(z) = \int_{\mathcal{M}} \varphi_{\lambda^2}(z-y) a(y)\,\mathrm{d}\mu(y).$$

By Cauchy–Schwarz in $L^2(\mu)$,

$$|q_\lambda^{\widetilde{\mu}}(z) - q_\lambda(z)|^2 \leq q_\lambda(z) \int_{\mathcal{M}} \varphi_{\lambda^2}(z-y) a(y)^2\,\mathrm{d}\mu(y).$$

Dividing by $q_\lambda(z)$ and integrating over z gives (141). The same argument applied to

$$J_\lambda^{\widetilde{\mu}}(z) - J_\lambda(z) = \int_{\mathcal{M}} y\,\varphi_{\lambda^2}(z-y) a(y)\,\mathrm{d}\mu(y)$$

gives (142), using $\|y\| \leq R_{\mathcal{M}}$ on $\mathcal{M}$. □

Corollary E.28. *Suppose a source class $\mathfrak{M}_N$ of probability measures of the form $\widetilde{f}\,\mathrm{Vol}_{\mathcal{M}}$ satisfies*

$$\inf_{\widetilde{\mu} \in \mathfrak{M}_N} \chi^2(\widetilde{\mu} \mid \mu) \leq C_\chi N^{-2\beta/k}.$$

For $\widetilde{\mu} \in \mathfrak{M}_N$, let $\eta_\lambda^{\widetilde{\mu}} := J_\lambda^{\widetilde{\mu}}/q_\lambda^{\widetilde{\mu}}$ and set

$$\mathcal{H}_N^{\mathrm{src}} := \{\eta^{\widetilde{\mu}} : \widetilde{\mu} \in \mathfrak{M}_N\}.$$

Then

$$\inf_{\bar{\eta} \in \mathcal{H}_N^{\mathrm{src}}} \sup_{\lambda > 0} \|\bar{\eta}_\lambda - \eta_\lambda\|_{L^2(q_\lambda)}^2 \leq 4R_{\mathcal{M}}^2 C_\chi N^{-2\beta/k}. \tag{143}$$

Proof. For any $\widetilde{\mu}$, $\eta_\lambda^{\widetilde{\mu}}(z)$ is a barycenter of points in $\mathcal{M}$, so $\|\eta_\lambda^{\widetilde{\mu}}(z)\| \leq R_{\mathcal{M}}$. Moreover,

$$\eta_\lambda^{\widetilde{\mu}} - \eta_\lambda = \frac{J_\lambda^{\widetilde{\mu}} - J_\lambda}{q_\lambda} + \eta_\lambda^{\widetilde{\mu}} \frac{q_\lambda - q_\lambda^{\widetilde{\mu}}}{q_\lambda}.$$

Taking square, multiplying by q_λ, integrating, and using $(a+b)^2 \leq 2a^2 + 2b^2$ gives

$$\left\| \eta_\lambda^{\widetilde{\mu}} - \eta_\lambda \right\|_{L^2(q_\lambda)}^2 \leq 2 \int \frac{\left\| J_\lambda^{\widetilde{\mu}} - J_\lambda \right\|^2}{q_\lambda}\,\mathrm{d}z + 2R_{\mathcal{M}}^2 \int \frac{|q_\lambda^{\widetilde{\mu}} - q_\lambda|^2}{q_\lambda}\,\mathrm{d}z.$$

Proposition E.27 bounds the right-hand side by $4R_{\mathcal{M}}^2 \chi^2(\widetilde{\mu} \mid \mu)$. Taking the supremum over $\lambda > 0$ and then the infimum over $\widetilde{\mu} \in \mathfrak{M}_N$ proves (143). □

Definition E.29 (Fixed-atlas source-density sieve)**.** Fix $\beta \in (0,1]$. Choose a finite smooth atlas $\{(V_j, Y_j, \psi_j)\}_{j=1}^{N_{\mathrm{src}}}$ on $\mathcal{M}$, where $Y_j : V_j \subset \mathbb{R}^k \to U_j \subset \mathcal{M}$ is a chart map and $\{\psi_j\}$ is a smooth partition of unity subordinate to $\{U_j\}$. Let

$$J_j(u) := \sqrt{\det(DY_j(u)^\top DY_j(u))}$$

be the volume Jacobian. For mesh size $h \in (0,1)$, partition each V_j into measurable cells $\{Q_{j,\ell}\}_{\ell=1}^{L_j}$ of diameter at most h, with $L_j = O(h^{-k})$. The source sieve $\mathcal{F}_h^{\mathrm{src}}$ consists of densities of the form

$$\widetilde{f}_h(y) := \sum_{j=1}^{N_{\mathrm{src}}} \mathbf{1}_{U_j}(y)\, \frac{b_{j,h}(Y_j^{-1}(y))}{J_j(Y_j^{-1}(y))}, \tag{144}$$

where each $b_{j,h}$ is piecewise constant on $\{Q_{j,\ell}\}_\ell$, the cell values are nonnegative, and the normalization condition

$$\sum_{j=1}^{N_{\mathrm{src}}} \int_{V_j} b_{j,h}(u)\,\mathrm{d}u = 1$$

is imposed. We write

$$\mathcal{F}_h^{\mathrm{src},+} := \left\{ \widetilde{f}_h \in \mathcal{F}_h^{\mathrm{src}} : \frac{f_{\min}}{2} \le \widetilde{f}_h \le 2f_{\max} \right\}$$

for the positive subclass, and define the corresponding source-measure classes

$$\mathfrak{M}_h^{\mathrm{src}} := \{\widetilde{f}_h \,\mathrm{Vol}_{\mathcal{M}} : \widetilde{f}_h \in \mathcal{F}_h^{\mathrm{src}}\}, \qquad \mathfrak{M}_h^{\mathrm{src},+} := \{\widetilde{f}_h \,\mathrm{Vol}_{\mathcal{M}} : \widetilde{f}_h \in \mathcal{F}_h^{\mathrm{src},+}\}.$$

Proposition E.30. *Under Assumptions 1 and 2, for every $\beta \in (0,1]$ and sufficiently small h, the source sieve contains an element $\widetilde{f}_h \in \mathcal{F}_h^{\mathrm{src}}$ satisfying*

$$\chi^2(\widetilde{f}_h \,\mathrm{Vol}_{\mathcal{M}} \mid f \,\mathrm{Vol}_{\mathcal{M}}) \le C_{\mathrm{src},\beta}\, h^{2\beta}, \tag{145}$$

where $C_{\mathrm{src},\beta}$ is a constant. Equivalently, with $N_h := \sum_j L_j = O(h^{-k})$,

$$\inf_{\widetilde{f} \in \mathcal{F}_h^{\mathrm{src}}} \chi^2(\widetilde{f} \,\mathrm{Vol}_{\mathcal{M}} \mid f \,\mathrm{Vol}_{\mathcal{M}}) \le C_{\mathrm{src},\beta} N_h^{-2\beta/k}. \tag{146}$$

Moreover, for small enough h, the approximating element may be chosen in the positive subclass $\mathfrak{M}_h^{\mathrm{src},+}$.

Proof. For the true density, define chart amplitudes

$$b_j(u) := \psi_j(Y_j(u)) f(Y_j(u)) J_j(u).$$

Since $f \in C^2(\mathcal{M})$ and the atlas is smooth and fixed, each b_j is C^β for every $\beta \in (0,1]$. Let $b_{j,h}$ be the cell-average approximation of b_j on the partition. Then $b_{j,h} \ge 0$,

$$\sum_j \int_{V_j} b_{j,h} = \sum_j \int_{V_j} b_j = 1,$$

and the standard cell-average bound gives

$$\|b_{j,h} - b_j\|_{L^\infty(V_j)} \le C h^\beta.$$

The associated $\widetilde{f}_h$ therefore belongs to $\mathcal{F}_h^{\mathrm{src}}$. Since each J_j is smooth and bounded away from zero on the fixed chart domains,

$$\left\| \widetilde{f}_h - f \right\|_{L^\infty(\mathcal{M})} \le C h^\beta,$$

using $\sum_j \psi_j = 1$. With $f \ge f_{\min} > 0$, this implies

$$\chi^2(\widetilde{f}_h \,\mathrm{Vol}_{\mathcal{M}} \mid f \,\mathrm{Vol}_{\mathcal{M}}) = \int_{\mathcal{M}} \frac{(\widetilde{f}_h - f)^2}{f}\, \mathrm{dVol}_{\mathcal{M}} \le f_{\min}^{-1} \mathrm{Vol}_{\mathcal{M}}(\mathcal{M}) \left\| \widetilde{f}_h - f \right\|_{L^\infty(\mathcal{M})}^2 = C_{\mathrm{src},\beta} h^{2\beta},$$

with $C_{\mathrm{src},\beta} = f_{\min}^{-1} \mathrm{Vol}_{\mathcal{M}}(\mathcal{M}) C^2$. Because $N_h = O(h^{-k})$, this is equivalently (146). Finally, the same L^∞ bound and $f_{\min} \le f \le f_{\max}$ imply $f_{\min}/2 \le \widetilde{f}_h \le 2f_{\max}$ for sufficiently small h. □

Proposition E.31. *Fix $\bar{t} \in (0, t^\star]$ and set*

$$\Lambda_{\mathrm{early}} := [\lambda_{\bar{t}}, \infty), \qquad \lambda_{\bar{t}} := \sigma_{\bar{t}}/\bar{t}, \qquad w_{\sigma,\bar{t}}(\lambda) := \frac{\sqrt{\lambda^2 - \sigma^2}}{\lambda^3} \mathbf{1}_{\{\lambda \geq \lambda_{\bar{t}}\}}.$$

Let $\mathcal{H}_N$ be a denoiser class on $\Lambda_{\mathrm{early}} \times \mathbb{R}^d$, and suppose that an estimator $\widehat{\eta} \in \mathcal{H}_N$ satisfies the following two properties for some $\eta^\dagger \in \mathcal{H}_N$ and some constants $C_{\mathrm{app}}, C_{\mathrm{est}}$:

(i) approximation:

$$\sup_{\lambda \in \Lambda_{\mathrm{early}}} \left\| \eta^\dagger_\lambda - \eta_\lambda \right\|^2_{L^2(q_\lambda)} \leq C_{\mathrm{app}} N^{-2\beta/k}; \tag{147}$$

(ii) estimation:

$$\mathbb{E} \int_{\lambda_{\bar{t}}}^\infty w_{\sigma,\bar{t}}(\lambda) \left\| \widehat{\eta}_\lambda - \eta^\dagger_\lambda \right\|^2_{L^2(q_\lambda)} \, \mathrm{d}\lambda \leq C_{\mathrm{est}} \frac{N}{n} \mathrm{polylog}(n, d, 1/\sigma). \tag{148}$$

Define the induced early velocity

$$\widehat{v}^{\mathrm{early}}_t(x) := \alpha_t x + \frac{1-t}{\sigma_t^2} \widehat{\eta}_{\lambda_t}(x/t), \qquad \lambda_t = \sigma_t/t.$$

Then the exact transfer identity

$$\int_0^{\bar{t}} \left\| \widehat{v}^{\mathrm{early}}_t - v^\star_t \right\|^2_{L^2(p_t)} \, \mathrm{d}t = \int_{\lambda_{\bar{t}}}^\infty w_{\sigma,\bar{t}}(\lambda) \left\| \widehat{\eta}_\lambda - \eta_\lambda \right\|^2_{L^2(q_\lambda)} \, \mathrm{d}\lambda \tag{149}$$

holds. Consequently, if $N \asymp (\sigma^{-1} n)^{k/(2\beta+k)}$, then

$$\mathbb{E} \int_0^{\bar{t}} \left\| \widehat{v}^{\mathrm{early}}_t - v^\star_t \right\|^2_{L^2(p_t)} \, \mathrm{d}t \leq C \, \sigma^{-k/(2\beta+k)} n^{-2\beta/(2\beta+k)} \mathrm{polylog}(n, d, 1/\sigma), \tag{150}$$

for some constant C.

Proof. By Proposition E.26,

$$\left\| \widehat{\eta}_{\lambda_t}(\cdot/t) - \eta_{\lambda_t}(\cdot/t) \right\|^2_{L^2(p_t)} = \left\| \widehat{\eta}_{\lambda_t} - \eta_{\lambda_t} \right\|^2_{L^2(q_{\lambda_t})}.$$

Since the affine term $\alpha_t x$ cancels in the velocity difference,

$$\left\| \widehat{v}^{\mathrm{early}}_t - v^\star_t \right\|^2_{L^2(p_t)} = \left(\frac{1-t}{\sigma_t^2} \right)^2 \left\| \widehat{\eta}_{\lambda_t} - \eta_{\lambda_t} \right\|^2_{L^2(q_{\lambda_t})}.$$

Set $\zeta = (1-t)/t$. Then $\lambda_t = \sqrt{\zeta^2 + \sigma^2}$ and the change of variables $\lambda = \sqrt{\zeta^2 + \sigma^2}$ gives

$$\left(\frac{1-t}{\sigma_t^2} \right)^2 \mathrm{d}t = -\frac{\sqrt{\lambda^2 - \sigma^2}}{\lambda^3} \, \mathrm{d}\lambda,$$

with $t \in (0, \bar{t}]$ corresponding to $\lambda \in [\lambda_{\bar{t}}, \infty)$. This proves (149).

Using $(a+b)^2 \leq 2a^2 + 2b^2$ in $L^2(q_\lambda)$ and integrating against $w_{\sigma,\bar{t}}$,

$$\begin{aligned} \mathbb{E} \int_{\lambda_{\bar{t}}}^\infty w_{\sigma,\bar{t}}(\lambda) \left\| \widehat{\eta}_\lambda - \eta_\lambda \right\|^2_{L^2(q_\lambda)} \, \mathrm{d}\lambda &\leq 2 C_{\mathrm{est}} \frac{N}{n} \mathrm{polylog}(n, d, 1/\sigma) \\ &\quad + 2 C_{\mathrm{app}} N^{-2\beta/k} \int_{\lambda_{\bar{t}}}^\infty w_{\sigma,\bar{t}}(\lambda) \, \mathrm{d}\lambda. \end{aligned}$$

Because $\bar{t} \leq t^\star = 1/(1+\sigma)$, $\lambda_{\bar{t}} \geq \lambda_{t^\star} = \sqrt{2}\,\sigma$, and hence

$$\int_{\lambda_{\bar{t}}}^\infty w_{\sigma,\bar{t}}(\lambda) \, \mathrm{d}\lambda \leq \int_{\sqrt{2}\sigma}^\infty \frac{\sqrt{\lambda^2 - \sigma^2}}{\lambda^3} \, \mathrm{d}\lambda \leq \frac{C}{\sigma}.$$

Combining with (149) yields

$$\mathbb{E} \int_0^{\bar{t}} \left\| \widehat{v}^{\mathrm{early}}_t - v^\star_t \right\|^2_{L^2(p_t)} \, \mathrm{d}t \leq C \left(\sigma^{-1} C_{\mathrm{app}} N^{-2\beta/k} + C_{\mathrm{est}} \frac{N}{n} \mathrm{polylog}(n, d, 1/\sigma) \right).$$

Balancing the two terms gives $N \asymp (\sigma^{-1} n)^{k/(2\beta+k)}$, and substitution gives (150). □

Remark E.32. The factor $\sigma^{-k/(2\beta+k)}$ in (150) originates from the weight mass $\int w \, \mathrm{d}\lambda \asymp \sigma^{-1}$, which is a structural feature of the denoiser-family approach: as $\sigma \downarrow 0$, the denoiser $\eta_\lambda(z) \to \pi_{\mathcal{M}}(z)$ and a nonvanishing approximation error floor is charged over the full band $\lambda \in [\sqrt{2}\sigma, 1]$ by the weight $w \asymp \lambda^{-2}$. A sharper route would first subtract the geometric projection and learn only the remainder $\eta_\lambda(z) - \pi_{\mathcal{M}}(z) = O(\lambda^2)$; this is not pursued here.

Proposition E.33 (Concrete early ERM from the positive source sieve)**.** *Assume Assumptions 1–3. Fix $\beta \in (0, 1]$ and $\bar{t} \in (0, t^\star]$. For sufficiently small $h \in (0, 1)$, let $\widehat{v}_h^{\mathrm{early,src}}$ minimize the early empirical loss*

$$\frac{1}{n}\sum_{i=1}^{n} \int_0^{\bar{t}} \left\| v_t(X_t^{(i)}) - (X_1^{(i)} - X_0^{(i)}) \right\|^2 \mathrm{d}t \tag{151}$$

over the source-sieve-induced velocity class

$$\mathcal{V}_h^{\mathrm{early,src}} := \{ v_t^{\widetilde{\mu}} : \widetilde{\mu} \in \mathfrak{M}_h^{\mathrm{src},+} \},$$

where $\mathfrak{M}_h^{\mathrm{src},+}$ is the set of source-sieve measures whose densities are bounded between $f_{\min}/2$ and $2f_{\max}$. Then

$$\mathbb{E} \int_0^{\bar{t}} \left\| \widehat{v}_h^{\mathrm{early,src}} - v_t^\star \right\|_{L^2(p_t)}^2 \mathrm{d}t \le C\sigma^{-1}h^{2\beta} + C\frac{h^{-k}}{n}\mathrm{polylog}(n, d, 1/\sigma), \tag{152}$$

for some constant C.

Proof. We separate the deterministic approximation and stochastic estimation terms.

Approximation. By Proposition E.30, for sufficiently small h there exists a source-sieve density $\widetilde{f}_h$ in the positive subclass such that

$$\chi^2(\widetilde{f}_h \,\mathrm{Vol}_{\mathcal{M}} \mid f \,\mathrm{Vol}_{\mathcal{M}}) \le Ch^{2\beta}.$$

Corollary E.28 transfers this source approximation to the corresponding denoiser family: for the induced denoiser $\eta_h^\dagger$,

$$\sup_{\lambda > 0} \left\| \eta_{h,\lambda}^\dagger - \eta_\lambda \right\|_{L^2(q_\lambda)}^2 \le Ch^{2\beta}.$$

Applying the exact early transfer identity (149) to the induced velocity $v^{\widetilde{\mu}_h}$ gives

$$\inf_{v \in \mathcal{V}_h^{\mathrm{early,src}}} \int_0^{\bar{t}} \|v_t - v_t^\star\|_{L^2(p_t)}^2 \, \mathrm{d}t \le Ch^{2\beta} \int_{\lambda_{\bar{t}}}^{\infty} w_{\sigma,\bar{t}}(\lambda) \, \mathrm{d}\lambda.$$

Since $\bar{t} \le t^\star$, we have $\lambda_{\bar{t}} \ge \sqrt{2}\sigma$, and hence

$$\int_{\lambda_{\bar{t}}}^{\infty} w_{\sigma,\bar{t}}(\lambda) \, \mathrm{d}\lambda \le \int_{\sqrt{2}\sigma}^{\infty} \frac{\sqrt{\lambda^2 - \sigma^2}}{\lambda^3} \, \mathrm{d}\lambda \le \frac{C}{\sigma}.$$

Thus

$$\inf_{v \in \mathcal{V}_h^{\mathrm{early,src}}} \int_0^{\bar{t}} \|v_t - v_t^\star\|_{L^2(p_t)}^2 \, \mathrm{d}t \le C\sigma^{-1}h^{2\beta}. \tag{153}$$

Metric entropy of the induced class. It remains to control the ERM excess over $\mathcal{V}_h^{\mathrm{early,src}}$. Let $\widetilde{f}_1, \widetilde{f}_2 \in \mathcal{F}_h^{\mathrm{src},+}$, set $\widetilde{\mu}_i = \widetilde{f}_i \,\mathrm{Vol}_{\mathcal{M}}$, and denote the associated denoisers by $\eta_\lambda^{(i)}$. Repeating the χ^2 transfer argument in Proposition E.27, using $\widetilde{\mu}_1$ as the reference measure, yields

$$\sup_{\lambda > 0} \left\| \eta_\lambda^{(2)} - \eta_\lambda^{(1)} \right\|_{L^2(q_\lambda^{(1)})}^2 \le C\chi^2(\widetilde{\mu}_2 \mid \widetilde{\mu}_1).$$

The positivity bounds $f_{\min}/2 \le \widetilde{f}_i \le 2f_{\max}$ imply both

$$\chi^2(\widetilde{\mu}_2 \mid \widetilde{\mu}_1) \le C \left\| \widetilde{f}_2 - \widetilde{f}_1 \right\|_{L^2(\mathrm{Vol}_{\mathcal{M}})}^2$$

and uniform comparability $cq_\lambda \le q_\lambda^{(1)} \le Cq_\lambda$. Consequently,

$$\sup_{\lambda>0} \left\| \eta_\lambda^{(2)} - \eta_\lambda^{(1)} \right\|_{L^2(q_\lambda)}^2 \le C \left\| \widetilde{f}_2 - \widetilde{f}_1 \right\|_{L^2(\mathrm{Vol}_\mathcal{M})}^2 . \tag{154}$$

Combining (154) with (149) gives the corresponding early velocity metric bound

$$\int_0^{\bar{t}} \left\| v_t^{(2)} - v_t^{(1)} \right\|_{L^2(p_t)}^2 \mathrm{d}t \le C\sigma^{-1} \left\| \widetilde{f}_2 - \widetilde{f}_1 \right\|_{L^2(\mathrm{Vol}_\mathcal{M})}^2 .$$

The class $\mathcal{F}_h^{\mathrm{src},+}$ is a bounded subset of a piecewise-constant atlas sieve with $N_h = O(h^{-k})$ free cell values. Therefore

$$\log \mathcal{N}(\varepsilon, \mathcal{F}_h^{\mathrm{src},+}, L^2(\mathrm{Vol}_\mathcal{M})) \le CN_h \log(C/\varepsilon),$$

and the same entropy bound holds for $\mathcal{V}_h^{\mathrm{early,src}}$ in the early space–time regression norm after the harmless rescaling $\varepsilon \mapsto \sqrt{\sigma}\,\varepsilon$, which changes only logarithmic factors.

ERM excess. The response $X_1 - X_0$ is sub-Gaussian under the CFM sampling model, and the induced velocities have a controlled envelope on the early window: the denoisers satisfy $\|\eta_\lambda^{\widetilde{\mu}}\| \le R_\mathcal{M}$, while the affine term is common to all candidates. Standard finite-dimensional least-squares sieve bounds, applied with the entropy above, give

$$\mathbb{E}\left[\int_0^{\bar{t}} \left\| \widehat{v}_h^{\mathrm{early,src}} - v_t^\star \right\|_{L^2(p_t)}^2 \mathrm{d}t - \inf_{v \in \mathcal{V}_h^{\mathrm{early,src}}} \int_0^{\bar{t}} \|v_t - v_t^\star\|_{L^2(p_t)}^2 \, \mathrm{d}t \right] \le C \frac{N_h}{n} \mathrm{polylog}(n, d, 1/\sigma).$$

Since $N_h = O(h^{-k})$, combining this stochastic bound with the approximation bound (153) proves (152). □

Corollary E.34. *Under Assumptions 1–3, fix $\alpha \in (0, 1]$ and set $\bar{t} = t^\star$. Choose the polynomially balanced mesh*

$$h = \sigma^{1/(2\alpha+k)} n^{-1/(2\alpha+k)},$$

and assume n is large enough that this h lies in the admissible small-mesh regime of Proposition E.33. Then

$$\mathbb{E} \int_0^{t^\star} \left\| \widehat{v}_h^{\mathrm{early,src}} - v_t^\star \right\|_{L^2(p_t)}^2 \mathrm{d}t \le C\sigma^{-k/(2\alpha+k)} n^{-2\alpha/(2\alpha+k)} \mathrm{polylog}(n, d, 1/\sigma). \tag{155}$$

Proof. Apply Proposition E.33 with $\beta = \alpha$ and $\bar{t} = t^\star$. With the stated h, the two polynomial factors satisfy

$$\sigma^{-1} h^{2\alpha} = \sigma^{-k/(2\alpha+k)} n^{-2\alpha/(2\alpha+k)}, \qquad \frac{h^{-k}}{n} = \sigma^{-k/(2\alpha+k)} n^{-2\alpha/(2\alpha+k)}.$$

The logarithmic factor multiplying the stochastic term is absorbed into the displayed $\mathrm{polylog}(n, d, 1/\sigma)$ term, yielding (155). □

E.6 Tube-risk theorem

This subsection assembles the early and late regression controls into the final statistical guarantee for the time-split estimator. The guarantee is stated for the tube-restricted squared velocity risk from Definition E.1, which is the quantity directly controlled by the approximation and ERM arguments above. The theorem first abstracts the early component through a prerequisite risk bound, while the subsequent corollary plugs in the concrete source-sieve construction from Corollary E.34. This gives the completed intrinsic-dimension tube-risk bound. The stopped synthesis in Section E.3 converts tube risk to ambient W_2 control when the estimator also satisfies the stated early-coupling, density-ratio, and terminal-moment assumptions.

Theorem E.35 (Time-split tube-risk synthesis with an early prerequisite)**.** *Under Assumptions 1–3, fix $\alpha \in (0, 1]$, $0 < r \le r_0 < \tau/4$, and suppose $\sigma \le \sigma_0$ and $r \le c_{\mathrm{thin}}\sigma$, where σ_0 is as in Proposition E.22. Let $\widehat{v}$ be a time-split estimator whose late component on $[t^\star, 1]$ is $\widehat{v}_{h_{\mathrm{late}}}^{\mathrm{late}}$ from Proposition E.25 with*

$$h_{\mathrm{late}} = n^{-1/(2\alpha+k)}.$$

Assume that its early component satisfies, for some prefactor A_{early},

$$\mathbb{E}\int_0^{t^\star} \|\widehat{v}_t(X_t) - v_t^\star(X_t)\|^2 \, \mathbf{1}_{\{X_t \in \mathcal{M}_r\}} \, \mathrm{d}t \le A_{\mathrm{early}}\, n^{-2\alpha/(2\alpha+k)} \mathrm{polylog}(n, d, 1/\sigma), \tag{156}$$

where the expectation is over both the training sample and an independent CFM test draw. Then

$$\mathbb{E}\big[\mathcal{R}_r(\widehat{v})\big] \le C\big(A_{\mathrm{early}} + 1\big) n^{-2\alpha/(2\alpha+k)} \mathrm{polylog}(n, d, 1/\sigma) + C(1+\kappa^2)\sigma^2. \tag{157}$$

The exponent of n is governed by the intrinsic dimension k; in this tube-risk bound, the ambient dimension enters only through logarithmic factors and embedding-dependent constants.

Proof. Taking outer expectation over the estimator randomness and splitting the time interval gives

$$\begin{aligned}\mathbb{E}\big[\mathcal{R}_r(\widehat{v})\big] = \mathbb{E}\int_0^{t^\star} \|\widehat{v}_t(X_t) - v_t^\star(X_t)\|^2 \, \mathbf{1}_{\{X_t \in \mathcal{M}_r\}} \, \mathrm{d}t \\ + \mathbb{E}\big[\mathcal{R}_r^{\mathrm{late}}(\widehat{v}^{\mathrm{late}}_{h_{\mathrm{late}}})\big].\end{aligned}$$

The early term is bounded by (156). Proposition E.25, with $h_{\mathrm{late}} = n^{-1/(2\alpha+k)}$, bounds the late term by

$$Cn^{-2\alpha/(2\alpha+k)}\mathrm{polylog}(n, d, 1/\sigma) + C(1+\kappa^2)\sigma^2.$$

Combining the two bounds proves (157). □

Corollary E.36 (Concrete low-Hölder tube-risk bound)**.** *Under Assumptions 1–3, fix $\alpha \in (0, 1]$, $0 < r \le r_0 < \tau/4$, and suppose $\sigma \le \sigma_0$ and $r \le c_{\mathrm{thin}}\sigma$. Let $\widehat{v}$ be the time-split estimator whose early component is the source-sieve estimator from Corollary E.34 with*

$$h_{\mathrm{early}} = \sigma^{1/(2\alpha+k)} n^{-1/(2\alpha+k)},$$

and whose late component is $\widehat{v}^{\mathrm{late}}_{h_{\mathrm{late}}}$ from Proposition E.25 with $h_{\mathrm{late}} = n^{-1/(2\alpha+k)}$. Assume n is large enough that these mesh choices lie in their admissible regimes. Then

$$\mathbb{E}\big[\mathcal{R}_r(\widehat{v})\big] \le C(\sigma^{-k/(2\alpha+k)} + 1)\, n^{-2\alpha/(2\alpha+k)} \mathrm{polylog}(n, d, 1/\sigma) + C(1+\kappa^2)\sigma^2. \tag{158}$$

The dominant statistical term exhibits the exponent $2\alpha/(2\alpha+k)$ controlled by the intrinsic dimension; curvature enters through the geometric constant and the $O(\sigma^2)$ residual, while ambient-dimensional dependence is only logarithmic or hidden in geometric constants. Eq. (158) *yields Eq.* (23)*.*

Proof. By Corollary E.34,

$$\mathbb{E}\int_0^{t^\star} \|\widehat{v}_t - v_t^\star\|^2_{L^2(p_t)} \, \mathrm{d}t \le C\sigma^{-k/(2\alpha+k)} n^{-2\alpha/(2\alpha+k)} \mathrm{polylog}(n, d, 1/\sigma).$$

The global early $L^2(p_t)$ risk dominates the tube-restricted early risk in (156). Hence Theorem E.35 applies with $A_{\mathrm{early}} = C\sigma^{-k/(2\alpha+k)}$, giving (158). □

E.7 Summary

This appendix section establishes two complementary parts of the manifold-aware theory and isolates the remaining interface between them. First, it gives a concrete intrinsic-dimension tube-risk guarantee for a time-split estimator. Second, it turns the tube-local anisotropic Jacobian bounds into a stopped trajectory-level stability theorem and an ambient transport synthesis. The latter synthesis requires three explicit estimator-side quantities: an early coupling budget, a scaled-tube density-ratio bound, and terminal moments. It does not require a priori confinement of the learned trajectory.

On the estimation side, the tube-risk analysis is self-contained in the thin-tube low-Hölder regime considered here. Corollary E.23 gives the late approximation bound for $r \lesssim \sigma$ and $\alpha \in (0, 1]$, Proposition E.25 gives the matching finite-dimensional late ERM bound, and Corollary E.34 gives the concrete source-sieve early control, with its structural prefactor $\sigma^{-k/(2\alpha+k)}$. Combining these pieces yields Corollary E.36. Thus the sample exponent in the proved tube-risk bound is intrinsic, $2\alpha/(2\alpha+k)$; ambient dimension enters only through logarithmic factors and embedding-dependent constants.

On the stability side, Theorem E.7 applies the block-Jacobian structure along mean-value segments up to a stopping time and shows that normal contraction absorbs the mixed tangential–normal interaction. Proposition E.8 consequently bounds the effective late-time amplification without a negative power of σ. The same stopped energy controls the probability that the learned trajectory separates from the population trajectory, while Proposition E.17 supplies an explicit Gaussian-type bound for population-flow tube exit. Corollary E.19 then converts these bounds into an ambient W_2 bound using a synchronous terminal coupling and a moment-tail argument. In particular, this route neither assumes learned-flow confinement in advance nor requires a separate coupling of tube-conditioned terminal laws.

What remains open is an end-to-end construction satisfying the estimator-side assumptions in Corollary E.19. Proposition E.10 and Corollary E.11 reduce the early contribution to learned-field regularity and a switching-time coupling bound. In the late regime, Lemma E.13 reduces the trajectory-to-reference distribution shift to the scaled-tube density ratio $\Lambda_{\mathrm{sc}}(\bar{r})$. As discussed in Remark E.15, tube-risk control alone does not yield this ratio, and the piecewise-defined late sieve should be refined to a continuous, slope-controlled, normal-matched tube class. Such a refinement is designed to preserve absolute continuity and control density-ratio propagation without changing the intrinsic complexity or approximation order. Terminal moments must likewise be verified for the learned flow. Accordingly, the appendix section proves the statistical tube-risk rate and the stopped dynamical transport mechanism separately, but does not claim that the stated estimator satisfies all assumptions of the transport synthesis. Establishing their compatibility through a suitably regular tube-adapted hypothesis class is the remaining step toward a fully end-to-end intrinsic-dimension transport theorem.

F Numerical verification

This appendix section documents the population-level diagnostic referenced in Section 4. Although narrowly scoped, the diagnostic tests the *analytic* Jacobian property characterized in Theorem 4.4 and Corollary D.6 on a manifold whose posterior moments are available in closed form. Because the diagnostic does not involve training a velocity field, it does not test the stability or estimation results of Section 5. Code is available at `https://github.com/oaksword/fm_vel_reg`.

F.1 Model and exact evaluation

We take

$$\mathcal{M} = S^1 \subset \mathbb{R}^2, \qquad Y = (\cos\omega, \sin\omega)^\top, \qquad X_t = tY + \sigma_t \eta, \qquad \eta \sim \mathcal{N}(0, I_2),$$

where $\omega \in [0, 2\pi)$, η is independent of ω, and $\sigma_t^2 = (1-t)^2 + t^2\sigma^2$ as in (4). The default source density is the uniform arclength density on S^1. To test the block decomposition beyond the rotationally symmetric case, we also consider the nonuniform angular density

$$p_{\mathrm{nu}}(\omega) = \frac{1 + 0.5\cos\omega}{2\pi},$$

which satisfies Assumption 2 with $f_{\min} = 1/(4\pi)$.

The Jacobian is evaluated using the exact identity (7) rather than finite differences; thus, no velocity field is estimated and no optimization is involved. For the uniform circle, consider $x = se_1$, where $s \geq 0$ and $e_1 = (1, 0)^\top$. The posterior density of ω given $X_t = x$ is

$$p(\omega \mid X_t = x) = \frac{\exp(\beta\cos\omega)}{2\pi I_0(\beta)}, \qquad \beta = \frac{ts}{\sigma_t^2},$$

which is the von Mises distribution with mean direction zero and concentration parameter β [15, Chapter 3]. Here I_0 denotes the modified Bessel function of the first kind of order zero. Consequently, $\mathrm{Cov}(Y \mid X_t = x)$ is available in closed form in terms of modified Bessel functions. We instead compute the covariance by deterministic quadrature over $65{,}536$ equally spaced angles, using log-sum-exp stabilization. This approach also accommodates the nonuniform density. We validate the quadrature against the Bessel-function closed form and find agreement to within 10^{-9} across the grid of σ values used below.

We sweep $\sigma \in \{0.2, 0.1, 0.05, 0.025, 0.0125\}$ and evaluate at the switching time $t^\star = 1/(1+\sigma)$, where $\sigma_{t^\star} = \sqrt{2}\,\sigma/(1+\sigma)$ and hence $\varepsilon_{t^\star} = \sqrt{2}\,\sigma$. Since the threshold ε_0 in Proposition 3.1 is not explicit, this sweep should be interpreted as a direct asymptotic check of the predicted rates rather than as a numerical determination of ε_0.

Table 1: Centerline diagnostic at $t^\star$ for $x = t^\star e_1$ with $e_1 = (1,0)^\top$, uniform density. The final column is the prediction $\sqrt{2}(1-\sigma)/2$ for $\sigma_t \|\nabla_x v_t^\star\|_{\mathrm{op}}$ implied by Corollary D.6.

σ	$\sigma_{t^\star}$	λ_N	λ_T	$\sigma_t \|\nabla_x v_t^\star\|_{\mathrm{op}}$	$\sigma_t^3 \|\nabla_x v_t^\star\|_{\mathrm{op}}$	predicted
0.2	0.236	-2.249	1.053	0.530	2.945×10^{-2}	0.566
0.1	0.129	-4.889	1.039	0.629	1.039×10^{-2}	0.636
0.05	0.067	-9.947	1.022	0.670	3.038×10^{-3}	0.672
0.025	0.034	-19.974	1.012	0.689	8.197×10^{-4}	0.689
0.0125	0.017	-39.987	1.006	0.698	2.128×10^{-4}	0.698

Table 2: Projected blocks at $t^\star$ for the nonuniform density $p_{\mathrm{nu}}(\omega) = (1 + 0.5\cos\omega)/(2\pi)$, evaluated at $x = t^\star e_2$ with $e_2 = (0,1)^\top$.

σ	$\sigma_{t^\star}$	$P_T \nabla_x v_t^\star P_T$	$P_N \nabla_x v_t^\star P_N$	$\lvert P_T \nabla_x v_t^\star P_N \rvert / \sigma_t$
0.2	0.236	0.987	-2.249	0.560
0.1	0.129	1.010	-4.889	0.461
0.05	0.067	1.009	-9.947	0.407
0.025	0.034	1.005	-19.974	0.380
0.0125	0.017	1.003	-39.987	0.367

F.2 Centerline sharpness

Table 1 evaluates the Jacobian at the tube centerline $x = t^\star e_1$ with $e_1 = (1,0)^\top$, a choice that is without loss of generality for the uniform circle by rotational symmetry. Here $\lambda_N := e_1^\top \nabla_x v_t^\star(x)\, e_1$ and $\lambda_T := e_2^\top \nabla_x v_t^\star(x)\, e_2$ are the normal and tangential quadratic forms of the Jacobian, with $e_2 = (0,1)^\top$ the unit tangent; these are the quantities controlled in (20), and λ_N is the measured form, to be distinguished from the envelope $\lambda_N^{\mathrm{flat}}$ of (18). Three predicted signatures are visible. First, the normal form is negative and closely tracks the centerline prediction $-(1-\sigma^2)/(2\sigma)$ of Corollary D.6, reaching -39.987 compared with the predicted -39.994 at $\sigma = 0.0125$. Second, the tangential form remains $O(1)$, decreasing slowly toward 1. Third, the scaled operator norm $\sigma_t \|\nabla_x v_t^\star\|_{\mathrm{op}}$ approaches the predicted value $\sqrt{2}(1-\sigma)/2$, which converges to $\sqrt{2}/2$ as $\sigma \to 0$. The relative error decreases from 6.3% at $\sigma = 0.2$ to 0.02% at $\sigma = 0.0125$, while the fitted log-log slope of $\|\nabla_x v_t^\star\|_{\mathrm{op}}$ against σ_t over the sweep is -1.10. The cubic ambient envelope is correspondingly loose: $\sigma_t^3 \|\nabla_x v_t^\star\|_{\mathrm{op}}$ decays by two orders of magnitude over the sweep.

F.3 Tangential–normal block structure

Table 2 repeats the diagnostic for the nonuniform density at $x = t^\star e_2$ with $e_2 = (0,1)^\top$, where the density gradient is tangentially nonzero, and reports the projected blocks of Theorem 4.4. The pattern is the one the theorem predicts: $\|P_T \nabla_x v_t^\star P_T\|_{\mathrm{op}}$ stays $O(1)$, the normal form is $-\Theta(\sigma_t^{-1})$, and the mixed coupling normalized by σ_t remains bounded and in fact decreases across the sweep. At the scaled centerline $x = t^\star e_2$, the normal column coincides exactly with that of Table 1: the reflection $\omega \mapsto \pi - \omega$ leaves the likelihood and the normal coordinate $\sin\omega$ unchanged while reversing the density perturbation $0.5\cos\omega$, so the posterior normal variance equals its uniform-density counterpart.

F.4 Integrated scaling and held-out samples

We then compare time-integrated centerline Jacobian norms with the two candidate scalar envelopes relevant to a Grönwall analysis. Table 3 integrates $\|\nabla_x v_t^\star(te_1)\|_{\mathrm{op}}$ over $t \in [10^{-3}, 1]$. The ratio of the measured integral to $\int \sigma_t^{-1} \mathrm{d}t$ rises from 0.487 to 0.763 across the sweep, and the measured integral increases by approximately $\log 2$ with each halving of σ at the smaller noise levels, consistent with logarithmic growth in $1/\sigma$. The cubic comparison satisfies the exact identity $\sigma^2 \int_0^1 \sigma_t^{-3} \mathrm{d}t = 1 + \sigma$, explaining the fourth column and the $\Theta(\sigma^{-2})$ growth of that envelope. The numerical cutoff at $t = 10^{-3}$ removes only a uniformly bounded early-time contribution and does not affect the small-noise scaling; the ambient Jacobian norm is well defined at $t = 0$. We note that these centerline measurements are a scaling diagnostic, not a spatial Lipschitz envelope or a transport stability bound.

Table 3: Time-integrated comparison on the centerline. The final column is the ratio of the first two integral columns.

σ	$\int \|\nabla_x v_t^\star\|_{\mathrm{op}} \mathrm{d}t$	$\int \sigma_t^{-1} \mathrm{d}t$	$\sigma^2 \int \sigma_t^{-3} \mathrm{d}t$	ratio
0.2	1.199	2.461	1.200	0.487
0.1	1.824	3.082	1.100	0.592
0.05	2.500	3.734	1.050	0.670
0.025	3.189	4.405	1.025	0.724
0.0125	3.881	5.086	1.012	0.763

Finally, we check whether the centerline behavior is also observed at random locations drawn from the late marginal. For 500 held-out samples $X_{t^\star} = t^\star Y + \sigma_{t^\star}\eta$, the empirical median of $\sigma_{t^\star}\|\nabla_x v_{t^\star}^\star(X_{t^\star})\|_{\mathrm{op}}$ equals 0.536, 0.629, 0.670, 0.689, and 0.698 across the sweep. These medians closely match the centerline values in Table 1: the relative difference is approximately 1.1% at $\sigma = 0.2$, and the remaining values agree to the displayed precision. The empirical upper-tail ratio also decreases: the 99th percentile divided by the median is 2.67 at $\sigma = 0.2$ and falls below 1.01 for $\sigma \le 0.1$. These observations support the centerline value as a description of the typical Jacobian norm in the sampled late marginal at the smaller noise levels.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 identify the tube-local $O(\sigma_t^{-1})$ velocity Jacobian bound and the finite-time anisotropic decomposition as the main contributions (Theorems 3.2 and 4.4). Section 5 and Appendix E.7 distinguish the proved stopped transport mechanism and intrinsic-dimension tube-risk rate from an end-to-end transport guarantee for one estimator, which remains open; Appendix F provides narrowly scoped population-level numerical checks.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 6 and Appendix E.7 explain that the intrinsic-dimension tube-risk guarantee and stopped transport synthesis are proved separately: an early coupling budget, a scaled-tube density-ratio bound, and terminal moments still need to be verified for a single estimator. The discussion excludes $\sigma = 0$, non-Gaussian perturbations, and OT-path CFM, while Appendix F explicitly limits its circle diagnostics to population Jacobian properties rather than learned-flow stability or estimation.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The standing geometric and density assumptions are stated in Assumptions 1–3; the principal geometric, stability, and statistical results have numbered statements and proofs in Appendices B–E. The stopped transport synthesis in Corollary E.19 explicitly states its additional estimator-side assumptions, and Appendix E.7 distinguishes this conditional result from the separately proved tube-risk guarantee.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Appendix F specifies the circle model, source densities, noise sweep, evaluation locations, and posterior-covariance quadrature used for the population-Jacobian diagnostics. The linked code repository provides the diagnostic script, reproduction command, and dependency list.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example

(a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Appendix F links the code at `https://github.com/oaksword/fm_vel_reg`; its README provides the reproduction command `python3 run_circle_jacobian.py`, and dependencies are listed in `requirements.txt`. The diagnostics use the specified synthetic circle model and require no external dataset.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Appendix F and the accompanying code specify the source densities, noise levels, time and spatial evaluation settings, quadrature resolution, and held-out sampling procedure. No velocity field is trained, so training splits, optimizer choices, and training hyperparameters are not applicable.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The main diagnostics in Appendix F use deterministic quadrature, validated against the uniform-circle closed form to within 10^{-9}, so repeated-run statistical error bars are not applicable to those evaluations. The additional diagnostic based on 500 held-out samples reports empirical medians and upper-tail ratios without confidence intervals or repeated-sampling uncertainty estimates; we acknowledge this omission as a minor limitation of the supplementary diagnostics, which does not affect the core theoretical conclusions established by the proofs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: Appendix F specifies the numerical diagnostic setup, but the paper does not report the hardware, memory requirements, execution times, or total compute used. The diagnostic script is lightweight and is expected to complete in a short time, although execution time has not been benchmarked or reported.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The research conforms with the NeurIPS Code of Ethics and consists of theoretical analysis and numerical diagnostics on synthetic circle data. It involves no human subjects, personal or sensitive data, or release of trained generative models.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: The paper develops foundational theory for flow matching on manifold-concentrated targets, supported by synthetic circle diagnostics, rather than an application or deployed generative system. We identify no direct path from these results to specific societal benefits or harms beyond those associated with generative modelling research in general.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The released assets consist of circle-diagnostic code and synthetic numerical outputs, not trained generative models, personal data, or scraped material. These assets do not present the high-risk misuse concerns addressed by this question.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The diagnostic repository identifies NumPy, SciPy, pandas, and Matplotlib as dependencies, and its README documents their licenses with links to the upstream license terms. No external datasets or pretrained models are used.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The repository linked in Appendix F documents the diagnostic code and synthetic numerical outputs in its README, including their scope, reproduction command, and dependencies. The README provides the full MIT license covering both the code and numerical outputs.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper involves no crowdsourcing and no research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper involves no research with human subjects, so no IRB approval is required.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs were used for writing, editing, and formatting assistance, and were not an important, original, or non-standard component of the core methodology or theoretical results.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.